UNIVERSITY OF SOUTHAMPTON
Faculty of Physical Sciences and Engineering
School of Electronics and Computer Science

Sustainable Knowledge in Hypertext

by

Mark William Rawlence Anderson
ORCiD: 0000-0001-7396-0721

Thesis for the degree of Doctor of Philosophy

30th September 2020

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Sustainable Knowledge in Hypertext

by Mark William Rawlence Anderson

Hypertextual in nature, the Web in its earliest form was technically limited and not capable of using the full richness of hypertext at that time. Despite subsequent advances in Web technology, some of the older hypertextual capabilities remain unrealised and hypertext/media appears to be treated more as a technology than a medium. For a hypertext docuverse that holds changing information, such as a knowledge base, paying heed to its hypertextual structure aids the long-term health and sustainability of the knowledge it contains. Wikipedia is the world largest public hypertext knowledge base. Constantly updated by humans and bots, it is an ever-changing knowledge store. Using Wikipedia as a context, this thesis investigates whether large collaborative hypertexts show signs of their contributors using deliberate hypertextual structure or are simply connecting 'pages' of digital content. The research also considers collaborative hypertexts in the context of social machines with regard to sustaining organisational knowledge as hypertext content. The results reveal under-use of processes available to sustain and improve an organisation's docuverse and a gap in organisational roles and skill-sets to apply those processes.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Mark William Rawlence Anderson declare that this thesis entitled *Sustainable Knowledge in Hypertext* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission;
- parts of this work have been published as: 'There and Here: Patterns of Content Transclusion in Wikipedia' in the Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT'17, 2017)[21]

Signed:.....

Date:.....

DEDICATION

To ELA, without whose love and support this thesis would never have reached completion

To MB and THN for their help and inspiration

Acknowledgements

This work was supported by the Web Science Centre for Doctoral Training at the University of Southampton, funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/L016117/1. But, there is more to a completed thesis than that dry statement: I have many to thank, and do so below.

Firstly, and rightly my thesis supervisors: Les Carr and Dave Millard. Thank you for your help, guidance and generous spirit. The path to a thesis is winding, tiring and full of the unexpected. Having supervisors who were open-handed, not judgmental of false starts, and who allowed me space to explore my subject area has been a true benison. My very sincere thanks.

My PhD cohort: Alan, Alex, Amy, Anna, Conrad, Emma, Faranak, Gert, Jess, Joanna, Johanna, Michael, Neal, Peter, Sami Sophie, Steve, Tim and our late and much missed colleague Laura. The long grind of MSc and PhD has been the easier for having such a large, friendly and supportive group travelling the same path. Your kind and good-humoured companionship has been a boon.

Time spent in the WAIS lab and the Web Science group was much enriched by many erstwhile colleagues there: Amal, Amber, Ashton, Briony, Bernard, Beltran, Callum, Charlie, Chris, EA, Ellena, Frode, Gefion, Henry, Iman, Ismail, Jacqui, Jo, Jonny, Manuel, Mark, Michał, Nick, Niel, Nikko, Richard, Rikki, Rob, Sofia, Stephen, Steve, Tom and others I have overlooked. Thank you all for your company.

To Frode Hegland, Chris Gutteridge and Paul Smart. A special thank-you for our interesting freewheeling lunches talking about hypertext, citations, data structures and knowledge systems; these have definitely informed this work.

To Alison Tebbutt (and previously, Lynn Oloro and Claire Wyatt) in the CTD office, and to Megan Chan in the WAIS office. Too easily we overlook all the background work that is so important to the completion of our studies. Thank you all for your help, ever rapid and supportive: less seen, perhaps, but certainly not unappreciated. My thanks too, to Karen, who largely unseen keeps the Lab so clean and whose ever cheery smile warms the start of a day's research.

To Gill Higgin. Thank you for listening ear, your assistance and advice, and for your willingness to step above and beyond administrative red tape to help with my work. I hope your own plans for study come to fruition.

To Nick Gibbins, Su White, Hugh Davis and Mark Weal. Thank you for your help and forbearance with my arcane hypertext questions.

To the team in University's Public Policy office, my thanks for the opportunity of a secondment to the Cabinet Office. My particular thanks to Yaryna Basystyuk for your help and support, making my short-notice secondment all the more productive and less stressful.

Further afield, beyond Southampton…

Very special and particular thanks go to Mark Bernstein. Firstly for writing Tinderbox, the tool that has enabled so much of my work and imbued me with an understanding of hypertext. More importantly, your help, support and friendship before and during my studies have been important in the completion of this thesis. Our regular correspondence mixing beta test reports, hypertext theory and culinary matters continues to be a joy.

To the Tinderbox and Storyspace communities. Your discussions and interest in hypertext has helped shape and sustain my own interest in the medium. I also thank those attending the dinner at Tinderbox Weekend 2013 in San Francisco—your encouragement to return to study and thus to give back to the community was the first step to the path to this thesis.

My thanks to Belinda Barnet, Rosemary Simpson, Cathy Marshall, Frank Shipman, Ben Shneiderman, Andrew Pam, Claus Atzenbeck, Stacey Mason and others in the hypertext community at large for fielding my many questions and helping track down old papers and software.

Particular thanks go to Ted Nelson. Thank you for taking time amidst all the calls on your attention for our correspondence, which was of much help to me. As importantly your inventions, insights, and long-time steadfastness in the face of opposition and uninterest have inspired me in my work and sustained it through the long trudge to its conclusion.

At the Cabinet Office, my thanks to Chris in the OIT for accepting my secondment and assigning me to the BI4BG rather than to general work. To the 'BeeGees' themselves: Glyn, Ian and Rob, thank you all for such a warm welcome into a small team and your openness to outside ideas. The opportunity to contribute to your work made by brief time with you all the more productive for my studies. You and the DKIM team—David, Galla, Dan, and Sarah—made the days in 1HGR fun as well as productive: all those hours on the train to and from London seemed less of a chore.

My thanks to Paul Fisher for stepping from my past to act as a reference when I applied to study; our time together on C-in-C Fleet's staff was the genesis of my interest in knowledge systems.

To my gallant proof readers: thank you to Clare, Richard, Gill and others who kindly stepped up to assist with proof reading as this thesis hurtled toward the final tape.

Lastly, and most importantly, my thanks to my wife, Elizabeth. The selflessness of your help, support and willingness to accept me doing this work that has demanded so much of my time and attention has been humbling. Thank you so very much for that gift, from the very bottom of my heart.

Footnote: I cannot help but think I will have left out someone or something deserving of mention above. If I do so it is not deliberate, but merely the act of a brain addled by much study. If I omitted you, my apologies. I thank you too.

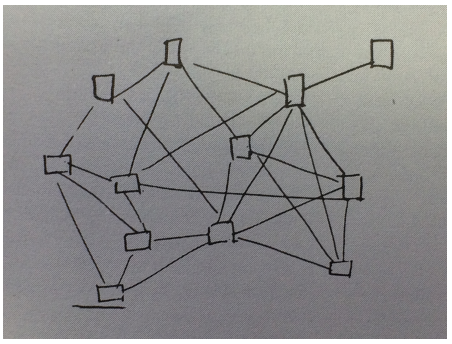

FIGURE 1: *Dream Machines, p31*
(reproduced with permission)

Hierarchical and sequential structures, especially popular since Gutenberg, are usually forced and artificial. Intertwingularity is not generally acknowledged—people keep pretending they can make things hierarchical when they can't.

EVERYTHING IS DEEPLY INTERTWINGLED

In an important sense there are no "subjects" at all; there is only knowledge, as the cross-connections among the myriad topics of this world simply cannot be divided up neatly.

Hypertext at last offers the possibility of representing and exploring it all without carving it up destructively.

Ted Nelson, *Dream Machines* [264, p.DM31]

The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries, or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.

The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.
[Section 6, paragraphs 1-2]

There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record.
[Section 8, para 2]

Vannever Bush, '*As We May Think*' [71]

Chapter 1

Introduction

Explaining the subject of a doctoral study to a lay audience can be challenging. In such conversations, the answer I receive most often is a variation of: '*Hypertext*? Isn't that the Web?'

For the generations who have grown up since the Word Wide Web's advent, it is easy to presume 'hypertext'[1] is just the a method by which we navigate the web, an affordance of that system. But to do so is to overlook the fact that technical limitations of the Web's early design meant some significant ideas of hypertext were left to one side and broadly forgotten by the Web—and the world at large. The Web is certainly a form of hypertext, but only *one* form of it. Hypertext is more than just the Web.

This lost understanding means that we also often miscast hypertext as a technology, when it is more usefully understood as a medium. This disconnect is further illustrated by the quoted examples below drawn from Web and social media posts. I have omitted the authors' names as the interest here lies in the words themselves rather than in the identities of their authors[2]. To dwell on the identity, or experience, of the authors of these quotes only distracts from the words themselves. Their general source is several queries running that scan for use of the word 'hypertext' on Twitter and on the Web.

This first item is a comment in response to an article on Medium.com which, unusually, is on the topic of transclusion. The article included this section:

> "Those old hypertext theory people had broader ambitions, though. They thought we might someday be able to pull live, updated pieces of other sites into our own websites, mixing and matching data or even whole apps as needed. This ability to include part of one web page into another was called "transclusion", *and it's remained a bit of a holy grail for decades*."
> (Medium.com, 26 March 2018: from 'The Missing Building Blocks of the Web' [97])

[1] Although the word 'Hypertext' might imply only text-based content, here I use it in the wider sense of 'Hypermedia' as per Landow [212, p.3]. The true interrelation of the two terms will be explored in Chapter 2.1.

[2] For provenance, the given references show the authors of the respective articles.

This drew a comment linked to highlight (italicised above) of the last sentence in the previous quote:

> "I don't see it. Every CDN[3], every image host, every advertisement, every iframe, every embedded font, every embedded video, it's all transclusion. In fact, Web 2.0 exploded with the idea of browser-based transclusion."
> (Medium.com, 26 March 2018: comment to a highlight in the article above [378])

The commenter seems unable to draw the distinction between a form of 'transclusion'[4] that reflects concepts such as software libraries and includes, and the form that is amongst the earliest ideas in hypertext: that of pointing to or *transcluding* (including) that content from its canonical source, thus both linking source and re-use and preserving the content unaltered. The technician's view reduces the hypertext to a method of simple code re-use. Indeed, they mention every aspect of a web page *except* the text. Instead the focus is on the styling of the content and serving ads in, on and around it.

Next, on Twitter, a tweet neatly defines the more general user's view of hypertext as experienced primarily via the Web:

> "Web browsers make use of a concept called hypertext, which was a thing developed in the 80s and at one time touted as the next big thing. While it made the Internet more accessible to more people, it's still just one protocol among many."
> ('Twitter.com', May 2018 [245])

Unsurprisingly, for the many for whom exposure to the word 'hypertext' means affordances that you click or touch to change content: hypertext feels like something under-the-hood such as a protocol. Again, it *enables something* rather than actually *existing as something*: protocol vs. medium.

The less common area of hypertext literature has a different problem. Few, if any of us, really understand how to write for—or read in—this 'new' but actually 50-year old medium. Twitter, once more:

> "There's a chicken-and-egg issue: no one understands what a hypertext book is, because there are so few of them; there are few because the tools are inadequate; there's no demand for better tools because no one wants to write hypertext books, because no one can read them…" ('Twitter.com', July 2019 [82])

[3] This a Content Delivery Network (CDN).

[4] The word is one of Ted Nelson's neologisms, which I will explain further in section 2.9.1

We may eschew paper for the convenience of digital 'paper' on screens, whilst still thinking in terms of paper. Yet, less obviously, much of our recorded knowledge is moving—indeed, has already moved—from physical paper to digital form. Many of our 'documents' may never be printed on paper and will increasingly be 'read' more by software than by humans. This change of canonical context offers both challenges and potential benefits. Information in the digital sphere need not be hampered by the enforced linearity of paper, digital or physical, and allows us more easily to start to use the benefits of a hypertextual world. In this vein, one more tweet, this time with a more positive view:

> 'Hypertext is an unavoidable format when reading digitally. We should be teaching students how to interact with hypertext and to use hypertext to increase their comprehension, not distract. What strategies can we teach students to support their digital reading success?' ('Twitter.com', February 2019 [222])

The first Twitter user quoted above might have replied to the last quote above that students already know how to click links, so why do they need training? Is not hypertext, or at least the Web, just a matter of pointing and clicking? Yet the last Twitter quote also raises the issue of whether we really understand how to use this new medium. This disconnect underlies my thesis. We created hypertext: we got the Web. In the scramble of the early days of the Web what useful hypertext features got left on the roadside?

Do we really understand the medium of hypertext? These quotes speak to the heart of an overlooked problem. Despite our use of hypertext in the context of the Web, many people (aside from the few who study hypertext) seem discomfited by the notion of hypertext as a medium. In my own field, hypertext appears to inhabit an interstice between Web Science's two core disciplines of computer science and sociology. The former tends to look at hypertext as being the output of technology and code, whilst the latter concentrates on the power structures within use of medium: both appear to spurn the perspective of the hypertextual medium itself. As our knowledge increasingly lives (only) in digital form it behoves us to understand the medium within which it resides, and understand how the linearity of print-age writing may not be the only, or best, way to record and sustain knowledge—especially that which is subject to ongoing change.

During my initial exploratory research I noticed that within my own multi-discipline of Web Science, the primary disciplines of Computer Science and Sociology have pulled me in different directions. With the computer scientists, there was much reference to Halasz's 'Seven Issues' [154] (which I revisit in section 2.4) a sense of hypertext being 'done'. Notably, there was a sense that the past is simply of less relevance because the *next* program will work better. For the sociologists there were political structures behind everything, which denies the reality of more basic human endeavours, such as knowledge of our everyday processes. For me, neither of these views addresses the central issue. Hypertext is over fifty years old and whilst we have in the Web a form of hypertext, we seem to still be using only part of what hypertext has to offer. Essentially, we are driving hypertext with its training wheels still attached.

We march boldly into a digital-first future with minds still shaped by ideas of the print era. This is not some zero-sum game between hypertext and the Web, if indeed such were possible given the Web's place in the (developed) world. Captivated by the wonderful new affordances of the Web, we have rather overlooked the transition from the print era to the digital era and with that inattention, we have neglected a medium that could help us make sense of this new era[5]. To distinguish between the Web and hypertext is not to denigrate the former, but to understand the broader possibilities of the latter. Without making this difference explicit, discussion of hypertext tends to circle back to a discussion of the Web.

Thus it is that in this thesis I wish to look at how we use hypertext as a medium to sustain our knowledge in the post-print age.

1.1 Background

Knowledge: the OED definition of the word is long and offers a many-faceted definition but in this thesis I will use it in the context of 'that which is known; the sum of what is known'[6]. The word turns our mind most easily to an encyclopædia or work of reference. As important, however, is the knowledge encapsulated in our cultural, civil and business organisations. The latter two are especially prone to rapid change, something difficult to capture in print technology. Metaphorically the handbook is out of date as soon as printed. Hypertext does, or should, free us from that constraint: free us not from the re-writing task but the mechanical process therein. In doing so, we should be able to become more efficient at sustaining our knowledge resources in a manner that reflects the present and not just the past.

The Web itself grows apace. Humanity is committing ever-increasing amounts of its knowledge, work and memories to digital form as Web-accessible data. Unlike pre-Web resources, this data is—or may be—richly interlinked. However, the extent to which the existing linkage is there by design is unclear; the same may be said of the degree to which its content is written down with fine-grained link accessibility and the hypertext medium in mind.

Though the Web is our largest hypertext, it is worthy of note that not all hypertext is part of the Web, or in the format of the Web. In many cases the Web's links are implicit (e.g. via query) rather than explicit within the coded data. But the distinction is moot in this context, as it is the nature and degree (or lack thereof) of linkage that is of interest. Hypertext can be implemented on other systems, and indeed was, prior to the advent of the Web as I will describe in section 2.3 (and in more detail in Appendix B).

Furthermore, the Web's Hypertext Markup Language (HTML) implements only some of Nelson's original vision of Hypertext and the ideas of Engelbart and Bush. Most notably, web links

[5] This is neatly encapsulated in visual terms by the 'Distracted Boyfriend' meme of 2017 [4].

[6] OED q.v. Knowledge, III.9.b [200]: III. 'The object of knowing; something known or made known': 9.b. As a mass noun. 'That which is known; the sum of what is known'.

are embedded in documents rather than being in a discrete link table and links have no semantic markers as to intent. A Web document describes only its outbound links and has no knowledge of any inbound links. The full and precise differences between the wider concept of hypertext and the Web's implementation of it is beyond the scope of this discourse but nonetheless does have impact on consideration of how knowledge is best maintained in Web-based hypertext.

Within the Web, Wikipedia is its largest, public, collaborative hypertext ranked at fifth[7] for popularity by Alexa.com [W91] [8] and the most popular content site that is not related to search, sales or social media. Wikipedia's largest sub-site, that for English language, has as at 12 August 2019[9] 5,208,948 discrete articles with 28,895,698 editor accounts—including 1,291 administrators—and is adding nearly 1,000 new articles a day.

The size and scale of Wikipedia alone shows how the Web has grown from a single page to a scale hard to imagine at the Web's outset in the 1990s. Elsewhere public and private organisations are accreting similarly large amounts of Web data. The Web not only connects 'big data'[10] but is itself big data.

Use of an interactive medium is not just consumption, but (co-)construction. In this context, hypertext does—or may—differ from the formalisms of 'print culture'[11]. Marshall McLuhan's essays *The Guttenberg Galaxy* (1962) [242] and *Understanding Media* (1964) [241] made the all too easily missed point that it is the *medium* that we should study, and understand, rather than its content[12].

McLuhan described as 'rear-view mirror' thinking the assumption that a new medium is merely an extension or amplification of an older one [289, p.98]. Writing from within the field of hypertext, Bolter's 1990 book *Writing Space* described the 'late print-age'[13]:

> "The term refers instead to the relationship of print technology to our literate culture. The printed book is no longer the only or necessarily the most important space in which we locate our texts and images. For all our communicative purposes, print is now measured over against digital technology, and the idea of perfect communication that our culture associated with print is under constant challenge."
>
> ([55, p.210])

[7] As at 12 August 2019.

[8] The separate list of 'W'-prefixed references to Wikipedia Uniform Resource Locators (URLs) is explained in section 1.6

[9] 'Wikipedia:Size of Wikipedia' [W675].

[10] Described by Wikipedia as"...a term for data sets so large or complex that traditional data processing applications are inadequate.", in article 'Big data' [W5].

[11] The latter term appears to have no canonical definition but is best described by Eisenstein, in her *The Printing Press as an Agent of Change* '[111].

[12] As in his most famous quote: 'The medium is the message' [241, p.9], i.e. pay heed to the medium rather than the message borne upon it.

[13] Here, the meaning of 'late' is as in 'towards the end' rather than '(almost) deceased' [55, p.210].

Thus Nelson's notion of 'intertwingularity'[14] is something to embrace rather than avoid. Richer linking aids serendipitous discovery, albeit with the complexity of more links to manage, though complexity of linkage is not bad in itself if authored with care. Hypertext (data) lives in a continually changing context. Even data about records of fact may require ongoing re-contextualisation to ensure they are accessible and understandable. This fettling[15] helps achieve an escape from—or at least an alternative to—the dependence on brute force search as a substitute for appropriate structuring of hypertext.

Indeed, at a time when Artificial Intelligence (AI) is gaining increased visibility in the wider community and Machine Learning (ML) is at a fledgling stage of truly understanding the subtleties of human language, well structured hypertext should provide a bridge to the technology like the semantic web. If we can annotate better the invisible thought processes that lie behind our knowledge, this can aid semantic understanding by machines. In turn this lessens the need to extract structure by brute-force probabilistic attack on data and may reduce the need for post-hoc clarification by human experts of algorithm-derived results. Such an approach to seeding semantic structure also offers a bridge across to areas of tacit knowledge[16].

When we consider how published knowledge is curated[17], we must acknowledge that hypertext is a change from the past behaviours. Whilst the physical nature of most curation tasks alters in practical action for digital assets, the same tasks generally remain. It is the last item, maintenance of context, aspect that is arguably new. Physical media may be re-indexed and re-racked, but there is no scope to dismember the books and swap the pages around. True, presenting old material via new media may refresh its context, but this goes to underscore the point that hypertext/hypermedia offer a new opportunity—and task—to the curator[18].

By contrast, refreshing or renewing content is possible, where pertinent, with hypertext. It is not that this must be done in all cases, rather that it may be done, in appropriate circumstances. This difference does not reflect ill upon traditional print-era curation activity. Instead, it adds to it and enriches the process. Given the nature of the hypertext medium and the fact it requires software to render human-usable output, maintenance of context is actually mixed in with all the other tasks, although it should still be recognized as a discrete activity to avoid being overlooked.

---

[14] q.v. epigraph at chapter head.

[15] The Oxford English Dictionary (online) defines *fettle* as 'To make ready, put in order, arrange.'. See the Oxford English Dictionary online, entry for the verb 'fettle' [3]. Originally relating to animal husbandry and subsequently to nineteenth century machines, in wider usage it has come to mean putting something in generally good order: this is most commonly echoed in the phrase 'in fine fettle'. 'Fettling' of a hypertext achieves the same as for an animal or machine: maintaining it in, or returning it to, good (working) order.

[16] Polanyi developed the concept in his 1966 book *The Tacit Dimension* [288]. Here I use it in the Knowledge Management meaning of Brown and Duguid [64], the 'know-how' of an organisation that is generally not written down and resides in its staff.

[17] Here I use the meaning of 'curator' (and words stemmed from it) as in the Oxford English Dictionary (online) [2]. Curator, n., II/3.:'A person who has charge; a manager, overseer, steward.'. This is as distinct from the more modern meaning, used in museums and galleries, where the curatorial role is about presentation of items more than the care of them.

[18] This drift in the meaning of the term is nicely considered in *Textual Curation* [197, pp.4–8]. I would agree that modern drift has obscured the older meaning of the word.

In collaborative contexts the share of edits is often seen to conform to the Power Law. However, Swartz's 2006 [337] analysis of Wikipedia editing shows a more complex picture where the main editors are mainly tidying and re-factoring content with individuals providing much of the content data (i.e. actual knowledge)[19]. Investigating the editing of links in Wikipedia, Boulain, Shadbolt and Gibbins [59] indicated actual content editing was only around one third of overall editing effort. This is not to imply that the other editing work is without merit, but does imply a significant and large overhead being needed to get usable content in a system supposedly as simple as a wiki[20].

Surprisingly, little effort has been devoted to building on that work, to investigate the unseen overhead of ongoing maintenance, in a curatorial sense, of large hypertext-based knowledge resources. Do we understand the true scope of the work, or the human skills required? In part the lack of progress here may be due to what Broussard's *Artificial Unintelligence* [63, pp.7–8] terms 'technochauvinism', that is 'the belief that tech is the solution'. Indeed, part of my argument here is that rather than naïvely presume the *next* iteration of technology will be better, we should learn to make use of that which we have. Another upbeat technical assumption encountered is linked to 'Linus's Law' that "given enough eyeballs all bugs are shallow" [296, p.30]. Originally a term relating to bug detection in open source software, the concept has spread to other technology-related areas including Wikipedia, in articles on 'How Reliable Is Wikipedia?' [W671], 'Encourage readers to edit the article' [W671], and in the Wikimedia lecture 'The Future of Libraries and Wikipedia' [W394, p.8]. At such distance from source that idea of the protection afforded by "many eyeballs" veers close to Broussard's point about hubristic faith in technology.

Missing is a more holistic[21] approach to the management of digital-era knowledge that cleaves neither solely to technical optimism nor to validating social theory. Much of the knowledge we need to run our public and private work may be humdrum in nature, yet it still needs to remain fit for purpose as the world around it changes.

1.2 Motivation

My first career was at sea, still (just) in the age of steam. No one had heard of the term 'socio-technical' but life afloat was very much that. As a signals officer in the Royal Navy (RN) I experienced the arrival of information networks and the step change in the speed of information flow. In turn, our tactics–our core business process documents—were a struggle to keep updated. I saw the difficulty of capturing changing knowledge in print-based systems and the problems of premature formalisation (linearisation) of complex subject matter. This sparked my enduring interest in the challenges of capturing and sustaining organisational knowledge.

[19] In fact content creation editing is likely even more nuanced than Swartz suggests, as noted by Darlby [95, pp.145–6]

[20] In fact, as my research will show, the technical and human editorial debt runs deeper than this.

[21] Here I use the meaning of 'holistic as in the Oxford English Dictionary (online) [176]. Holistic, adj., a. 'Of or pertaining to holism; characterized by the tendency to perceive or produce wholes'.

I left the RN just as the Web emerged, and by 1995 I was coding and commissioning hypertext for Web use. Since then, various employments[22] have all been closely linked to use of the Web and hypertext. From the early 1990s and continuously since, I have also taken an active contributory or admin/moderator role in online forums[23]. As an adjunct to that activity, I created a succession of Web-based FAQ and tutorial sites since the late 1990s. The longest running of these, started in 2004 is 'aTbRef' [20], a continuously-updated public guide to Eastgate's Tinderbox hypertext tool [110]. Those experiences have been invaluable in learning how to structure hypertext knowledge resources for anonymous use[24], and giving practical experience of the oft-unstated assumptions of both makers and users of hypertext.

My use of Tinderbox software from 2004 broadened my understanding of hypertext beyond its Web context and introduced me to lesser known forms like spatial hypertext and hypertext literature. Since then I have also been a lead tester of both Tinderbox and Storyspace [328] and given community support for the programs[25]. Engagement in Tinderbox-based consulting work has included integrating data with SIMILE Timeline [318][26], SIMILE combined with Google Maps[27] and experimentation with stretchtext in a Web context, including folded blog articles[28]. In 2010–2011 I also helped with documentation, testing and support of the erstwhile Tinderbox-based 'Emberlight' project [235].

A continuing and consistent observation from the above experience is that Web/hypertext resources need regular review and revitalisation to remain pertinent—unless they are purely archival in nature. Even in the latter case the linkage to and from the resource (or data for search affordances) may still require review. A further lesson is that to write for the reader as opposed to for oneself requires far more considered authorship: it may also involve presenting facts in more than one form to aid differing approaches by readers to the resource (context, vocabulary, etc.). Different sub-groups of readers may need different structural affordances to aid them navigate around a hypertext resource[29]. In turn, this may impose hitherto unseen requirements on the structure of the source hypertext and even to the manner of its writing.

---

[22] Co-founded a Web-based start-up: Brainworks plc 1995-2001. Digital Asset Management consultancy 2000-2013. Tinderbox-based [110] consultancy 2010-16.

[23] Initially on CompuServe, then Usenet and more recently web-based forums. It is an opportunity both to return the kindness of strangers back when I knew less and to learn from answering the questions of others.

[24] That is a practicality, not a virtuous act. Online resources are available at all hours and to people of differing languages, expertise and interests. Catering for anonymous access requires giving more thought as to how the reader will approach the work.

[25] On a voluntary unpaid basis. See the Tinderbox forum [109] and the (private forum) Tinderbox Backstage forum [108].

[26] 'Kandahar Timeline 1979-2010' [332], the original work underlying this, is a web timeline of events (now dark) in Kandahar undertaken for a journalist studying the genesis of the Taliban. This involved integrating exported Tinderbox data with SIMILE Timeline [318].

[27] The Kandahar project led to further integration of Tinderbox, Timeline and Google Maps: e.g. experiment 'The Itinerary of HM Queen Elizabeth II' [19].

[28] Ted Goranson's blog [145], written in Tinderbox, uses parenthetical content that may be both nested and transcluded to multiple insertion points within the hypertext.

[29] There are echoes here of Bush's concept of trails [71, section 7]

My personal experience, over some twenty-five years, of providing hypertextual knowledge resources on the Web and in other forms has shown me hypertext's usefulness both in documenting knowledge subject to change and in early capture of organisational tacit knowledge.

A good craftsperson keeps their tools fettled and in good order, the better to practise their craft. However, my observation from experience is there is no such craft for knowledge—at least not in the digital context. As the print era falls astern, we need to consider how to sustain our knowledge. An assumption that sustaining knowledge is a quotidian task for 'somebody else' is a harmful one.

General understanding of hypertext appears hampered by the fact that few people create deliberate hypertexts (as opposed to those who use web publishing to link together resources, e.g. blogs, etc.). I have long been surprised that we fail to make more of hypertext's affordances and my research here is an attempt to move use further towards a more informed and productive use of the hypertext medium.

Drawn as I am to that challenge, from a thesis perspective it became clear in early reconnaissance that this would be building upon sand. This is because there is no real consensus on effective use of the hypertext medium, as reflected by a dearth of published research or writing on the subject. As business follows sources of revenue and profit, so academic focus follows the areas funded by research money. Hypertext has long been outshone by its more captivating sibling, the Web—about which a lot *is* studied and written. This forces my thesis into a necessary but more difficult task: to indicate an absence of understanding and activity in an area (hypertext) that few find interesting and which many assume is a finished episode in the march of progress. I say 'indicate' rather than 'prove', as this is not a simple empirical measure.

1.3 Scope

To investigate the reality of the issues above, in Section 1.1, it is illustrative to observe them in the context of a Web hypertext. Non-Web hypertexts of scale do not exist in the public domain, and it seems appropriate to do this underpinning research on a resource against which others can easily revalidate.

One aspect of this investigation is whether—as a discrete role—the fettling of a hypertext is necessary or whether the wisdom of crowds will resolve emergent issues. This term makes distinction from the purely technical aspect of maintenance, as opposed to the wider task of maintaining the hypertext's coherence as a resource. Likewise, curation can be seen as simply preserving the (past) status quo and does not necessarily speak to the need to keep resources relevant or to re-arrange content.

In general discussion, there seems a strong faith (not backed by evidence) that large open systems will simply look after themselves. The implied evidence appears to be that systems with minimal

hypertextual assistance still persist (without insight as to why). Thus a starting point to investigating the longitudinal health of Web hypertexts is to look for signs of signs of maintenance in a public Web hypertext.

To do this, it was desirable to select a hypertext that met a number of criteria. Ideally, it should:

- Offer content that is publicly and anonymously accessible: access should not require membership or deliberate login (though the later might be needed for some features such as actual editing).
- Be edited by multiple, non co-located, authors. The latter aspect is important to increase the likelihood that collaboration occurs within the website.
- Contain (mainly) hypertext articles. This ruled out many of the larger public websites that are social media of some form (e.g. Facebook, Reddit). These mainly contain discussion threads consisting of short comments. Data changes little once created. Forum sites such as stackoverflow.com present the same issue in that they consist of very small pieces of content.
- Contain articles recording knowledge (all or in part). This discounted both large web stores such as Amazon and search portals, such as Google. Although they contain much content, sales copy changes little while site listings pages change constantly (even if the individual sites do not.
- Content should be in English or contain significant parts in English. This reflects the fact that evidence of transclusion may need to be investigated by reading the source code of pages. English sections can be used to identify patterns in languages not known to the researcher.
- Offer bulk access to site data. This obviates the necessity of using screen-scraping—assuming that is allowed—which can be complex and brittle if the public data lacks semantic structure.

Meeting all these criteria, Wikipedia was selected as the research target. As an encyclopædia, its content covers matters of absolute fact, such as the location and name of the local county town, as well as subjects where the knowledge is not yet unambiguously defined or agreed such as with Climate Science. Wikipedia is also collaboratively created and maintained, as anyone may edit it: most users log in to a Wikipedia account to edit the website, although it does still allow anonymous edits (the user's IP address is recorded for such edits). Wikipedia also makes its data available via public anonymous access, either via an Application Programming Interface (API) or as static dumps of data. As a wiki, the most complete datasets can include every edit as longitudinal data and thus current and past states can be rendered as a web page or viewed as a difference from preceding or succeeding revisions of the page.

1.4 Research Question

Contributors to a public hypertext such as Wikipedia do not necessarily record their maintenance activities, but finding deliberate use of hypertextual techniques would indicate evidence of editing activity with a mind to the hypertext's long-term use.

Having chosen Wikipedia as the locus of study, the question then arising is how to detect evidence of deliberate acts to help the longitudinal quality of the hypertext, such as to sustain the knowledge therein. This thesis thus poses the following question:

Does editing of Wikipedia show signs of sophisticated hypertext techniques?

It is most common for investigations to focus on the current state of a website, such as a wiki. Thus, how it looks and is constructed at the time of the observation. This does not, of itself, give any insight into the consistency and resilience of the hypertext as a whole over time. The anbove question is explored and further refined into a series of narrower scoped questions in the introductory section of Chapter 3.

1.5 Contribution

The research questions above have formed the content of my theses. Unglamorous as this task is, I feel strongly that it is needed to forward research into knowledge within hypertext. By describing the current state of our use of the hypertext medium, I offer a clean slate—a *tabula rasa*[30]—upon which others may start investigations into more productive use and understanding of the hypertextual medium within which the knowledge of many organisation now resides.

With such a base, I believe it will enable investigation of what neglected features should be used, how and by whom. Without it, any study will be hampered by those invested in saying other subjects are of more pressing interest. If Broussard's technochauvinism is not to rule, the digital era of knowledge needs a more outward-looking perspective, and to consider who speaks for a hypertext—as discrete from either the technical maintainers or the authors of particular subject matter therein.

1.6 Hypertextual Elements: Bibliographies & Glossaries

This thesis document is written in traditional long form as is required. However, the document also uses a lot of internal hyperlinks, albeit these are only available if reading the document in

[30] Latin: literally a 'scraped tablet'. Figuratively, as here, an absence of preconceived ideas ergo a clean foundation safe to build upon.

PDF form. The reader is strongly recommended to keep a digital copy to hand if reading the paper form of the thesis in order to have access to the extra semantic structure the digital copy will offer.

To accommodate the large number of Wikipedia references used for research, there are two bibliographies. These and other hypertextual affordances of this document are explained in Appendix O.2.2.

1.7 Summary

In this introductory chapter I have given a description of the problem area. Our societal move away from paper to digital media as a primary form of record creates new challenges in the way we sustain the knowledge stored therein.

This chapter continued by setting out thee background to the subject, my motivation in addressing it and the pertinent past expertise I bring to bear. Having defined a scope to the task I have then stated a research question which will be explored in subsequent chapters. A short description was also given as to some of the arrangement of content within this document.

In next Chapter 2 I will review literature covering the origins of hypertext, its important founders, and look at some of the early pre-Web systems. In addition, some features and processes germane my research are described.

I decompose my over-arching Research Question in Chapter 3, into the questions that formed the frame of this thesis' investigation. The chapter then explores my decisions relating to the choice and acquisition of the source data used for the subsequent research and analysis. It also explains my choice of what aspects of that content I chose to study in answering my research questions. Reflecting these choices I then discuss my selection of methodology and attendant ethical considerations are described.

In Chapter 4 I will report on my first piece of research into evidence of use of hypertextual techniques in Wikipedia content. The findings here led to the framing of the second piece of research, reported in Chapter 5. This, too, uses Wikipedia as the its data source but, in this case, the contributions by bots to the hypertext. This is in counterpoint to the efforts of human editors reported in previous Chapter.

My research is then synthesised and discussed in Chapter 6 and the outcome considered within the context of a social machine. I explore whether the latter is useful model for describing and informing use of hypertext in maintaining knowledge.

Chapter 7 will review the findings of the previous chapters and reflect on current challenges for sustaining hypertext. Lastly, Chapter 8 will summarise my findings, contribution and conclusions.

Chapter 2

Literature Review

> *'Ten years ago, I asked a hypertext conference, "Where are the hypertexts?" That's still a good question. I still don't know the answer. Since 1982, I've been associated with Eastgate, publishers of hypertexts and builders hypertext systems. We have lots more hypertexts today than we did in 1982 or 1989. But our virtual shelves are far too bare. What's taking so long?*'
>
> ACM Hypertext Conference 1999, keynote 'Where are the hypertexts?' [375]

In this Chapter, I will start by giving an overview of the hypertextual medium, before a review the literature describing the early history of hypertext, the advent of the World Wide Web, the invention of wikis and the concept of Social Machines. The starting point is a brief description of the concept of hypertext, before giving a pre-history of hypertext followed by some detail of hypertext's three main progenitors: Nelson, Bush, and Engelbart.

From that base, I will then review the period from 1960 to the 1990s looking at some of the notable hypertextual systems and their features. This will cover the period from Engelbart's NLSand Brown University's ground-breaking HES through to systems like Southampton's Microcosm and Graz's Hyper-G that emerged at the same time as the World Wide Web (further per-system detail is given in Appendix B.1). This period is reviewed in outline, with greater detail in Appendix B.1. The section finishes with notes on hypertext research from outside the Western Europe/North American context, primarily (then Soviet) Russia.

Having reviewed these early systems, I turn to Halasz's 'Seven Issues' for hypertext. These were raised at the very first ACM Hypertext conference in 1987, though Halasz actually revisited this topic again twice. Though the 'Seven Issues' have been much referenced and used for a yardstick of hypertextual advances, I have not found reference to the three iterations as a group and thus offer a mapping of the three and a review as to their completion.

I will then briefly cover two strands of lesser known hypertextual work: Electronic Literature (ELit) and narrative in hypertext, and Spatial Hypertext. The former is essentially the genre of

creative writing in hypertext. The latter covers a powerful but little-seen technique using 2D maps of hypertext nodes and inferring relationships from relative placement on the map. Neither is directly part of the research here, but the techniques in those areas are pertinent to the considerations of the research findings. These two topics are covered in detail in Appendix B, sections B.2 and B.3.

My review of early hypertext concludes with a reflection on the degree to which Bush's 'trails' have been implemented and the (missing?) tools to support use and maintenance of trails.

Moving to the early 1990s, the focus shifts to the nascent 'World Wide Web' and I will review hypertext developments pertinent to its birth and growth and the genesis of the wiki, the seed of the Wikipedia platform.

This leads to a reflection on Transclusion, now in the context of the Web and wikis, and how it has diverged from its inventor's vision with the advent of 'web hypertext'. This acts as background for the research in Chapter 4.

In support of the research in Chapter 5, I then look at the background of software 'bots'. This is followed by a section on the Semantic Web and Social Machines as background to Chapter 6.

Before moving to the literature it is useful to reflect on the above in the context of the concept of knowledge that may reside within a hypertext. The definition of knowledge I have referenced in the introduction admittedly is broad in scope. However, this thesis is not an investigation of the narrower contextual definitions of knowledge, but rather it simply reflects that our hypertexts hold organisational and process knowledge that is not captured in a single node: "that which is known; the sum of what is known" [200, III.9.b]. To delve more deeply into that definition is of nugatory value and ultimately un-revelatory as a narrower scope simply obfuscates understanding of the whole.

2.1 The Origin of Hypertext Terminology

The neologism and the concept of 'hypertext', as we now it know it today, is an invention of the latter half of the twentieth century. This is Theodor ('Ted') H. Nelson's canonical description, when he introduced the term 'hypertext' in 1965:

> "As "philosophy" I want to speak of two major things. First, complex file structures (like the ELF[1]) make possible the creation of complex and significant new mediam [sic], the hypertext and hyperfilm. Second, evolutionary file structures (like the ELF) make it possible to keep track of things that have been changing, without our awareness, all along. These include the major categories of human thought, which will go on changing.

[1] Evolutionary List File (ELF)

> ...
>
> "Let me introduce the word "hypertext" to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper. It may contain sunmmries [sic], or maps of its contents and their interrelations; it may contain annotations, additions and footnotes from scholars who have examined it.
>
> "Let me suggest that such an object and system, properly designed and administered, could have great potential for education, increasing the student's range of choices, his sense of freedom, his motivation, and his intellectual grasp. Such a system could grow indefinitely, gradually including more and more of the world's written knowledge. However, its internal file structure would have to be built to accept growth, change and complex informational arrangements."
>
> ('Complex Information Processing' [259, p.96])

Nelson actually defined a parallel word 'hypermedia' to cover a wider range of content than text (e.g. images, audio, video, etc.). Hypertext is the more widely used term and has come to cover both meanings in general usage. This likely reflects that at the time interaction with computers was via punch-card programs and teletype-style printer output. Though we tend to experience hypertext in a computerised visual context today, it is worth remembering that the early hypertext pioneers were working without these modern affordances.

In this thesis I follow Landow's approach [212, p.3] of treating the terms hypertext and hypermedia as interchangeable and will thus generally use the term hypertext generically for all forms of hypermedia.

Hypertext consists of two primary elements: pieces of media (most usually text) and the links between them. Over time terminology has varied and it is useful to cross reference the various terms. Firstly, terminology for the media.

Notes, nodes, lexia, writing spaces, documents, books, objects, and (on the Web) pages are all terms with essentially the same meaning in terms of a hypertext: a discrete part of hypertext corpus. Though some systems give addressability at a finer scale *within* such a part, it is nonetheless the essential (media) unit of a hypertext. The term 'nodes' came from graph theory where the graph is composed of nodes and linked by 'edges'. 'Lexia' is a nod to literary critic Barthes [32] as described by Landow [212, p.3]. Notes, writing spaces, documents and books all reflect the purpose of erstwhile systems; regardless of size, these would all have been discrete unit within their source hypertext systems.

Nelson himself generally refers to documents, which makes sense from the perspective of Xanadu. However, other hypertext systems have not adopted his Edit Decision List (EDL) concept. In the early days of smaller, non-networked system, a 'document' might also contextually mean a computer file containing a complete hypertext. 'Object' is a more abstract term and one which also covers non-textual media. There simply is no agreed canonical term for the discrete parts of media

within a hypertext. I will generally use the term note or lexia but use other terms as contextually appropriate.

The second element, links, is more simple. This are called links in almost all cases. Those using graph terminology may in some papers refer to the term 'edge' to describe a link. I will generally use the term 'link' to refer to linkages between items within a hypertext. Note too, the term 'trail' which generally is used to refer to a number of links that connect so as to form said trail. The terms 'path' and '(guided) tour' have also been used in a similar context.

Nelson coined another term I will use here, that of the 'docuverse' (Nelson's italics):

> "All of storage near and far must therefore become a united whole— what is now called a "distributed data base." Actual locations become essentially invisible to the user; or, in that traditional phrase, "You don't care where it's stored." The documents and their links unite into what is essentially a swirling complex of equi-accessible unity, a single great universal text and data grid, or, as we call it, the *docuverse*."
>
> (Literary Machines 87.1, p.2/53 [266])

In Nelson's Xanadu, all lexia are in a single docuverse, whereas the hypertext systems we have today are somewhat different. However, the concept of a single space for all content of a hypertext is still useful. At Web scale, we might see it as all web connected 'pages'. For a project like Wikipedia, we might regard it as all content on that domain, or the content of a single language wiki; or, in a work context, an organisation's document cloud.

2.2 Bush, Engelbart and Nelson

Nelson may have named this subject, but today's general concept of hypertext also draws upon the work of two other pioneers working in the twenty years preceding Nelson's paper: Vannevar Bush and Douglas ('Doug') C. Engelbart. These three hypertext pioneers we not of one type. Bush's work background was the US East Coast. He studied engineering at MIT and stayed in that field working for and with the government. He is rather an establishment figure by the time he joins our story. Engelbart was from a more humble US West Coast start. His education was interrupted by Word War II. After US Navy service, he returned to complete studies in engineering before turning to computers[2]. Nelson's background was more bohemian, the son of a film director and an actress. He had a liberal arts education at Swarthmore and after initial work in the North-East, he moved out to the West Coast.

Chronologically, Bush's work is the earliest, dating from July 1945 after World War II had ended in Europe. Engelbart read and was inspired by Bush's essay whilst still serving in the Pacific [280, pp.235–6]. Nelson did not come across Bush's work until after developing his own concept of

[2] This was before there was a concept of the subject of 'Computer Science'

inter-connected texts [272, p.170]. I think it is to hypertext's gain that Nelson and Engelbart brought quite differing pieces of the puzzle to hypertext. Engelbart brought a disciplined engineering eye and a focus on efficient synergy of man and machine. Nelson brought a wider, more prophetic view, informed by the humanities[3] and which encouraged others to think more expansively about the possibilities for hypertext.

Though these three men differed in background, experience, and outlook, their varied contributions combined to underlie, and give inspiration to, modern hypertext. The World Wide Web is essentially a large hypertext, albeit of limited capability in pure hypertext terms, and thus their thinking that lies at the base of today's primary medium of intellectual interaction today.

2.2.1 Hypertext's Prehistory

Before discussing the early pioneering work in hypertext, I should note some aspects of it can be seen echoed further in the past. None of these earlier examples were seen as a 'hypertext' in their time but nonetheless show some similarities in the use of non-linear organisation of media and high levels of inter-connection.

An early example often mentioned, is of the Talmud, a collection of writings that are a commentary and debate upon Hebrew Scriptures, and Jewish law and religious practice. Talmudic writing displays rich marginalia and cross references create a network of implicit links within the book. In 1659, Johann Amos Comenius created a dual-language encyclopædia in Latin and High Dutch called *Orbis Sensualium Pictus* ('visible world'). The two-column layout had the same text, in each language, set side by side offering an instant translation. Every section included images annotated against the text, allowing a visual translation and explanation[4] [86]. These examples are mentioned by DeRose and van Dam in their retrospective paper on FRESS [104].

Giacomo Leopardi's commonplace book of thoughts *Zibaldone di pensieri* was posthumously published in 1898. It too offered rich cross-referencing, with the scope for non-linear readings that foreshadow hypertext [330].

Punched cards were an early form of programming. Used in early modern computers, they also have a lesser-known hypertextual aspect. Punched cards were invented in 1804 for use in the Jacquard Loom. They were also used in Babbage's Difference Engine in the 1840s. In the 1880s, Dr Herman Hollerith introduced the use of punched cards for the US Census. His company eventually lead to the creation of Internation Business Machines (IBM), the 'Machine' part referring to the card tabulating machines rather than yet-to-be invented computers. A lesser known side-shoot of this technology was a number of pin-sorted card systems in the first decade of the twentieth century [80, p4–5].

[3] When he defined hypertext he was working as a sociology lecturer at Vassar.

[4] The work is perhaps better known because in 1705 it was translated to use Latin and English as the two languages.

Pin-sorted card systems, such as the McBee system [18] offered the pre-computer world a means for a non-linear investigation of information. An interesting cross-link to hypertext is that a McBee card[5] was used to store the National Endowment for the Humanities (NEH) grant [179] which gave Brown funding for van Dam's File Retrieval and Editing System (FRESS).

In the 1890s, Belgian lawyers Paul Otlet and Henri La Fontaine, who shared interests in bibliography and international relations, found an interest in the Dewey Decimal Classification system. Seeing it as limited, they resolved to expand this system to cover much broader range of classification[6]. The result was their Universal Decimal Classification (UDC) and in 1895 Otlet then founded the *Institut International de Bibliographie*[7] (IIB). The UDC was used to create their *Répetoire Bibliographique Universel* (RBU) that by the end of 1897 had over 1.5 million index cards [387, p.76]. Just after the First World War, Otlet secured funding to move the RBU to offices in Brussels to create the *Palais Mondial* which by 1921 was receiving 2,000 visitors a day [387, p.158]. It also accepted external requests. For 27 francs anyone in the world could make a request for information via telegraph or mail resulting in over 1,500 requests from all over the world [387, p.160].

In 1928 Paul Otlet and the French architect Le Corbusier published a pamphlet describing their idea of a 'Mundaneum' [387, Ch.8]. Thus Otlet's *Palais Mondial* was renamed the Mundaneum. Otlet gave greater detail of his cataloguing system in his 1934 book *Traité de Documentation: Le Livre sur le Livre* [284].

Between the World Wars, writers began to explore plot generation. In 1928, William Cook created *Plotto* [90], a book of narrative plots and variations thereon. A staggering 1,462 plots were offered to the aspiring writer—albeit linear by time of delivery. More interestingly, from a hypertextual viewpoint, some two years later Doris Webster and Mary Hopkins wrote *Consider the Consequences!* [371] and early form of Choose Your Own Adventure (CYOA) book featuring decision trees [304] to drive branching storylines. Whilst the first offer many choices of plot, the latter pulled them into the narrative itself rather than just in its planning.

Meanwhile, in Germany Emmanuel Goldberg was working on new Information Retrieval (IR) technology. His 'Statistical Machine' was intended for rapid locating of information stored in microfilm [69, pp.154–162]. The technology soon crossed to the United States where it was adopted by Hollerith's IBM and likely informed Vannevar Bush's thinking on IR mechanisms (q.v. section 2.2.2).

Author H. G. Wells is best known for his science fiction, but he was also an essayist and socialist and saw emergent information technologies assisting in spreading peace and prosperity. In 1937 he released a series of essays *World Brain*, noting:

[5] The article about the grant shows the record card along with sorting pins in a stack of such cards.
[6] Done with Dewey's assent [387, p.82].
[7] Known in English as the International Federation for Information and Documentation (FID)

> "The whole human memory can be, and probably in a short time will be, made accessible to every individual . . . It need not be concentrated in any one single place. It need not be vulnerable as a human head or a human heart is vulnerable. It can be reproduced exactly and fully, in Peru, China, Iceland, Central Africa, or wherever else . . . It can have at once, the concentration of a craniate animal and the diffused vitality of an amoeba." ([373, p.76 (Essay II)])

It is an interesting co-incidence that Otlet, Wells and Goldberg all appear to have attended the same 'World Congress of Universal Documentation' in Paris during 1937, although there is no sure record that the three met in person [387, p.210]. Otlet continued to struggle to find funding for his plans and his fledgling Mundaneum was ransacked by invading German troops in 1940. He died in 1944, unrecognised, only a year before Bush's famous essay (described in section 2.2.2). It was some decades before Otlet's place in the pre-story of hypertext was brought back to light. Luckily some of the Mundaneum survives and in 1998 it moved to a permanent display in Mons, Belgium. It is not clear if hypertext's pioneers knew of Otlet's work, but there many interesting cross-links from him to those supporting the written word, such as the library philanthropist Andrew Carnegie [387, p.118], and those envisioning it like W.E. DuBois[8] [387, pp.167–171] and Otto Neurath[9] [364, pp.96–103, 387, pp.167–171].

Was the Mundaneum a hypertext? Digitally, no, as the digital computer was yet to be invented. However Otlet's UDC gave a vision of interconnectedness that is at the heart of hypertext and it certainly offers an adumbration of a future such as the World Wide Web (I will return to this connection in section 2.7).

Overlapping the dawn of hypertext is another interesting project, Father Roberto Busa's '*Index Thomisticus*' [70] which involved computerised production of a printed concordance of St Thomas Aquinas' writing. With backing from IBM, the work took from 1949 until 1980. The cross-referencing nature of the original *Index Thomisticus* has led some Catholic observers to treat Fr. Busa as a hypertext pioneer, but that is to be over-generous. In this case the computer's role was more an efficiency aid to the massive process of producing the cross-reference work which was then used for a printed concordance[10].

The literary field offers up examples of pre-hypertext experimental non-linear writing. Jorge Luis Borges' 1941 volume of short stories *The Garden of Forking Paths* [57] included an eponymous tale where the narrative breaks general norms of temporal procession and narrative exposition. The apparently jumbled nature of its construction hinted at multiple reading paths through the

[8] Long hidden from view, a monograph of DuBois' infographics for the 1900 Paris Expo were recently published [361]

[9] Neurath created the Vienna Method of Pictorial Statistics [364, p.61] and founded the Isostat Institute [364, p.113]: he used bold, simple, stylised information display.

[10] A web-based online version [187] of the work was released in 2005, though it is the earlier work that gets mentioned in the context of hypertext. For example, 'The Priest Who Invented Hypertext' [363] claims somewhat speciously that 'Fr. Busa really did work on hypertext at least 15 years before Nelson'. This is also a good example of how the press, like much of the general public, generally have little understanding of what hypertext is.

whole. Here in the context of hypertext, the story's meaning is of less relevance than the manner of its exposition, when compared to contemporary norms of more linear narrative. Marc Saporta's 1962 *Composition No.1* [305] was a novel in a box. Just predating the first definition of hypertext, the 'book' was presented as a box with 150 unnumbered sheets of text, which truly could be read in any order. It is not suggested that these literary experiments had any direct impact to early (computer software) notions of hypertext, but as shall be seen in Section B.2 the combination of computers and hypertext opened new opportunities for narrative experimentation, and in media beyond printed text.

Differing aspects of modern hypertext are indeed seen echoed in past works. Looking more widely, researchers have found echoes of hypertext in areas as diverse as the texts of ancient European civilisations [213, pp.1–2] and the Wampum of the American First Tribes [169, pp.97–97]. That said, these are retrospective observations based on the modern notion of hypertext. I do not suggest that some of these latter, more obscure connections, influenced the birth of modern hypertext. Yet, they do expand our understanding that hypertext did not spring from the void.

Past connections notwithstanding, the advent of the digital computer in the late twentieth century shifted significantly our potential to exploit interconnected and non-linear text, as well as other new forms of media emerging from the shadow of paper and print. The time had come for hypertext to step forward as a concept in its own right.

2.2.2 Vannevar Bush

Bush's primary contribution, his essay 'As We May Think' [71] essentially predates Nelson's definition of hypertext by 20 years but he put some important ideas into play and inspired others towards what coalesced into hypertext. Writing in 1945 as World War II was drawing to a close Vannevar Bush's essay 'As We May Think' [71], written for *The Atlantic Monthly* magazine, reflected his experience of managing the ever-increasing flow of information becoming available to the individual.

In the inter-war years Bush was involved in U.S. Government work on creating analogue computers used for the likes of calculating ballistic trajectories for artillery use and on information retrieval processes. The latter work led to him applying for a patent for his 'Rapid Selector' machine in the late 1930s, since seen by some as the first electronic IR machine. However, the patent was turned down, with the aforementioned Emmanuel Goldberg's work being cited as prior art [387, p.209]. Despite this, generally his commercial work had more success and the company he founded went on to become today's Raytheon [297].

Bush's inter-war scientific and engineering work also led to a series of posts supervising a range of U.S. Government research activities. By 1938, he headed the National Advisory Committee for Aeronautics (NACA) before going on to chair the U.S. National Defense Research Committee (NDRC). In 1941 he then became director of the U.S. Office of Scientific Research and Development (OSRD) and during World War II directed the efforts of thousands of scientists working for

the country's national war effort. He was still at the OSRD when he wrote 'As We May Think'. In the essay, he reflected on the human pattern of associative thinking:

> "The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries, or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.'
>
> "The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature." ([71, Section 6, paragraphs 1–2])

Arising from such patterns of thought, Bush raises the notion of individual 'trails'. He also considered how these might be shared, proposing the 'Memex' device. Pertinent to note is that Bush was already exploring some of IR aspects used in the essay, as early as 1932–33 [178, p.57][279, pp.42–43]. Bush's proposal of a 'Memex' device thus builds upon that earlier IR work. The Memex, perhaps the most talked about aspect of his essay was thus more than the distillation of his wartime experience. He states:

> "Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, "memex" [sic] will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory."
> ([71, Section 6, paragraph 4])

Despite the generally favourable view from today's perspective of hindsight, there was some objection at the time. Psychotherapist John H. Weakland wrote to Bush, in private correspondence:

> "(1) Wouldn't the fact that association patterns are thoroly [sic] individual make a general use of the memex difficult? (2) How would the tremendous bulk of information already recorded be made usable, especially for a searcher who wants to

> branch into lines of thought and knowledge that are new and quite unfamiliar to him? (Weakland, 27 August 1945)" ([279, p.60])

Weakland's fist question suggests an individual's train of thought, as recorded in their trails, might be (too) difficult of others to follow and thus gainfully reuse. The point about "individuality" likely also reflects the fact that what is 'intuitive' is not always common to every observer. Clearly, we have little problem following today's web links, but we are also educated to not worry the fact that we do not always know exactly what lies at the end of the link. The *reason* for such links may still be less clear, so Weakland's point is not without merit. However, as Bush's trails have never actually been implemented in Memex form, Weakland's question remains unanswered.

Weakland's second question shows a reasonable concern for its time, as issues of scale were more immediately challenging than today. Though the mimeograph stencil had been invented in the 1880s and Chester Carlson invented the photocopier in 1938, Weakland was nonetheless writing before the availability of large scape digital copying and before the arrival of technology like Optical Character Recognition (OCR). Still, the issue of scale is arguably still being resolved. We can replicate and store at scale more easily, but finding items can still be hard. Personal network or intra-organisational search is often a pale shadow of the power of Web-based search engines but even the latter can return excessive or incorrect data, depending on the user's best-guess search query.

It is also interesting that despite Bush's background in large-scale organised work that his idea of the Memex, with its storage and retrieval of 'trails' is not hierarchical. Indeed, the concept of trails led to a further observation:

> "There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record." ([71, Section 8, paragraph 2])

Bush could see value in the exploration, revelation and exchange of discrete perspectives of our corpus of knowledge—a point to which I will return in Section 2.6.

A point often overlooked is that Bush's ideas for the essay were broadly formed *before* World War II. As early as 1939 he had notes that contained all the contents of the essay he eventually had published over five years later bar the introduction and conclusion [279, p.52].

Bush's work was not entirely without criticism. Bawden [34, p.2] observes that some revisionist authors have suggested Bush failed to give proper recognition to the inventors of some of the technology he adopted. For instance, he cites Buckland's [69] biography of Emmanuel Goldberg's IR and his 'Statistical Machine'. That seems to be harsh as there is no clear evidence Bush was directly influenced by Goldberg, though likely he would have come across Goldberg's work. Indeed, closer reading of Buckland [69, pp.219–225] does not really reflect the tenor of Bawden's

charge. Buckland writes that it is not clearly the case Bush's Microfilm Rapid Selector (and by inference the Memex that builds upon it) was a case of plagiarism [69, pp.222–3] but he does go on to highlight Bush's neglect of referencing the works of others that had fed into his own work [69, pp.224–5]. Despite some carping, it is nonetheless the case that Bush's writing give his readers few clues as to previous thinking that might have helped shape his view.

There is also some indication that the essay was in part teasing what Bush felt to be backward-looking librarians [178, p.58]. His early IR work was driven in part by his dissatisfaction with existing practice [178, p.57]. Given that the essay had been in draft for some years, since before World War II, his time with NACA, NDRC, and thee OSRD would have given Bush ample opportunity to see weakness in contemporary information management.

Bush revisited the Memex concept in 1959 'Memex II' [72, pp.165–184] and again in 1965 'Memex Revisited[11] [73, pp.196–216], updating some technical aspects to reflect advances in technology. Despite the passage of time and the ennui Bush himself may have come to have felt for his original essay, it does not diminish from the inspiration that the original essay gave to many who read it. In '*From Memex to Hypertext*' [280] Nyce and Kahn draw together a range of essays that explore further both Bush's latter work in the subject and contributions from other early hypertext pioneers including both Engelbart and Nelson.

2.2.3 Engelbart

Douglas ('Doug') Engelbart had indeed read Bush's essay whilst still serving in the United States Navy (USN) as world War II drew to a close [30, p.44 and footnote 7], though it served more as general inspiration [280, p.236] than a blueprint for his own future work (as sometimes implied on general summaries of this subject). Engelbart was more influenced by the work of the American linguistic theorist Benjamin Lee Worf [29, pp.36–38][12]. This lead him into looking at language from a symbolic perspective and consideration of how symbolic logic might best be used in allowing effective interchange of information between man and computer.

Recently married, in 1950 Engelbart had an epiphany, choosing to dedicate himself to find a way to maximise his contribution to mankind. Seeing that the new field of computing could dramatically increase his contribution, he returned to study a computing PhD at Stanford. Feeling his interest in language and symbolic structures made him a poor fit for academia and that industry offered a faster route to making a contribution, he took work in the US West coast's tech industry which at the time was driving innovation due to massive government seed funding and grants[13]. He joined Standford Research Institute (SRI) in 1957 and started on what would become his life's work.

[11] Not published until 1967

[12] Another interesting cross-connection is that Nelson was also aware, separately, of Worf's work [30, p.38, footnote 1].

[13] For instance, the US Office of Naval Research (ONS) alone was funding 40% of all basic US research[29, p.6]).

The concept of what became his life's work is laid out in his 1962 paper 'Augmenting Human Intellect: A Conceptual Framework' [112]. He sought to enhance the quality of interaction between people and computers and did not shy from a challenging approach where he and his growing team would be part of the experiment by forcing themselves to use the tools they were developing, rather than simply to observe others doing so. Compared to Nelson's flights of fancy, Engelbart's work may seem dry and yet it produced some impressive results.

Engelbart's commitment to his whole team using the tools they were creating was challenging, as the task of making and understanding the tools they were creating precluded simply leveraging the work of others. New tools needed to be imagined, built and then understood but that experience and the feedback loop it created fostered a team well informed in terms of understanding the interaction of person, tool and task—a point all too easily overlooked today when other people test new code. This is the so-called 'bootstrapping' approach [29, p.24], not mentioned explicitly in the 1962 paper (above) but implicit in it nonetheless[14].

His concentration on practicalities involved experiments in de-augmenting capabilities [115, pp.197–201], including tasks like trying to write using a pencil and then doing so with the pencil taped to a house-brick (illustrated at [115, p.199, fig.8]). He wrote:

> "My workstation-design philosophy was to fix it so what you are looking at is positioned for best viewing, and the devices you use to control the computer should be located where it is best for you to operate them. Don't get caught in the anachronism that because we got used to paper and pencil and that technology, we ought to be able to have our controls right on the surface of the thing that we're working on. It may end up that that way is best, but don't make an a priori assumption. So we didn't.
>
> "One way of explaining to somebody why it could make a significant difference if you can do things faster, is to provide a counter example. So, I had them write with a brick taped to their pencil (Fig. 8), because it's only a matter of happenstance that the scale of our body and our tools and such lets us write as fast as we can. What if it were slow and tedious to write? A person doesn't have to work that way very long before starting to realize that our academic work, our books-a great deal would change in our world if that's how hard it had been to write."
>
> ([115, pp.198–200])

Such experiments led to innovative results such as the mouse and the one-handed chording keypad[15]. Both the subject of much erstwhile interest, only one of these went on to have success—the mouse. The chording keypad, required training and proficiency, whereas the interaction was

[14] The description is derived from a computing term of art—the 'bootstrap technique'— which refers to a small initial loading program: it traces back to at least 1953 [68, p.1273].

[15] Coincidentally and separately, Nelson had been thinking of a similar one-handed input device, though this was not apparent until the two first met at SRI in 1966 [272, p.166]

a mouse was simple to acquire. Even so the mouse had to await the arrival of cheap PCs in the 80s before it became a common object, the invention having predated its market[16]. All these innovations fed into what became the oNLine System (NLS) system which started c.1963 and was enhanced and refined, based on the bootstrapping approach of the SRI team.

At the 1968 Fall Joint Computer Conference (FJCC) in San Francisco, Engelbart's SRI team gave a demonstration of their tools and software, NLS forming the core of activities. In a tour de force that has become known as the 'Mother of All Demos'[17] they demonstrated windowed working, the mouse, linked text (hypertext), word processing, outlining, collaborative editing and review, re-configurable views, detailed addressing, video conferencing and more. For most in the audience this was all new and many of the techniques were years ahead of their first widespread use. It was also NLS's first public outing. Here Engelbart describes NLS's shared Journal files and video conferencing:

> "Citation practices—If in reading an item online the user comes across a citation link to a passage in another Journal item he can follow it directly to specified passage in the cited item. In writing a Journal item there is a great feeling of freedom about citing other items, given these expectations of easy accessibility. This freedom fosters the use of very short dialogue contributions; when several succinct citation links directed at specific passages can be easily used for nouns in a sentence, one can communicate important contributions very quickly."
>
> (NLS teleconferencing features [113, p.175])

This describes shared files for collaborative working at a time when few computers were networked and fewer used shared files between users. A notable point about NLS is it used an underlying outline structure, which is at odds with most later hypertext systems.

It is worthy of note that at a time when few computers were networked, NLS was a time-sharing multi-user system. The SRI, the parent institution for Engelbart's Augment lab, was also one of the first four nodes or Network Information Centers (NICs) on the **A**dvanced **R**esearch Projects **P**gency **Net**work (ARPANet).

Despite all this innovation, Engelbart was on the wrong side of history. The early dominance of time-sharing terminals was giving way to network PCs and with that software was moving away from structures like NLS towards the smaller programs familiar to computer users today. Government funding was withdrawn in 1975. A side effect was that labs like PARC also began to poach Engelbart's staff and by 1977 little was left of the original SRI team.

[16] Perhaps forgotten is that the mouse's original task was simply to position the cursor into a desired box on the screen. It was not until 1974 that the (Xerox) Palo Alto Research Center (PARC)'s 'Gypsy' programme used the mouse for point-and-click commands [174, p.201].

[17] '...a calming voice from Mission Control as the truly final frontier whizzed before their eyes. It was the mother of all demos.' [216, p.42], popularised as Andy van Dam quoted it in a keynote to the 1987 ACM Hypertext conference.

The austerity of Engelbart's methodology does not make his work as approachable as some of his contemporaries. His papers 'Toward High-Performance Knowledge Workers' [116] and 'The Augmented Knowledge Workshop' [114] give a flavour of this serious approach—which belied his demeanour in person. This is unfortunate as the powerful core of his concept is easily overlooked. All too often we will wait for the polished final product that may never arrive. The H-LAM/T or Augment process forced the SRI team to live through the inadequacies of their fledgling tools whilst they worked to improve them. This is a model of core engagement that can be seen in some organisations today (across a range of fields) and which resonates strongly with my early training at sea.

From their initial meeting in 1966 and despite their very different approaches, Nelson and Engelbart continued to communicate and became good friends. Nelson, as we will now see shared the same difficulty in getting others to adopt his grand vision.

2.2.4 Nelson

Theodor ('Ted') Holm Nelson is the third of the hypertext pioneers—though definitely not third in order or importance. Inventive from an early age, Nelson's involvement with computers and ultimately hypertext begins with him at Harvard, reading 'Social Relations' (Sociology) and taking a computer course at Harvard at the beginning of the 1960s. Nelson was amongst the small group of people who did not ascribe to the then prevailing notion that computers were only suitable for subjects like mathematics and engineering. Moving on to teach sociology at Vassar he coined a name for his concept of non linear text.

It first appears as a flyer for a talk given at Vassar (reproduced in *Possiplex* [272, p.149]) in January 1965. During that year he produced two papers on the subject. Best known is 'Complex Information Processing: A File Structure for the Complex, the Changing and the Indeterminate' [259], already quoted above in section 2.1. In fact, the first published use of the term hypertext was earlier that year, in a lesser known short paper 'The Hypertext' [260] given to the 1965 Conference of the International Documentation Federation:

> "This paper proposes a new medium, the hypertext. This is a generic term for texts (and combinations of texts with other materials) which, because of their structure, require automatic handling and display devices. The hypertext will typically be non-linear, branching, and large, with various options to the user."
>
> (as reproduced in [272, p.154])

1966 saw the first 'Xanadu Manifesto' [261] (reproduced here, in part, at Appendix A), laying out Nelson's putative Xanadu hypertext system. Later in the same the year he would first meet Doug Engelbart and get to see the latter's NLS [272, p.166]. It is unclear to what extent, if any, the latter had on Nelson's collaboration at Brown on Hypertext Editing System (HES) (described below) though his recollection to Barnet [30, p.98] would suggest it was not discussed.

In early 1967, between January and April, Nelson produced a series of eleven short 'Hypertext Notes' [265, pp.15–31]. They are interesting as they show the hypertext concept being fleshed out and proposal for new terms of art for the medium. In 'Hypertext Note 1' (23 Jan 67), Nelson refers to other proto-hypertext systems:

> "Hypertexts are inevitable, and will come into being just as fast as text handling systems come into the field. Several systems suited to handling some hypertexts exist already; Magnuski's DOC system at Project MAC [290], Bernstein and Slojkowski's program documentation system[18] at Bell Labs[38], and Engelbart's [NLS] editorial console at Stanford Research Institute." ([265, p.17])

Dated 6–10 March 1968, Nelson's 'Hypertext Implementation Notes' [262][19] begin to add actual process detail for a working hypertext system. They were likely written for HES or with HES in mind[20].

An innovation included in the notes is 'Stretchtext': the basic concept is that particular passages may 'stretch' to reveal a more detailed exposition. Nelson envisaged the hypertext user drilling down and rising back up to the top level text as required. Many sources incorrectly cite *Dream Machines* [264, p.M134] as the initial source but a much clearer vision is given in these notes (pp.14–18) [262, pp.14–18]. The very first reference, though much thinner in detail, is 'Hypertext Implementation Note #8' of 29 April 1967 [265, p.28].

Stretchtext is elegant in concept but not so easy to implement. Simple in theory, it is linguistical challenging, as I can attest from implementing stretchtext in an online blog [145] in 2011. The issue is not technology but the flow of text and the user experience but Nelson's early notes offer a vision of what it could be.

Nelson's intent with hypertext was far from theoretical. In 'No More Teachers' Dirty Looks' (1970)[21] [263], he criticises the fledgling use of 'Computer-Assisted Instruction (CAI)' for its lack of vision and proposes alternatives, including use of hypertext. In 1974 he self-published his book *Computer Lib/Dream Machines* [264][22], a book which was to inspire many in the forthcoming personal computer revolution.

---

[18] Actually called Program Management System (PMS).

[19] With Nelson's permission I have recently added digital transcriptions of these notes to the same archive [262].

[20] Nelson himself is unclear on this (author's personal correspondence with Nelson). The last page of the document ends: "(Note that we discussed these matters in a very confused way last June. This was the purpose of the various acknowledgement backpointers and "Wilco bit".)", which might indicate Nelson and van Dam were discussing Nelson's hypertext concepts in mid-1967. van Dam and Nelson met up and started discussing what led to HES after the 1967 Spring Joint Computer Conference (SJCC) (18-20 April). 'last June' in Nelson's notes would have been June 1967, which fits the narrative.

[21] This is also reproduced at pp.DM130–135 in the 1982 reprint of his book *Computer Lib/Dream Machines*.

[22] 1982 reprint edition by Tempus/Microsoft Books (1982), updated the original 1974 version. It mainly changed some of the early technical details of computers which were now already outdated. The book also took the interesting form of being two discrete two texts, one reversed and bound to the other—turn *Computer Lib* over and by rotating the book it lets you read *Dream Machines*.

A second book by Nelson, *Literary Machines*[23] [266] laid out his vision for his Xanadu system. Chapter 2 is notable for the running headers and footers, giving the chapter 3 parallel text streams. On page 2/53 he describes the docuverse: conceptually, this becomes the single repository of the Xanadu system within which authors can cross-connect texts or transclude as allowed by the original author. Chapter 2.9 presciently addressed issues of attribution, pluralism and copyright that have yet to be resolved in today's public web. Nelson foresaw the need for robust system of micro-payments to allow reproduction whilst acknowledging copyright. I well recall the difficulty of implementing e-commerce in the mid-'90s and even today *micro*-payments are not realistically possible due to current transaction-processing cost models.

Another hypertext system Nelson has proposed is ZigZag [357], which involves sequences of nodes intersecting in a multidimensional space (see video [340], and demo [10]). It moves away from most other systems, in terms of structure, but (unintentionally) foreshadows the later semantic web concept of data triples. It was later explored further by Goulding, Brailsford and Ashman in 'Hyperorders and Transclusion: Understanding Dimensional Hypertext' [147].

In a recent video 'SHOWING CONNECTIONS—the general problem' [317], whilst repeating his argument for visible connections (i.e. links) in hypertext, he rightly widens the scope to timelines. Timelines are a powerful means to show parallel strands of events. Though outside the purview of this thesis, timelines are another neglected area of knowledge retention and visualisation.

In 'As We Will Think' (1972) [267] Nelson compares and contrasts Bush's Memex with Xanadu, the section 'Hypertext' [267, pp.255-255] is particularly pertinent. Nelson teases out the different ways links may work and thus how trials might be used. The article as whole is an interesting reflection on Bush's 'As We May Think' some twenty-seven years after publication.

Nelson has struggled to get backing for his Xanadu system, in part perhaps reflecting his desire for design control after his experience with HES at Brown (q.v.). For a period in the late 1980s Nelson worked with a team at Autodesk on a new version of Xanadu but the collaboration ended and 1992. Since then Nelson has continued to work on his own to bring his ideas to fruition.

His 1995 paper 'The Heart of Connection: Hypermedia Unified by Transclusion' [269] is the clearest published exposition of his transclusion concept[24] and he illustrates his idea of "Deep Intercomparison Viewing—Transclusion and Link Display through Transpointing Frames".

Nelson returns to the Xanadu concept in 1999 'Xanalogical Structure, Needed Now More Than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-use' [270]. The paper includes illustrations of two digital, limited, implementations of Xanadu. Of note is direct reference to EDLs which form part of the current Xanadu demonstrator.

[23] First self-published in 1980 and a further nine times between 1981 and 1993 (according to the 199 frontispiece.)

[24] The term was first mentioned in a later edition of *Literary Machines* (version remains unclear but likely Ed. 93-1), but as the book is no longer in print few will have seen it.

Nelson has not given up on Xanadu. A 'Xanaviewer' [388] demonstrator is available online with some sparse documentation [291]. Whilst Xanadu itself has yet to become a working system, its concepts remain inspirational. The viewer is shown in Figure 2.1, a diagram I created as part of the presentation of my own paper on transclusion[25]. The figure actually shows the original definition of hypertext being transcluded, using Xanadu[26].

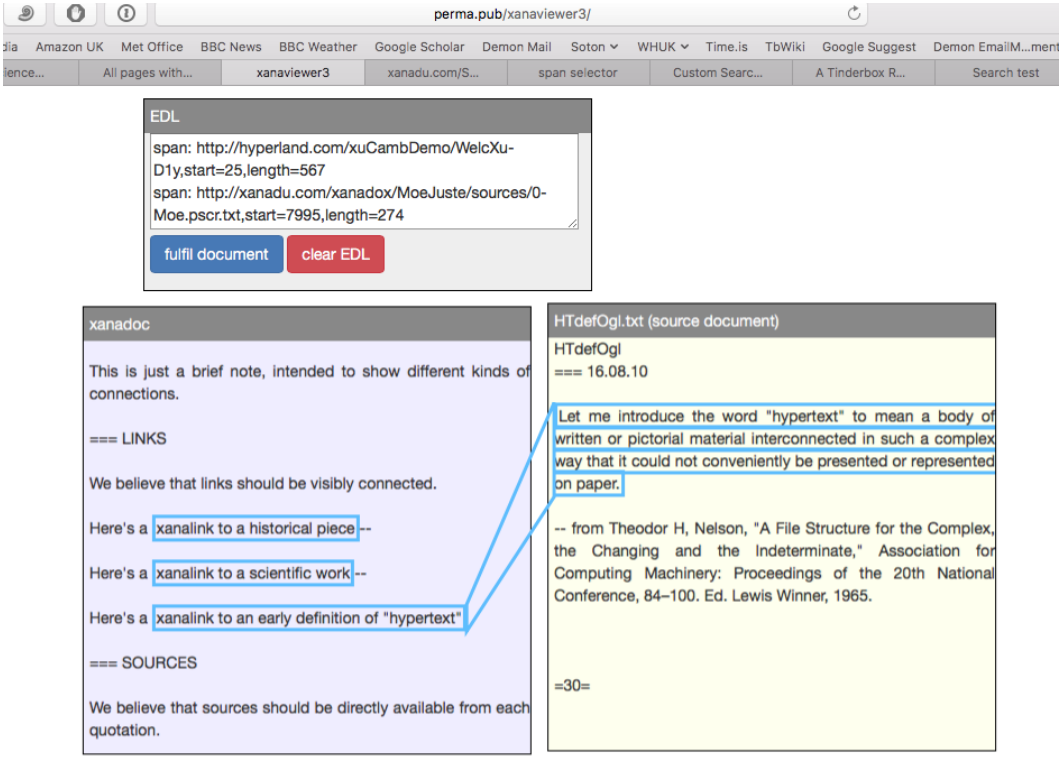

FIGURE 2.1: Nelson's Xanadu 'xananviewer3' displaying transcluded text.

Transclusion is explored further in Section 2.9 and Appendix C.

2.2.5 Hypertext Pioneers Redux

Three very different personalities underpin the birth of modern hypertext. All are now best remembered for their ideas rather than their hypertext systems, if this is not unfair to Nelson who is the only one of the three still alive today. Though Engelbart did get NLS operational, neither the Memex or Xanadu were ever properly released as commercial products. Indeed, NLS (latterly renamed 'Augment') was not widely adopted and Engelbart is best known for the computer mouse and the 'Mother of All Demos'.

[25] The paper was based partly on research presented in Chapter 4 [21] at ACM Hypertext 2017

[26] The process is hard to use, as once a document is declared to the system, it cannot be edited. Rather, a new version of the document must be written and uploaded

It is too easily forgotten that all three systems and their attendant concepts were far ahead of contemporary thinking and a likely market. Engelbart's mouse had to wait until the arrival of affordable PCs in the late 80s to become the iconic peripheral device it now is, and it is already becoming less visible as touchpads and haptic interface screens have rendered its physical form, if not its task, obsolescent.

NLS suffered from two impediments: the skill required to operate the system and the fact it used the time-sharing approached to shared computing. Both these were a poor fit for the newly burgeoning personal computer market and its wider audience. Although Bush revisited the Memex concept twice, interest waned in the machine itself. The idea still excites some and whilst not trying to build the original memex—which would now be very dated—Gemmell *et al.* showed an interesting conceptualisation in their 'MyLifeBits' [139]. A side issue of note is their reflection on data size and the variation between expectation and reality.

Finally, Xanadu remains in limbo and yet to be formally implemented and launched. Still, it would be wrong to see that as a record of failure as all three systems, even if only in conceptual terms, have inspired later work in the hypertext medium.

Drawing on the work of Bush, Engelbart and Nelson, and the increasing degree of access to computers, there was a burgeoning of hypertext systems for around twenty-five years from the late 1960s.

2.3 Pre-Web Hypertext Systems

In the post-Web world is is easy to miss the fact that there was a diverse and active hypertext research community. A variety of approaches were being trialled, noting that this was in an era when networking of computers was an exception, other than large infrastructure projects like ARPANet. An extended discussion of pre-Web hypertext systems is at Appendix B.1, covering 28 discrete systems. Here, I wish just to note some of the innovations and features that these systems brought to the field of hypertext.

- NLS (early 1967) was the first working hypertext system, albeit a lab-based research system.
- HES (late 1967) is held to be the first system to run on a commercially available computer, as opposed to specially built hardware. It is also notable for the first use of a 'back' button allowing the user to retrace their steps, for instance if following a wrong link.
- FRESS (1968), which replaced HES is notable for the first use of two-way links and for an 'undo' button for use in text editing.
- ZOG (1972) was deployed at sea in the mid-1980s, which was notable given the technology of the time. It was later commercialised as KMS.

- The Aspen Movie Map (1978) was the first large hypermedia project, making use of then-new laser-disks.
- Guide (1982) was the first commercially sold hypertext program for the general market. Prior to this NLS (above) was available as a time-sharing service, and KMS was offered as consulting service. By comparison, Guide was a program the computer could run on their own personal computers.
- Hyperties (1983) was the first program to use anchor text for links (rather than the previous number-selection method). Hyperties was also a first in being used for a book '*Hypertext Hands-on*' [316] that also included software with a hypertext version of text (and the Hyperties) program.
- Intermedia (1984) is noted for its introduction of a suite of applications which allowed linkage between their content.
- NoteCards (1984) introduced a history list, aiding the user to see their trail of visited items within the current session.
- Document Examiner (1985) introduced the concept of a bookmark, now a staple of web browser use.
- Storyspace (1987) was the first hypertext system designed specifically to assist creative (hypertext) writing.
- Hypergate (1987) pioneered the use of 'breadcrumbs' to assist with navigation.
- Microcosm (1988) brought support for multiple link bases.
- Hyper-G (1988) introduced shared link bases.
- Aquanet (1991) drawing on Notecards, this is arguably the deliberate spatial hypertext system.

Though many of the features mentioned above are in use now, notable omissions are two-way links, and the multiple link bases seen in Microcosm. More than that, the concept of a a hypertext as a non-linear multi-document form has been lost somewhat from the Web experience. A more familiar perspective, post-Web is that a hypertext in an emergent property arising from linking of discrete documents via the Web.

At the end of the 1980s there were a number of useful summary articles about the systems of the time. Nielsen's 'Two Basic Hypertext Presentation Models'[27,28] [360] which had the sub-heading 'Card sharks vs. holy scrollers: the two main ways of presenting interlinked information'.

[27] The phrase 'Card Sharks and Holy Scrollers' is attributed to Raskin when presenting his paper[295] at the 1987 Hypertext conference paper, though the actual phrase is not used in the paper itself.

[28] This was given further context by Bernstein's blog 'Card Sharks and Holy Scrollers' [76]

The Card Sharks were those systems that either used fixed-sized lexia that filled the screen, each replacing the last, like ZOG and PROMIS, or where a form of Stretchtext was used such as in Guide. The Holy Scrollers were those systems where lexia were of arbitrary length and which could be scrolled to access the whole content. The latter form is more familiar to us know and the former somewhat quaint, but this was the technology of its time in flux. Passing output to a *screen* was a relative novelty and today's rich rendering choices were still a long way off.

Young's 'Hypermedia' article appeared in MacWorld in 1986 [393]. A year later Conklin published 'Hypertext: An Introduction and Survey' which was a more serious survey of then-available hypertext systems.[88]. Lastly, Müller-Prove's thesis 'Vision and reality of hypertext and graphical user interfaces' [254] has a useful, illustrated description of the key early hypertext systems

2.3.1 Research from Elsewhere

In researching early hypertext it is noticeable that the work appears to be centred on the United States and in Western Europe. This offers up a question as to whether hypertext research was happening further afield before the arrival, and uptake, of the Web during the early 1990s.

I found very little from the Far East, the earliest hypertext paper being from Japan in 1990, Hirakawa [175] on hypermedia for design support. There is then a gap until the early 2000s and Nakakoji's papers [257, 258], relating to Spatial Hypertext (q.v. section B.3).

In Eastern Europe[29], in the Warsaw Pact countries and USSR, there is also scant literature. There are a few papers[30] indicating hypertext experiments in the late 1980s, the earliest being 'HYPERLOG' developed in Moscow c.1988 [395] along with [140, pp.192-193] 'HYPERNET', a visual front end for BAsic Hypertext System (BAHYS) [208], and 'SEMPRO'[31] [140, pp.192-193]. In Kiev, Grinchenko was working on 'HYSPY'[32] from c.1990 [148].

The Soviet work centres on analysis of information networks. Unsurprising perhaps, in the context of a large, centralised command economy system. Peters' *How Not to Network a Nation* [287] indicates how difficult it was to implement computing systems within the USSR.

Proving absence, i.e. looking for what is *not* present, is difficult. The fact that some of the areas above use non-Roman alphabets and were also socially and politically isolated in the last half of the twentieth century means there may well be further papers, indeed earlier papers, as yet hidden from the wider view. However, the small volume of papers I found points to hypertext (and via it the Web) being very much an invention of Europe and English-speaking North America.

[29] Europe was divided by the 'Iron Curtain' until the revolutions of 1989 opened borders. The Union of Soviet Socialist Republics (USSR) itself lasted until 1991.

[30] This does not preclude there being earlier papers in published in the USSR in Russian and not published more widely.

[31] From NOVINTEKH, a joint Soviet-Finnish-Bulgarian venture.

[32] Some papers refer to it as 'GISPI'.

2.4 'Seven Issues' for Hypertext

In 1988, Halasz, wrote a paper for the *Communications of the ACM* magazine known as the 'Seven Issues' paper [154]. This reflected from Halasz's experience with NoteCards [157] on which he was working at Xerox PARC. It had shown that users struggled with correct use of typed links[33]. For instance, Trigg's PhD thesis [353, pp.2–3] had identified over 80 [sic] discrete types of link that hypertexts might use; Trigg also worked at PARC with Halasz.

At the time, hypertext search was still a nascent feature, meaning users were over-dependent on the author's links. This latter issue, of the problem of search and query, became Halasz's first 'issue'. His resulting 'Seven Issues' are illustrated at figure 2.2, left column.

Ironically, the World Wide Web was proposed not long after Halasz presented a follow-up 'Seven Issues: Revisited' in 1991, included escaping 'the tyranny of the link' [155, p13]. The latter issue was a compression of three previous topics (search and query, composites and virtual structures) from his 1986 paper [157]. His point was to elevate use of the, then novel, search-derived links to be on a par with traditional deliberate (typed) links created by the hypertext's author. In the light of today's Web this seems unexceptional but was not at the time. Even if unintentionally, Berners-Lee's World Wide Web upended the status quo as regards hypertext linking (mainly as a result of what it could not support due to technical limitations of the time).

Halasz revisited the list a third time in a 2001 post-Web paper in the Journal of Computer Documentation (JCD) [156]. Furuta also presented a reflection on these lists, 'Halasz's ' 'Seven Issues' in Context' [128].

In figure 2.2 I have linked the issues together to show how the list, and its focus, changed over time. The problematic aspect is that these lists look—from a technical perspective—as if the problems are all resolved; thus, for some, hypertext as a subject was 'done' and attention drifted to the Web. Yet in the current Web paradigm, links are one-way and embedded in the Web 'document', at least at point of render in a browser. This (technology-limited) approach precluded adopting the rich systems of typed, non-directional links stored in external link tables that were being developed in the mainstream hypertext research community.

In his keynote for the 1987 Hypertext Conference [96] van Dam also gave a list of issues to concern the hypertext community. This reflects the fact that the end of the 80s was a time when hypertext was a subject of interest and a number of different comporting systems were around. The Web had yet to arrive.

At Halasz's 1991 keynote, he also described a number of 'dimensions' were there potential divergences in view or practice. The latter are shown in Figure 2.3.

[33] Bernstein recalls Trigg reporting NoteCards users found choosing the correct link type simply too hard and so tended to use no link typing at all, but this cannot be further sourced. Personal email correspondence with Mark Bernstein, Eastgate Systems Inc. (Sep 2015).

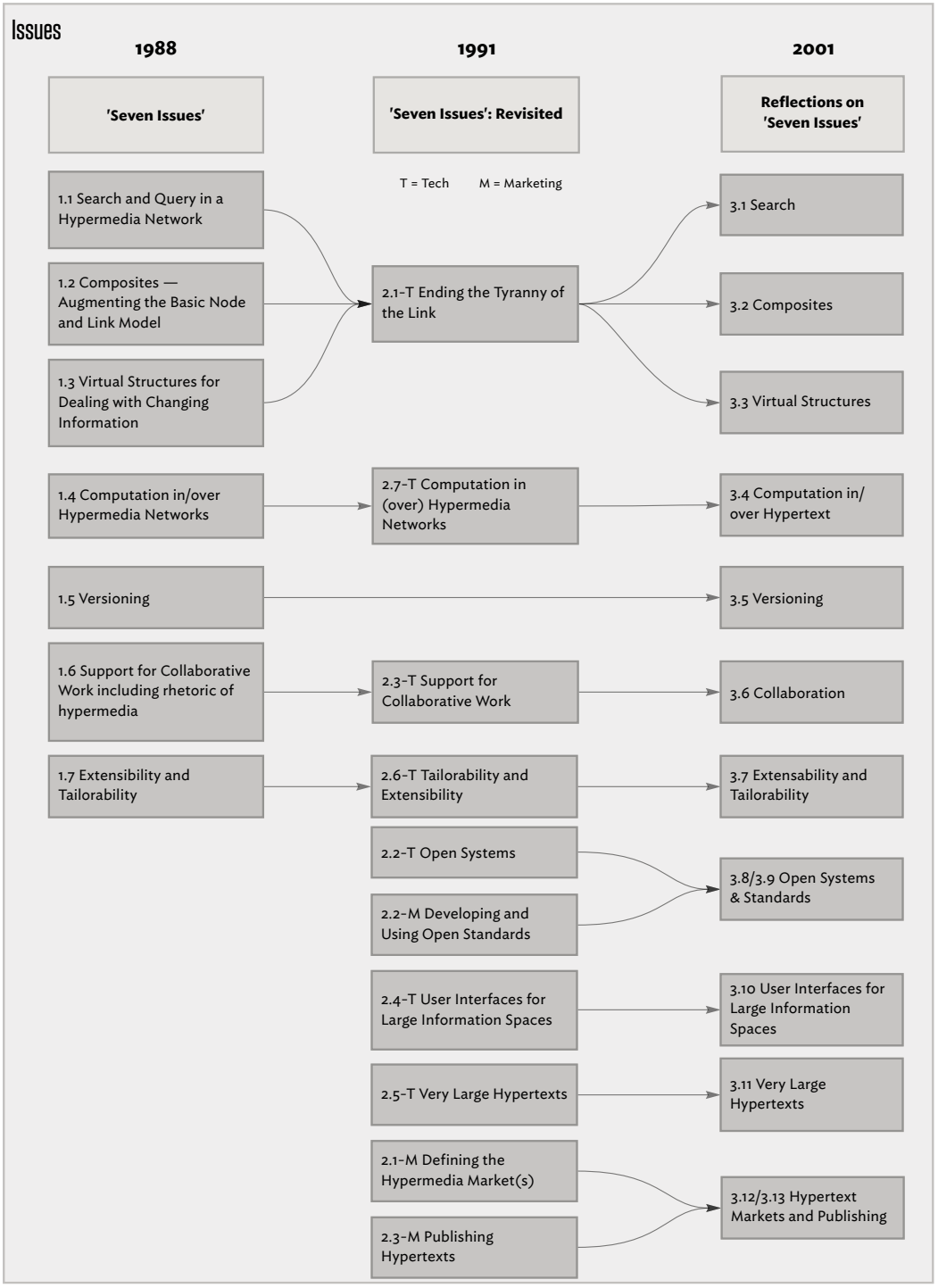

FIGURE 2.2: Halasz's '7 Issues' Mapped: 1988–1991–2001 [154, 155, 156].
Issue captions are as per source papers.

2.5 Narrative & Spatial Hypertext

As the 1990s started, the arrival and increasing predominance of the World Wide Web, diverted research efforts away from mainstream hypertext. Despite this two areas of hypertext research

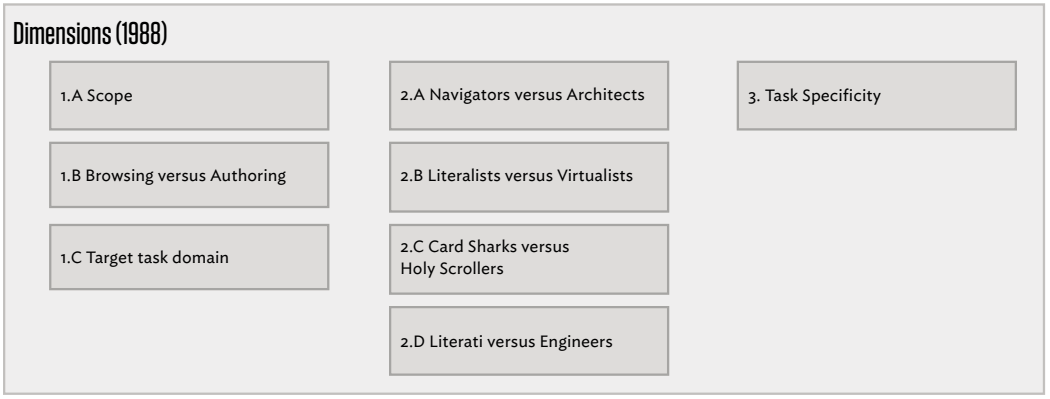

FIGURE 2.3: Halasz's '7 Issues': Dimensions (added in the 1991 [155] iteration)

continued in the margins. These areas were hypertext narrative and spatial hypertext, which are explored in greater detail within Appendix B.

Non-linear narrative reached beyond technology out into various parts of the Humanities where people were experimenting in new forms of literary structures. As well as those interest in post-modern narrative and critical theory there were was also interest in the creative area of ELit. Although the latter made its way onto the Web, the earliest Web was not as feature rich as the pre-Web hypertext is superseded. Hypertext narrative is discussed in more detail in narrative in Section B.2.

Spatial hypertext is an interesting divergence of the medium, whereby a note's position on a 2D (or 3D) space has a meaning from which inferences may be drawn, with or without the additional affordance of links. Spatial hypertext in Section B.3.

2.6 Bush's 'Trails'

Bush's notion of 'trails' [71, Section 6, paragraphs 1–2] are much mentioned and cited, but this has not been reflected in hypertext thinking with few papers on this topic. Trigg's 'TEXTNET' from 1986 [356] included a notion of 'paths' which are explicitly references to Bush's trails [356, pp.3, 6, 12]. Paths could be "…used to generate sequential walks over a subset of the network". Two years later Trigg described the 'guided tours' in PARC's 'Notecards' program that allowed an author to define a tour that another person could follow through the hypertext with the option to 'jump' out at any time [354, pp.219–221].

Zellweger's 'Scripted Documents' [397, 396] only allowed use of preset paths through its corpus. It offers a useful list of other contemporaneous programs using trails [397, p.6]. Hammond & Allinson's 'Guided Tours' [163] generated a tour based on a number of questions to the user and Atkinson's 'Hypercard' [144] allowed for programming so as to offer a degree of choice (the onward route selected might depend on a checking a variable set elsewhere or perhaps on data called directly front the user.).

As part of Nyce & Kahn's Memex retrospective [279] Trigg's 'From Trailblazing to Guided Tours' [355] noted that a limitation thus far was that systems were single-user so the available corpus of work was that available in the current system and that Notecards had recently added a collaborative mode added allowing user's to interact with other users work. Bear in mind that hitherto a hypertext was perforce a personal system as networks were in their infancy. Whilst users might share a hypertext *sequentially*, collaborative systems removed that barrier—and added new challenges for design and understanding. The addition to Notecards also lead to the 'history files'—a record of where the reader had been, within the hypertext—something now familiar to users of a web browser.

Furuta *et al.* 'Hypertext Paths and the World-Wide Web: Experiences with Walden's Paths' [130] discusses how a path based approach might be used in a Web [sic] hypertext context for teaching school children. An interesting side-note was the tension between some parts of the system wanting fixed media (e.g the authors of content wanted the content to stay how they set it) and some wanting more variation (keeping the experience fresh for the students)

Trailblazer [351][34], was a Chrome-only plug-in extension to track (logged-in) browsing use. Recent history is displayed as a radial tree graph. Although its designers saw it as a means for recording/showing browsing history, it shows promise as way to record passage through a hypertext allowing a maintainer to spot loops and missing connections.

2.6.1 Ways to Visualise Trails

Links and trails are often referred to as abstract affordances. On the Web page or many hypertext systems links are clickable anchor text or images, resulting in new content replacing old, the role of the link is a means of travel from node to node. Less explored are ways in which the graphic arts may inform how we conceive linkages, and also visualise our path through hypertextual spaces.

In his book discussing comic book structure, McCloud's *Reinventing Comics* [35] links the notion of hypertext and the comic book form [239, pp.215–229]. He gives a strong visualisation of a narrative thread in his article I Can't Stop Thinking! #4 [186]. That visual style is echoed in Shiga's *Meanwhile* [312] where the normal linear page, if not panel sequence, is exploded and the readers journey bounces them round the work from page to page. More recently, Sousanis' thesis, in graphic novel form, *Unflattening* [325] (2015) explores notions of the constraints of 'flatness' and how to expand beyond the two-dimension, as on the printed page.

This perspective may seem frivolous, but it is worth bearing in mind that our eyes are a major human sense and we should be open to exploring other forms of literary exposition for ideas that

[34] Started c.2103 by trailblazer.io (site now dark), the tool was last reported as being in the process of converting to open source status as at 2017.

[35] Also see McCloud's *Understanding Comics* [238].

might help with the presentation of trails. Recall Weakland's initial questioning of the Bush's Memex trails and the ability to generalise in a way accessibly to all (most?).

2.7 The Advent of The World Wide Web

Easily overlooked in the early days of the Web was possible longevity of all this new form of content. Initial web hypertext focussed on the basic, but then little-understood, task of encoding web page documents, interlinking them, and serving them to the Web. The early web authoring and browsing tools were under continual development. They were also subject to the disturbance of early commercial entrants looking to exploit market share in this new arena by adding divergent and competing features. Examples of this were the sometimes conflicting and even mutually incompatible technical developments of Netscape's and Microsoft's web browsers, each trying to push for dominance of a new market.

Web hypertext standards were also emergent and not necessarily well understood. Creating and serving a working web site was not a simple endeavour for any but the most technically adept. Thus much early Web content was written and 'published' to the Web with little thought to the longer term. That was not by intent but rather because it was not obvious that the Web would become the pervasive method of data presentation and information access that it is today. For all but the largest sites, early web pages tended to be static HTML pages making linking comparatively simple. The editing effort needed to re-link existing content in static sites was a disincentive to making regular structural changes to hypertexts.

The advent of 'Web 2.0' at the turn of the 20th Century, saw a move to greater use of database-driven websites and the emergent activity of 'blogging'[36]. Blogs typically showed a 'home' page that was one or more recent articles, posing a problem for web author wanting to link to an individual article once it rolled off the blog's front page. This issue led to concept of the 'permalink'[37], a permanent static hyperlink to a particular web page or entry in a blog, normally by embedding the host database's unique article ID into the link. A permalink also becomes important when addressable content on a web 'page' is actually the output of a database query. More important than the exact technical implementation was the simple manner of matching a unique persistent URL to an asset other than static page without having to use an explicit query-style URL. A benefit, for the human user, was a simpler—and often shorter, more memorable—URL.

However, those changes only solved in part the problem of maintaining an evolving web hypertext 'longitudinally', i.e. over time. In this context, sites such as blogs are essentially akin to a paper newspaper—each article (edition) is printed and then archived. The paper goes to the

[36] Blog is a contraction of the term 'web log', a website where an author records personal views observations and links. The origin of blogging is further described by Rebecca Blood [54].

[37] An early reference to this technique is by Jason Kottke [205] in March 2000, followed shortly after by Matt Haughey's description of implementation of the method in the Blogger blogging system [170]. As shown by the URLs here, hitherto 'static' webpages tender to involve an outline form of storage of content.

archive (in paper or digitised from) whilst as long as the blog's server remains active, permalinks provide ongoing access to the old (essentially archived) pages. A more difficult challenge is a web hypertext where the structure of the subject matter evolves over time: a resource where internal structure is as important as external boundary linkages.

Repositories of organizational tacit knowledge and documentation of tools or processes have just this problem. Moribund articles can be 'archived' similarly to static news pages (as above), given that they do not change. They become less important except as a static reference of fact. The challenge is posed by maintaining areas where the subject matter changes in context. This task affects not only the internal structure and linkage of the hypertext, but can also affect the (known) inbound links and the search/exploratory affordances offered at the hypertext's boundary. In most cases that boundary implies a single website, but in largest sites this may be a sub-site of a larger website (or web domain) and may involve material stored over more than one server.

2.8 The Wiki

In 1994, Ward Cunningham developed the first 'wikiwikiweb' to simplify intercommunication between programmers, using the software to create a site in 1995[38]. The key features of the wiki were open collaborative access for editing (with optional logon). It offered a reduced mark-up system, compared to HTML, with simple automatic intra-wiki page link creation. The site's data is stored in a relational database and presented via a series of scripts and a web server as an editable website [93]. Importantly, whether access is logon-mediated or not, the wiki is essentially a commons belonging to all (possible) editors who are free to (re-)edit and to add new pages as needed.

Cunningham notes some of his design's antecedents in the original wiki, including HyperCard [94]. This shows the interrelation of the generic notion of hypertext (linked content) with software design. Wikis have been used publicly, for business and personal use; in the latter case TiddlyWiki [348] creates a lightweight personal hypertext stored within in a single text file.

Cunningham's ideas were also publicised in Leuf & Cunningham's *The Wiki Way: Quick Collaboration on The Web* [215]. Typical of books of the time, this offered a brief overview of the concept and affordances (Chapters 2 & 3, only 33 of the 435 pages) followed by details about practical implementation of a wiki. AS such this gives less insight into the 'why' of wikis that the 'how'

The 'wikiwikiweb' software was subsequently made open source as 'wikibase', since when the wiki concept has been further implemented in a range of languages. The term 'wiki' initially referred to Cunningham's own website but has since, with his approval, been used to refer generically to any website using wiki software.

[38] The site was initially an adjunct to the 'Portland Pattern Repository' before growing to accommodate other topics before the latter split to the (now closed) 'meatball wiki'. The original site [92] is still online, albeit now read-only.

Wikipedia is the largest wiki-based website to emerge from the concept. Originally Wikipedia ran on Perl-based UseModWiki software, itself based upon WikiBase. Development by the Wikipedia community resulted in a PHP+MySQL-based derivative of UseModWiki that was subsequently published under GPLv2+[39] [142] licence as MediaWiki [243], the current software used to create Wikipedia. Mediawiki supports active transclusion of whole, or part, articles as well as substitution[40].

2.9 Transclusion

2.9.1 Nelson's Transclusion

Ted Nelson's original concept of Hypertext transclusion [266, p.4/76] [259, p.96], at least in his 'Xanalogical' form, is a means of preserving provenance by linking content transcluded into the current with its original source context such as might be useful for scholarly cross-referencing. Nelson describes it as "reuse with original context available, through embedded shared instancing" [269, p.32]. In Nelson's words transclusion is indicated (speaker's emphasis) "...with a strap or beam or bridge VISIBLY connecting them .....with pages VISIBLY connected, side by side..." [273][41] [270, pp.3–7] to describe verbatim the function of the re-use of (part of) another document in the current document. However, the term has since found a wider meaning, beyond its creator's intent, and without his 'visible links', even if he still dissents [274, 2:24–2:29].

Nelson's original vision is tightly bound to his other ideas over assertion and retention of artistic copyright. To Nelson, as the son of a film director, this (visual) provenance was important both to indicate prior art and to underwrite copyright. Reflecting this, Nelson's Xanadu concept also involved early ideas for micro-payments such provenance might imply [266, p.5/13]—before a suitable infrastructure existed (the Internet was then in its infancy).

However, the underlying concept works in a wider context such as the Web. Transclusion[42] retains the canonical form of the referenced source text: if the original is correct, so too will be the instance within the transcluding page. In a hypertext, including Web hypertext, it is possible to mark and link the transcluded passage. In large well-interlinked hypertexts the use of transclusion offers an effective tool for assisting in the maintenance of coherent description. It otherwise avoids fragmentation of content and provides a single canonical source for a given subject. The

[39] The '+' essentially implies 'GPLv2 or any later version', i.e. currently GPLv3.

[40] Essentially, a one-time transclusion, after which the page is updated to use the substituted content and saved in that form; source linkage is discarded.

[41] WSI Lecture, Southampton University 3 February 2106, see video at 43:50 [273].

[42] Transclusion, a neologism of Ted Nelson, describes the hypertextual inclusion of a whole or partial resource with the body of another resource. Nelson first used the term in *Literary Machines* [268, un-numbered preface p.5, footnote]. Barnet [30, p145, note 14] reports this was in the original 1980 edition but lost from subsequent reprints. Nelson gave a more explicit notion of transclusion in his 1995 article "The Heart of Connection: Hypermedia Unified by Transclusion" [269].

latter can then be transcluded, or included, into other webpages either as discrete pages of content or as inline content (which can be addressed directly via HTML anchors).

Transclusion can work in two possible ways, depending on whether the target content changes if the source content changes. In Nelson's model, a transclusion is fixed. He does not state it that way but the Xanadu system requires a new resource (document) for any change, even a single character typo. As transcludes link to a resource, the transclusion is essentially fixed. Another method is that used by the Web where the receiving programme transcludes whatever is at the called Uniform Resource Indicator (URI). Thus if the source changes, the transcluded content changes. The second model breaks the provenance implicit in Nelson's model but it is what the Web uses. A definition of transclusion such as that by Glushko, certainly leaves the exact meaning open to interpretation:

> "the inclusion, by hypertext, of a resource or part of a resource by another resource"
> (*The Discipline of Organizing* [141, p.231])

Thus transclusion offers a simple means of re-using (aliasing) pieces of original content—normally text—without the scope for errors that arise from re-entry of data (typing errors) or from lack of coherence when the same descriptive text is re-written in multiple discrete contexts. Looking beyond textual use, transclusion essentially implies (verbatim) re-use. The functioning of Transclusion is described further in Appendix C.

2.9.2 Transclusion on The Web

Transclusion is dissimilar but convergent with the software coding notion of an 'include'.It derives from use of code 'libraries', which originally sought to share data between systems[43]. This led to coding languages using libraries to separate core code from less-used task-specific libraries. Libraries are called via an 'include' statement, giving the notion of an 'include'. Thus an emergent property from erstwhile technical limitation has yielded a general re-use principle. Drawing on Alexander's patterns for architecture [15], in the 1980s software engineering also adopted the idea of (re-usable) patterns, further encouraging use of inclusion.

Although Nelson worked on early computers, his writing [259, 266] makes no mention of includes. Whilst the origin of includes slightly predates Nelson's description of transclusion, his concept appears a genuinely separate derivation, albeit convergent with includes.

Conveniently, the notions of libraries and includes have meant that effecting transclusion cannot have been too alien to coders of early hypertext systems and subsequently, of the Web. If the

[43] The term's exact genesis is unclear but has roots as far back as the American SAGE air-defence system in the 1950s, which used the JOVIAL language (amongst others). JOVIAL had the concept of a 'COMPOOL' ('COMmunications tag POOL') allowing access by multiple systems to working data (discrete from implementation code). Essentially this gave common data at a time when main memory was expensive and limited, and data sharing was a problem. Later COBOL initially had no standard library, but added library support in the 1970s (COBOL-74) and since then library support has become more common.[225]

original Web concept appears not to address re-use directly, it is implicit in the fact that multiple access to Web assets allows for such behaviour. In hindsight, it should be recalled that the new Web was a new and unproven network with bigger implementation challenges than transclusion. Arguably, the earliest web page transclusions were via the 'src' attribute used by the `<img>` tag in HTML 2.0 [44] [45]. These allowed insertion of images into what was then otherwise strictly text; the same source image could be used by *multiple* web pages.

Inclusion also found its way to the Web via 'Server Side Includes' (SSI) in the mid 1990s[46]. SSI solved a problem of delivering customised content at a time when delivery was only via static web page. The advent of HTML 4.0 in late 1995 saw the advent of the iframe (inline frame)[47] that allowed whole HTML pages to be transcluded/included into other pages without using a formal 'frameset', where every frame in the document window is a complete web page. The iframe allowed inclusion (albeit of another page). Again at this point the inclusion was static, although the iframe content could be refreshed independently of its host page.

Although Webpages had allowed 'frames' iframes were a newer and subtler frame variant that allowed inclusion (transclusion) of content from other sources in a less visually intrusive way: at the time, hiding ' the 'chrome' (surround) of a frame was not easy for most webpage authors. Whilst this solved an aesthetic issue for the mass-market it also made it easier for the less well-intentioned webpage maker to use (or substitute) nefarious content to which the webpage user was likely unaware. Whilst the web of today is (arguably) less insecure in this regard, at the time this aspect of iframes inhibited more general use of transclusion on the public web. By comparison, inside organisational networks were there was (supposedly) greater control over page content, iframes could be used with more freedom. It is worth noting the latter context was still dominated by one vendor—Microsoft.

Regardless, it was Microsoft's next innovation that moved the Web away from considerations of iframe use. With the release of version 5 of Microsoft's Internet Explorer web browser[48] came the introduction of the XMLHttpRequest. In turn, this led to the development of AJAX [12] (Asynchronous JavaScript and XML), which used a number of discrete Web features to allow dynamic web page content without the need for explicit page reload—such as was required with static-only pages. Reflecting the rising use of (X)HTML on the web and growing XML data resources, De Iorio's 2009 work [105] looked at implementation of transclusion in an XML context.

The usefulness of iFrames was offset by the security hazard posed for the unwary. Previously, Web pages with frames made it hard to disguise the frame boundaries and bad actors could

[44] The W3C describe initial use of images in Mosaic v1 circa December 1992 in 'Marc Andreessen makes a brief appearance on WWW-talk' [382]. Marc Andreessen proposed the IMG tag for HTML in early 1993 [22].

[45] The tag was present in the HTML 2.0 specification (RFC 1866) of 24 November 1995 [182]. Also see W3C HTML 2.0 Specification [380].

[46] Netscape's server added JavaScript in November 1994 [275].

[47] Originally a Microsoft-only feature in Internet Explorer, iframes were adopted as part of the HTML 4.0 specification (18 December 1997) [381].

[48] 'Microsoft Delivers World's Fastest Modern Browser Available Today' [248].

script the browser and make access between different frames or iframes to subvert content, either to inject unwanted content or place malicious code. The iframe, with its more subtle inline usage made in-page content inclusion less intrusive for the page designer but difficult for security staff to secure against attack. The arrival of AJAX and its derivatives, with the ability to load content into an active page made some of existing frame use unnecessary, and newer Web methods began to eclipse use of frames and iframes. Despite iframes' lack of built-in security[49] against misuse,they are still being used as a context for web-based collaborative hypermedia research, such as Bouvin & Kloknose's 'Webstrates' paper [61] at ACM Hypertext 2016. The 21st Century Web page can certainly technically implement content transclusion on the Web. Bouvin & Klocknose's method is an attempt to do just that, to return some of the former richness of hypertext to the current Web. The Web won out against richer erstwhile hypertext systems like Hyper-G and Microcosm though its simplicity. Whilst the Web experience has become visually richer and more interactive it has yet to regain all of earlier hypertext's features.

In 'From XML Inclusions to XML Transclusions'[105], Di Iorio & Lumley investigated using the comparative formality of XML as a means to implement transclusion in a Web context. Their end goal was for better documents within an enterprise workflow using their Transclusions in Enterprise Documents (TED) system. They recognise the temporal access, i.e. the potential change in the source document, and in their work they differentiate between 'static' and 'live' transclusions.

Tchernavskij's 2016' dissertation 'Transclusion in Software for the Complex, the Changing, and the Indeterminate' [339] nods to Nelson's first ACM paper on Hypertext and describes his 'In-Plenary' instructional program which uses transclusion.

Lastly, it is worth noting a very common form of transcluded content that is easily overlooked—embedded tweets. Twitter does not allow editing of tweets, though they can be deleted, so the source is (relatively) stable and tweets are now commonly embedded into other pages (using HTML and JavaScript AJAX). It does offer provenance of a sort but it is as likely to be due to ease[50] leading to a copy-cat approach: if all around are embedding tweets with a consistent style, those who do not begin to seem egregious in their methods.

2.10 Bots

Alan Turing's 1950 paper 'Computing Machinery And Intelligence' [359] contained prophetic discourse of computer 'thinking' powers not possible at that time. Such considerations inspired the first AI 'chatbots' such as ELIZA [372]. However, there is a broader concept of a 'bot' beyond

[49] The security issue is less one of the frame or iframe per se, as the lack of browser scripting structure designed from outset to control malicious behaviour, even if such are obviously different, before the fact, from legitimate use of the same techniques.

[50] Twitter makes it easy for others to re-purpose its content in a manner that advertises Twitter's brand

'intelligent agents'. It builds on the notion of software agents. Although there are also (ro)bots that are not solely software, it is only software-based bots that are pertinent here.

Whilst discussing building agents to use in a Unix environment, Etzioni, Lesh & Segal [118] included this description of a 'softbot' or '*soft*ware ro*bots*':

> "A *softbot* is an agent that interacts with a software environment by issuing comments and interpreting the environment's feedback. A softbot's effectors are commands meant to change the external environment's state (...). A softbot's sensors are commands meant to provide the softbot with information about the environment (...)." ([118, p.10])

Etzioni and Weld [119] developed the idea further with an 'Internet softbot'. Writing from an AI-informed perspective, Wooldridge and Jennings' [379] explored the theory and practice of agents, including the above definition as an example of 'a weak notion of agency' [379, pp.116–7].

At the same time, Franklin & Graesser have described a taxonomy for autonomous agents [127]. They reference Wooldridge and Jennings concepts and Eztioni and wild 'Internet softbot'. However, their taxonomy it is somewhat abstracted the emphasis on AI concerns rather than more general automation is distancing in terms of bots as the lay user might encounter them in the likes of a collective system like Wikipedia,

By comparison, Leonard's [214] *Bots: The Origin of New Species* is a very approachable summary of the history of bots and how they escaped early Artificial Intelligence (AI) labs to find a life in online games and as digital door-wedges in the early days of less pervasive and persistent online networks[51].

Geiger's article 'The Lives of Bots' [133] actually focuses on a single Wikipedia bot and in doing so reveals a number of the social and functional issues raised by bots operating in a single online community. His paper 'Are Computers Merely "Supporting" Cooperative Work' for a doctoral colloquium [134] looks towards establishing an ethnography of bots. Importantly, in a socio-technical context, is the consideration of the bot makers and operators as well as the bots themselves. In his paper 'Design by bot. How Application Programming Interfaces (APIs) have enabled new modes of software development that complicates traditional distinctions between developers and users' [135], Geiger notes:

> "A developer may have to request access to an API or get authorization to run a bot, but this barrier is often far lower than getting a patch merged, much less full commit access. Bots can be written in most modern programming languages, and

[51] This at a time when dial-up modem access was common it was rare for anyone to be fully online 24 hours a day. Bots were an early way to simulate a persistent presence.

> for specific platforms (like MediaWiki), frameworks have been developed in dozens of languages to make it easy for a novice developer to write their own bot in their favorite language. Bots do not require or allow the developer to interact with a platform's code, and as such, bot developers can treat the underlying software platform as a black box. However, bots raise new issues in the development of software, as bot developers cannot rely on the same regimes of sovereignty that make a feature embedded into a platform instantly implementable. Bot developers struggle with new issues of legitimation and negotiation, . . . " ([135, p.2])

Essentially, in systems allowing bots that are more than trivial, the bots potentially have a distorting effect, with a privileged bearing on the system's development—even if not by intent.

Long *et al.*'s 2017 CHI paper ' "Could You Define That in Bot Terms"?: Requesting, Creating and Using Bots on Reddit' [219] is most revealing. They comment:

> "We saw how many people simply have no idea what bots *are* at a functional level, how they act and react to data, or how sophisticated they can and cannot be. This included little understanding of the technical fundamentals of bots; budding botsmiths did not appreciate the additional infrastructure required to host and operate bots, nor understand the most basic of programming and networking competencies required to realise their idea." ([219, p.3496])

They also note that both parties—those building bots and those requesting them—are often talking past each other for want of a common terminology.

More recently, Gorwa & Guilbeault [146] 2018 paper discussed a taxonomy of bots across social media[52]. In their conclusion [146, pp.22–23], they note that use of quantitative methods alone fails to give a clear picture and that so far there is a lack of the needed balancing qualitative work in understanding the use of bots.

2.10.1 Bots within Wikipedia

There is some literature relating to bots in Wikipedia but as will be seen, mostly looking at sociological issues in the interactions of human and bots. Missing from the discourse is any real consideration of whether bots are deliberately used in a fashion that helps the hypertext as a whole, as opposed to any editor or group of editors. My point here is not about the merit of such papers within their intended context, but to highlight the lack of attention shown to the hypertext itself.

[52] I would content this category also encompasses Wikipedia as its attempts to follow current news and trends means (parts of) the hypertext as a whole functions like a form of social network.

Bots are recorded in Wikipedia as early as the end of 2002 (see section 5.1.2) but took some while to achieve acceptance and for their role to be formalised (see section 5.1.3). It is not until 2009 that the role of bots begins to be discussed. Geiger's [132] short paper usefully charts the rise in bot edits, specifically in the anti-vandalism role.

The general role of bots in collaborative hypertext is discussed in Section 6.4.1.

2.11 The Semantic Web And Social Machines

The original vision of this is Tim Berners-Lee's description in his book '*Weaving The Web*' (2000). He wrote (author's emphasis):

> "I expect these tools to develop into a common new genre on the Web. Real life is and must be full of all kinds of social constraint—the very processes from which "society" arises. Computers help if we use them to create abstract *social machines* on the Web: processes in which the people do the creative work and the machine does the administration." ([35, p.172])

Relevantly, a few pages later he wrote (author's emphasis):

> "The first step is putting data on the Web in a form that machines can naturally understand , or converting it to that form. This creates what I shall call a *Semantic Web*—a web of data that can be processed directly or indirectly by machines." ([35, p.177])

This usefully makes the link between human and machine via the Web, indeed via the Semantic Web. However, it appears the term 'social machines' actually pre-dates Berners-Lee's initial use of the term [310, p3], expanded in Berners-Lee, Hendler & Lassila's 2001 paper 'Semantic Web' [36]. Nonetheless, Shadbolt *et al.* hold that social machines as we see them today are very much in the spirit of Berners-Lee's use of the term. Indeed, Berners-Lee's description would now be seen as akin to a canonical definition and it is the definition I have followed within this thesis.

2.11.1 The Semantic Web

The affordances of wiki were not lost on those looking at the Semantic Web. Tolksdor & Simperl's 'Towards Wikis As Semantic Hypermedia' [350] which showed how the then disparate areas of the Web and semantically structure data might be tied together more closely.

In 2006, returning to the subject of the Semantic Web, Shadbolt, Berners-Lee & Hall [309] revisit the 2001 paper ([36] above), noting (my italics):

> "As we build ever more complex computational artifacts and information infrastructures, we observe that large-scale behavior emanates from small-scale and local regularity. We need engineering methods to ensure that our structures conform to reliable and repeatable design requirements.... New rules of interaction such as peer-to-peer protocols result in new macro behaviors—behaviors we can exploit and also analyze. These micro rules can occur at different levels of abstraction—*the rules of Wikipedia are beguilingly simple but lead to overall coherence*. Local-scale changes in Web architectures and resources can lead to large-scale societal and technical effects. How so?" ('The Semantic Web Revisited' [309, p.101])

That observation on the 'rules of Wikipedia' seems an overly-optimistic statement, albeit written in the days when Wikipedia was still ramping up in size. Wikipedia today has a multitude of 'essays', i.e. (informal) rules, often inconsistent with each other and some editors are anything but happy, as will have been noted in my previous chapters. Coherence is not a description that now springs to mind in the context of Wikipedia: the face—the article–may have coherence but behind lie layers of breakage and unresolved technical debt. There are reasons of erstwhile expediency for the latter situation, but it does not change the reality.

In 2010, Boulain, Shadbolt & Gibbins offered a solution for richer communally-editable resources with what was termed a "semantic open hyperwiki" [60]. Atypically, the design gives thought to transclusion re-sue of content and this issue of context into which a transclusion may occur. As importantly is the blending of traditional wiki and semantic wiki functions. Such enrichment of a wiki with semantic data is not seen widely. Since 2013 Wikipedia has had access to a semantically structured 'Wikidata' [W390] though the use of the latter's data is via existing transclusional methods using reference code, as explained in an article on 'Inserting Wikidata values into Wikipedia articles' [W709]

Bizer, Heath & Berners-Lee described the 'Linked Data: The Story So Far' [53] explaining the use of Linked Data on the web and of the 'Web of Data' arising from such activities, including the 'Linking Open Data' Project [5]; Heath & Bizer went on to cover this and attendant technologies in greater detail in their book *Linked Data: Evolving the Web into a Global Data Space* [171].

Berners-Lee & O'Hara 2013 paper 'The read–write Linked Data Web' [37] makes the case that whilst the term 'Semantic Web' has gained little public stature, the systems to effect it are in place in the 'Linked Data Web', and outline some of the core technologies. The technical process for linked work may be in place but it is also noted that more human concerns over commerce and privacy continue to create siloed pools of data without the long envisaged seamless interconnection.

Kaffee *et al.* [194, 193] have begun mapping language representation in Wikidata in comparison to populations of native speakers, showing some significant imbalances. Wikidata's semantic structure is being investigated to look at the possibilities of bots being used to even out these disparities. Being structured data, bot activities ought to be more easily audited for errors or

inappropriate edits in as fashion that is much harder in the free-text of articles in the Wikipedia wikis.

2.11.2 Social Machines

Compared to the abstraction of the Semantic Web and linked data, the notion of social machines appears to offer a space for more varied interaction of humans and machines. I would emphasise the last part of Berners-Lee's initial quote above: "processes in which the *people do the creative work* and the *machine does the administration*'. The pertinence of this is that it is not, perhaps, where the meaning of social machines has since ended up.

Bryant, Forte & Bruckman's 'Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia' [67, p.9] (2005) is written from a sociological frame. It charts the progression of new Wikipedia editors. As such, it casts non-editing work (i.e. hypertext structure underpinning the visual content) as an almost accidental by-product. In fairness, this myopia—in the form of an unwillingness or inability to look outside the authors' disciplinary norms—is as prevalent in writing from the technical perspective as it is from the humanities. Lost in the middle is the new medium. The paper does note the ongoing change from the paper age, but they are less clear as to what follows from that

In their introduction to 'Observing Social Machines Part 1: What to Observe?' [101], De Roure *et al.* note "Very many examples of Social Machines are proffered by the emerging research community, from Wikipedia to Ushahidi, Galaxy Zoo, reCAPTCHA and Mechanical Turk."[53] [101, p.901]. Further, in the same paper, they comment:

> "This scenario illustrates a degree of automation involving machine-to-machine communication without human mediation, automatic assembly, and the presumed bot detection algorithm illustrates a built-in observation mechanism."
>
> ('Observing Social Machines Part 1: What to Observe?' [101, p.902])

In that social machine scenario, the human is not even present, a nuance that will become of more relevance below. Smart & Shadbolt's 'Social Machines' [321] looks beyond the use of social machines as a portmanteau term and tries to define more closely the essence of social machines. Extracting their descriptions from the more fulsome source list (these are the bolded terms therein), they state: "social machines ... are *social* machines ... are systems that draw on the complementary contributions of both human and machine elements ... are forms of biotechnologically hybrid systems ... are Web-based systems ... may function as social systems" (*ibid.*, p.4). But, except insofar as the Web is a form of hypertext, there is no hypertextual relationship between actors implicit in this description.

[53] See 'Ushahidi' [362], 'Galaxy Zoo' [131], Google 'reCAPTCHA' [299], and Amazon's 'Mechanical Turk' [16]

Meanwhile, in a follow-up to their earlier paper (above), De Roure *et al.* make this aspirational observation:

> "We believe that the notion of the Web as an ecosystem of interacting social machines is a useful analytic and design paradigm, providing an alternative conceptualization to models such as crowdsourcing, collective intelligence, and human computation. In other words, we see a community of living, hybrid organisms, rather than a set of machines which happen to have humans amongst their components. Social machines often change in behaviour (the perpetual beta principle of the social web) and their successes and failures inform the design and construction of their offspring and successors."
>
> ('Observing Social Machines Part 2: How to Observe?' [102, p.13:1])

From an overview perspective, such 'hybrid' settings make sense but there is little clarity as to what this looks like from a human perspective. Contrast this with the confusion seen in the limited human-machine interaction seen in my analysis of Wikipedia 'Bots requests' (see Section 5.3).

In Hendler & Berners-Lee's 'From the Semantic Web to social machines: A research challenge for AI on the World Wide Web' [172] (2010) the subtitle indicates a drift to an AI-centric view of Social Machines, with humans in less of a central role. They conclude:

> "We hope we have demonstrated that exploring this new generation of social machines can take AI researchers and others into the design of new algorithms and interfaces; into new approaches to distributed inference and query; and into developing declarative social machinery, including policy-aware systems of privacy, trust and accountability." ([172, p.160])

Somehow the human part of the socio-technical social machines has been forgotten. This shift is seen as clearly in Hendler & Mulvehill's *Social Machines* [173], though there is a chapter on 'The Limits [sic] of Humans', this is clearly a technologists machine lead view of social machines.

What seems strange is to see those who first described the semantic web ignore the human side of the interface. Do we have so little faith in understanding our (human) thinking that we only trust black-box algorithms to divine the meaning? This seems a failure of imagination and one sign that the hypertextual medium is not to the fore here. If our human understanding and our links between objects (explicit or implicit) other than to better configure a subsequent algorithm, this seems less of a social machine and more a case of the human working to feed the machine.

The concept of social machines is quite central to Web Science, my field of study. Hall & Tiropanis observer in 2012: "In this light, we can define Web Science as the theory and practice of social machines." [162, pp.3863–3864]. Web Science diverges from the machine centric view as its ambit very much acknowledges the role of the human actor. This is somewhat at odds with

the description of social machines as seen in the grant application for the SOCIAM Project[54]. It states:

> " "Machines" used to be programmed by programmers and used by users. The Web, and the massive participation in it, has dissolved this boundary: we now see configurations of people interacting with content and each other, typified by social web sites. Rather than dividing between the human and machine parts of the collaboration (as computer science has traditionally done), we should draw a line around them and treat each such assembly as a machine in its own right comprising digital and human components—a Social Machine. This crucial transition in thinking acknowledges the reality of today's sociotechnical systems. This view is of an ecosystem not of humans and computers but of co-evolving Social Machines."
> ('The Theory and Practice of Social Machines' in SOCIAM grant application [323])

Once again, although both humans and machines are mentioned, the relationship is envisaged in terms of machines, the humans simply being part of the machine. This seems naïve, in overlooking the significant differences between the two group's abilities.

By comparison, Smart (2017) defines a 'Knowledge Machine' as "A knowledge machine is a social machine that engages in a form of knowledge-related activity. Such activities include those associated with the elicitation, acquisition, and representation of knowledge, as well as those associated with the discovery of knowledge and the development of intelligent systems." [320, p.3]. Though a somewhat abstract paper in scope it does seem to offer more hope for inclusion of a human element into the social machine as the reference to the manner of dealing with knowledge for once seem closer to the to human and the machine actors.

Examining co-operation with Social Machines, Applin notes: "the Social Machines construct encourages exploration of distributed cooperative computing as a new means for connecting people and making possible new processes and outcomes not previously achievable" [24, p.67]. Thus, by providing a networked, shared environment—often, but not always, via the Web—that offers collaboration it avoids the historical need for co-location to get instant or near-instant sharing of results. This is an important point, discrete from consideration of relative inputs of human and machine. A significant upside to participation at distance is the ability to improve the diversity of view and experience of those engaging with a social machine.

Weber's 'A Common Language' [370] gives a historian's view of the emergence of collaborative 'social machines', drawing a line through from ARPAnet in 1960s. It notes the need to both recall how we got to this point and understand the meta-structure of these new machines. Whilst the technologists may tend to focus more on process and outcome, it is important to understand the context and history of the emergence of these new social machines. Not least, it offers a means to understand the perspectives and biases extant as these new forms of endeavour emerge.

[54] In 2012 a group of UK institutions led (then) by Southampton University started the SOCIAM (**SOCIA**l **M**achines) project.

2.12 Summary

In this chapter I have described the origin of hypertext, and the its pioneers: Bush, Engelbart and Nelson. With that context I have described earlier ideas that predate formal hypertext systems, before looking at the three progenitors of hypertext in more detail and their varied contributions to the field and the systems that then arose prior to the advent of the Web (and described in additional detail in Appendix B).

I have touched on how the initial development has given hypertext a potentially Western cultural bias—or received a lack of input from other cultural traditions. I then reviewed the seminal 'Seven Issues' from just prior to the Web. The arrival of the Web shifted research focus and I then touched on two surviving strands of the original field, Narrative & Spatial Hypertext (both then described in greater detail in Appendix B, Sections B.2 and B.3).

I then turned to describing the advent of the Web, and re-consider Nelson's transclusion in the context of the Web. This leads into consideration of the wiki concept of Web hypertext such as it used by large collaborative systems like Wikipedia.

Bots were then described in terms of the concept and their genesis so as to give some context to my study of bots in Chapter 5. Lastly, as a similar exercise for the discussion in Chapter 6 I look briefly at the Semantic Web and Social Machines.

In the next Chapter, 3, I will start by refining my research question into two pieces of primary research and the manner in which they will be investigated. I will then proceed to describe my research methodology, methods of investigation and the selection of the data used as the subject of my research.

Chapter 3

Methodology & Methods

'*On the Internet, no one knows you're a dog*'

Gerge Steiner, cartoon, *New Yorker* magazine, 5 July 1966. [224, vol.1, p.616]

Доверяй, но проверяй

Russian proverb ('*Trust, but verify*')

In this chapter, I will first explain the manner in which my Research Question is to be undertaken. I will then discuss my methodological approach used for my my two pieces of research both conducted using Wikipedia as the data source, as subsequently reported in Chapters 4 and 5. Lastly, I will describe my methods for acquiring and using the research data.

3.1 Research Approach

In the Introduction, I stated my over-arching research question thus:

Does editing of Wikipedia show signs of sophisticated hypertext techniques?

The aim of my investigation was to use a large, collaborative, publicly editable hypertext as the research context for my investigations. For the reasons stated in Section 1.3, I had selected Wikipedia as the most suitable candidate hypertext. From that decision, my next task was to choose the mechanism within Wikipedia to test the hypothesis within Wikipedia.

3.1.1 Knowledge in Hypertext

This might seem an odd entry into a discussion of methods, but (re-)connecting the present with the less techno-centric thinking of pre-Web hypertext is a necessary grounding to what I wish to explore. in this thesis. Academe embraces narrow distinctions, the better from which to build argument. Whilst Socratic argumentation may improve the rigour of output, it can also corrosive of, and intolerant towards, improvment or innovation: "four legs good, two legs bad" [283, p.38], brooks no dissent.

My aim here is not narrow debate on the meaning of 'knowledge' as a context of argument, but instead to consider the reality that knowledge resides within our organisations' docuverses. Importantly, this is especially true of the evanescent system knowledge that is the sum of its parts, as distinct form the facts residing within any given node. Rather, I wish to address the harms that become such system-encapsulated knowledge if we fail to attend to the needs of the docuverse as a whole. Should we do so, we are collectively the poorer for it whether complicit in the act or not. Poor organisational resources fail the entire organisation, not just those working on the resource; the harms multiply unhelpfully for all within the organisation.

3.1.2 Finding Knowledge And Structure in Hypertext

The most commonly discussed aspect of Wikipedia is its users' edits or—more precisely—the counts [W641] of their edits, be it through the lens of type of edit being made (e.g. revisions, minor corrections), the type of editor (style, gender, nationality, human or machine, etc.), and other perspectives. Thus, regardless of the researchers' viewpoint, much research and commentary into Wikipedia, both qualitative or quantitative, essentially rests on consideration of abstracted counts of edits. Edits are the changes made to discrete content pages within Wikipedia by both human and machine user accounts[1]. Edits may occur in any of Wikipedia's multiple namespaces, though primary research focus tends to be on edits in the 'main' namespace which holds the wiki's public articles, which form the content experienced by the ordinary web visitor to Wikipedia. Wikipedia's documentation offers a description of the basic editing process [W46] and the policies [W723] attached to editing the Wikipedia wikis.

With a few exceptions, all edits to a wiki, in any namespace, are logged[2] allowing later identification of which accounts made which edits. Editors may leave edit comments to inform others of the nature of the changes they have made, though these comments are of variable quality—and in areas of disputed articles may be disingenuous in intent. Edits else where in the system, such as changes to user profiles is sadly less diligently recorded with important state changes passing un-noted.

[1] Machine user accounts are generally referred to as a 'bot'. Such accounts always have at least one associated human user account. This aspect is explored in more detail in Chapter 5.

[2] The logged-in user account making the edit is recorded, or their IP if editing anonymously.

Early reconnaissance of Wikipedia when preparing for my thesis showed that the abstracted perspective of edit counts actually obfuscates much of the real functioning of the Wikipedia social machine. Sets of article edits may be useful for providing a large data corpus for training algorithms, but in such cases the source and context of an individual's edit data is less used than its contribution to the volume of the dataset. Likewise, edit counts may be useful if trying to map the social behaviour of interacting edits within a large group of editors, but such a lens offers little insight in to the content underlying such activities.

Regardless of their worth in other context, edits therefore do not give useful insight in to the structure (or lack of it) in a large hypertext, as they speak of activity within the hypertext but not of the hypertext's content or structure. I will return to the subject of edits in Chapter 6, but for the reasons set out above, I discarded edit analysis as a suitable method for my primary research.

A hypertext's link are another structural aspect that can be studied. But at scale these too, give limited insight into the context upon which my research is focused. Certainly, we can measure network graph statistics and show on how well cross-linked the hypertext's corpus is. However, such observations are devoid of any useful connection to the *content* of individual nodes within the hypertext. Why so? Consider, article A might link to article B, and reach article X by no more than N in number sequential links. But, it does not tell us if A linking to B is useful or even relevant to the knowledge within A. If B holds canonical information useful (necessary?) to the correct understanding of A, and B is changed, is A then updated or its maintainer(s) informed? A's (network) propinquity to X may speak as to the speed or efficiency of navigating the hypertext via links but it provides little if any information relevant to assessing the efficiency and robustness of knowledge encapsulated within the hypertext. Wiki links have no semantic content, for instance to indicate the intent of the linkage; most (intra-wiki) links are essentially a form of 'see more' or 'see also' pointers between discrete articles (documents).

Wikis also have external links, i.e. those links terminating outside the wiki. These are used primarily for referencing and provenance purposes[3]. However, analysed without the anchor text and its context, such links tell us little about the content they connect. If an article has N external links, we do not know if they service all the source text that needs such referencing or if there are actually further sources in the same article lacking such a link. Knowledge of the latter difference significantly alters the utility, indeed implicit veracity, of such simple link counting.

One might investigate such external reference links but this would be a significant undertaking across one, let alone several languages. Each article with external links would first need to be checked for correct anchors and missing links. All links would then need to be resolved (i.e. successfully de-referenced) and the targets verified as appropriate reference sources. For multiple languages' wikis this would need first-language or very good second-language skills. Ergo, in the context of this research, the wiki's links are less informative than might be assumed by the casual observer, without significant activity *outside* the primary subject of investigation. Studying

[3] Other uses include linking out to the websites of organisations, companies, etc., that have a discrete article page in the encyclopædia.

Wikipedia links—as in their existence and prevalence—was thus rejected as a suitable subject for this research.

A wiki is essentially a bundle of discrete articles with no inherent structure or index, but a structural property of Wikipedia that could be investigated is the categorisation [W631] of articles[4]. Amongst other roles, categories function as a proxy index and so might be a useful aspect of investigation.

Categories that are applied to a Wikipedia article are listed at the bottom of the article's webpage[5]. In principle, studying these categories should give insight into how users are structuring the content, especially as a feature of wiki software is that there is no inherent outline or index to the overall wiki, unless constructed by the wiki's users. Categories can be created by any user, and may be nested hierarchically in a form of index. However, delving deeper there are some important issues that need careful consideration relating to using categories in a research context.

Firstly, the use of categories covers (at least) two significant but discrete activities. Firstly, they act as a proxy index to article content. Secondly, they act as an organisational tool for back-of-house content management, e.g. marking an article as in the purview of a particular 'WikiProject' [W715] or as needing some particular form of review. There is seemingly neither a structural method nor community intent to de-interleave these discrete categorisation roles in a fashion as might aid more efficient analysis. Being able to tease out these separate activities would reduce noise and ambiguity in the results.

Adding to the ambiguities, Wikipedia categories are also folksonomic in creation, organisation and management. As noted, any user may create a category and, optionally, nest it in the existing category structure. As the community refines or retires categories, there is thus a constant gentle churn of the inventory of categories and the allocation of categories to articles. Whilst some of the best organised WikiProjects, or other user special interest groups, may apply categories diligently and consistently across groups of articles, this behaviour does not occur consistently across the wiki. There is nothing wrong with this folksonomic method of soliciting the organisational information, but it is not clear that the process has any rigorous oversight for correctness, completeness, coherence, and coverage, other than by the general assumption of there being 'many eyes' on the problem [296, p.30].

Article pages for any category will list all articles with that category applied to them, but application of a category to an article is a human-mediated task—often entirely a manual one. A category is only as accurate in coverage of its subject matter as the background (manual) editorial effort applied to the category. Thus per-category article listings cannot be *relied upon* fully to reflect their subject: unknown are the number of pertinent articles that are un-categorised, or incorrectly categorised. As such, categories, do not represent—in current Wikipedia form—an

[4] See also Wikipedia Help article 'Category'.

[5] In some special cases, e.g. forms of maintenance, a Category may not be listed on the visible page. The category's use can still be checked via other means, but hiding back-of-house activities avoids cluttering article pages.

effective method to study the sustenance and preservation of knowledge, regardless of their usefulness in other contexts. Given these factors categories were adjudged an unpromising research approach

Taking a rather different approach, I could have looked at user behaviour and intent via interviews with editors, but this too throws up some challenges to effective research outcomes. How to define suitably representative editors? Interviewing may indicate what editors *think* they are doing, or wish to do, but there is then significant separate work would still be needed to map the editors' presumptions about their work to the reality of their recorded actions. Bot editors are an additional problem as they cannot be interviewed. Although they have a human operator, bots form a far smaller group than that of general human editors, despite the edits they contribute. This means that even if bot operators were receptive to the study recruiting a representative group is harder, not least because bots are rather opaque in function to the outside observer (as will become apparent in Chapter 5).

Combined, these issues throw some doubt on the wider relevance of findings drawn from a comparatively narrow samples. Additionally, a more significant limitation is that for all the research I uncovered relating to Wikipedia, there was little insight into editing activity from the standpoint of hypertextual structure (other than as an emergent property). In turn this means there is little substantive information against which to verify the findings of such interviews. Both approaches—interviews and review of data—have merit, but the question is which order of doing these tasks will lead to more robust findings. Though of great use in rounding out our knowledge of this editing activity, I judged an interview approach far more likely to yield insights of significance *after* the real nature of the editing task was better understood and documented than at present.

I then looked for artefacts of the hypertext that might give a stronger clue as to deliberate work towards the hypertext as a whole rather than the visually rendered article text. This led me to consider transclusion as a focus for study. Transclusion is essentially the re-use of a single piece of source content (text, images, etc.) in more than one other place within the hypertext as a whole. Employing transclusion is a deliberate use of hypertext/hypermedia technique, given the structural implications of re-using canonical sources. A simple Web-based transclusion example is the insertion of images into a web page: for instance, a website's logo will likely use a single image file in multiple pages and the image content (the file's pixel data) is *transcluded* from a single source to multiple pages. Moreover if the source file's content changes, all the pages using it echo that change. The same re-use principle applies to text or other media. Transclusion of textual data can be very narrow, akin to semantic data inserts), such as a significant date, or it might involve pulling whole/part articles from elsewhere in the wiki and rendering that content inline into the current article.

I have already described transclusion in more detail in Section 2.9, as well as additional discussion in Appendix C. Wikipedia's mediawiki software uses transclusion extensively, even if not apparent to editor or reader. Wikipedia's mediawiki software's concept of transclusion use varies

notably from Nelson's original vision (as in Section 2.9.1), a difference which I will explore in mote detail in Chapter 4, but it is not a difference that would hamper investigation of transclusion.

Having considered the challenges of the various hypertextual aspects described above, and noting the deliberate hypermedial nature of transclusion, I chose analysis of the occurrence of content transclusion as the target artefact for my initial research. Another factor in that decision was that studying transclusion appeared less prone to ambiguity as to (the observed) purpose of its use compared the other approaches above, not least because the investigations would not be at such an abstracted level. A risk attached to this framing of the research is how little is on record about transclusion use in Wikipedia, with research more likely to be talking in theoretical terms of what might be done rather than what has been done. It may therefore be that little can be found or that what transclusion there is cannot be unambiguously detected. Thus, not without risk this seems an untested area worthy of explorations.

In revising my overarching question into two discrete pieces of research I will describe two primary sub-questions in order, noting that they were actually defined sequentially, the second drawing on the initial findings after the first piece of research was completed.

Reflecting the considerations above, I planned the first piece of research to focus on transclusion. As Wikipedia's mediawiki software uses transclusion widely as part of its general infrastructure, I narrowed the scope of investigation to analysing only instances of transcluded *article content*. This framing thus deliberately excluded 'back of house' areas of Wikipedia such as the namespaces for templates, categories, general administration and such. Further, I construed 'content' to be only textual content—stored in the article itself—rather than other types of media, such as images, maps, sound files, etc. Thus, I was essentially asking "Are Wikipedia authors deliberately re-using (textual) article content as part of their structuring of the wiki's hypertext?". As a result RQ1 was formally framed thus:

> **Does Wikipedia show signs of editors making deliberate use of content transclusion?**

As stated, the evolution of Research Question within my overarching hypothesis was sequential. Having done the initial data extraction and analysis for RQ1, and acknowledging the lack of structural work undertaken by human editors, I wanted to look at ways in which other editors might be working to improve the hypertextual structure within Wikipedia's 'docuverse'[6]. Bots, Wikipedia's the non-human editors appear in the literature about Wikipedia as a discrete type of editor both in terms of role [133, 207, 197] and interaction with human and other bot editors [358, 326, 349]. Bots are more often discussed in humanist content, looking at social structure: the detail of their actual work presumably being regarded as quotidian. This means less has been written about the finer detail of what they do within the hypertext structure, outside their flagship role in controlling vandalism of articles.

[6] A neologism of Ted Nelson, its definition is discussed in Section 2.1.

Early results from the first research task (fully described in Chapter 4) indicated less content transclusion than expected, implying structural work was occurring by other means. One possible resource for this could be bot editors. As with the planning for RQ1, I considered interviews with interviews with bot operators. The weakness of such an approach was that without a clearer understanding of the status quo of bot activities, it would be hard to situate properly the views of (human) editors and bot operators. Such an overview of bot activity and organisation proved absent. More fruitful would be to look at the tasks undertaken by bots and requests made of bots by the wider community of editors. The non-interventional nature (q.v. Section 3.2.1) of such an investigation would dovetail well for RQ1, with all research avoiding any direct communication with users and forcing an insights to be taken from what the hypertext chose to make available to the visitor. In this latter instance, the Wikipedia API [W3] was included as a potential source as it also required no overt communication with Wikipedia editors.

Usefully, not only would researching bot activity in this manner fill an existing gap in the literature but, more pertinent to my thesis, it would indicate whether bots were contributing to the structural quality and integrity of Wikipedia. Such research would also give some insight into whether humans and bots were addressing different or similar parts of this task, in harmony or otherwise. Therefore, RQ2 was defined thus:

> **Is there evidence of automated 'bot' processes being used to support the hypertext quality of Wikipedia?**

3.1.3 Research Question Refined

The outcome of the above activity of refining of the overall thesis question into actionable research questions was to generate the two items listed below:

- RQ1. Does Wikipedia show signs of editors making deliberate use of content transclusion? (Research in Chapter 4)
- RQ2: Is there evidence of automated 'bot' processes being used to support the hypertext quality of Wikipedia? (Research in Chapter 5)

These separate pieces of research, both using Wikipedia data, were conducted separately and sequentially. While both pieces of research were undertaken within Wikipedia, each required a slightly different approach due to the differing nature of the each task. These difference will be explained in the rest of the chapter below.

3.2 Selection of Methodology

3.2.1 Contact With Editors or Tacit Observation?

With the lack of a clear picture of of the status quo in the areas of research interest, this posed a choice. Firstly, I could interview Wikipedia users ('editors') and use their observations about their work with Wikipedia. This would give a view of what they *think* they are trying to achieve and could inform later study of their actual work as reflected in Wikipedia's contents. The second approach would be to look at the present structure of Wikipedia content, as reflected in its current contents. That picture could then be used to frame later interviews with editors if necessary. As stated above, both approaches have merit, the challenging is sequencing when neither has significant prior work in the research context.

Having considered both paths, a non-interventional approach was chosen lest participation in the subject hypertext caused active users to change their behaviour. Wikipedia offers free and anonymous access, and even logged into Wikipedia, merely reading data gives no immediate sense of the researcher's intent. By comparison, leading with editor interviews about an area of the wiki that could be likely improved offered the chance that the editor(s) might act on the subjects arising in the interview and change their editing behaviour. That would then make it harder to look at the overview as their could have been significant recent change, indirectly caused by the interviews.

Both courses of action were equally plausible in terms of execution at the time of initial consideration. However, by sampling the data first, there was a very limited chance of the observed resource being changed by that observation; that same could not be said for an approach of starting with interviews. With this in mind, the overarching tenet of my Wikipedia research was to do so as a tacit observer rather than as an active participant. A further factor influencing that choice was the thesis' aim of looking at hypertext as a *medium* rather than as a technology or as a refection of human users' activity and behaviour.

The chosen approach also simplified considerations as to if or when I might engage with other editors during the research. It seemed simplest and most consistent with the above choices that I would not make deliberate contact with other editors during the research and not to take actions (e.g. active editing of articles) that might generate such contact. This imposed some constraints on possible opportunities for additional information, as described in the next section below, but nonetheless gave a clear rubric as to contact with the observed subjects.

3.2.2 Differing Considerations for RQ1 & RQ2

Despite my observational role, the first two Research Questions (RQs) each required a different perspective. Having extracted and cleaned the data for RQ1 my first task was quantitative, to tabulate the occurrence of transclusion markers in mediawiki code. This produced metadata allowing a qualitative analysis of selected subsets[7] of the data to assess qualitatively the reasons for

[7] Selected subsets were used due to the sheer scale of the source data.

transclusion use and to look for link patterns. Wanting the flexibility to repeat the process in subsequent years, it increased the importance that my research not be announced to the Wikipedia community lest it trigger a change in their editing behaviour.

For the initial stage of RQ2 (Section 5.2), the initial task was an expanding iterative search to find and annotate sources that would identify bot accounts, their administrative bureaucracy and their actions. Much iteration was needed given the inconsistent quality of much of the relevant non-article pages within the Wikipedia namespace. In this task, the non-interventional approach precluded me asking for help finding better sources. However, to have asked for help would risk the observed activity changing and thus hiding the true state of the research target.

Having located the primary sources there was a further iterative task to find, clean and store metadata in order to achieve a consistent metadata dataset for all identified bots. This enabled me to make a qualitative assessment of the bots and their stated or inferred—due to lack of data—purpose. For the same reasons as above, incomplete and unfinished Wikipedia sources required multiple iterations across the dataset to drive out errors and reveal previously missed bot accounts (whose metadata then recursively had to catch up to the state of the main body of notes). With a resolved list of bots assembled, I then undertook iterative qualitative coding of the their tasking to assess its relevance to helping maintain the overall hypertext.

In the case of the second stage of RQ2 (Section 5.3), the first part of the research required multi-pass reading of every discrete thread in a large archive of requests for bot assistance. This was data unstructured apart from being broken into discrete threads, although the Wikipedia archival process had broken many of the outward threads. Incomplete logs, broken links, and accumulated technical debt in Wikipedia's back-of-house systems made this task more complex than assumed from initial exploration. The lack of a consistent structure in the source material required a flexible approach. The incomplete and inconsistent source data required investigation involving intentional incremental formalisation.

3.2.3 Methodology

I considered a number of methodological approaches, but initially discarded them all for their rigidity of process. The challenge faced in my research was the proof of *absence*, but an absence of what? From initial reconnaissance of Wikipedia data it was clear that for both RQs 1 and 2 there would be no simple metric to use. As shall be seen in Chapter 4 Wikipedia's use of transclusions for much of its assembly and rendering of article pages meant that de-interleaving content transclusion from all other forms of transclusion would be a complex task. In the second case, as seen in Chapter 5, the initial test of data showed a sort of 80/20 distribution. The available query of user roles yielded *most* of the bot accounts for the current time. But the remainder, whilst hinted at by references all over the target data, lacked a simple method of summary.

Thus, the two pieces of Wikipedia-based research required an open, flexible approach that avoided premature formalisation. As important to the process would be the tool used collate and code the

data. In each of the two Wikipedia investigations, there was a two phase process. The first involved locating the relevant data and then extracting sufficient metadata to enable review for coding. As it was not possible to predict accurately the nature of that metadata, I followed the general tenets of Content Analysis and Thematic Analysis as described in Flick, von Kardorff and Steinke *A Companion to Qualitative Research* [125], Flick *An Introduction to Qualitative Research* [124], and Markham and Baym *Internet Inquiry: Conversations About Method* [227]. Rigidity of approach was tempered by Shipman and Marshall's 'Formality Considered Harmful' [313] that reflects on the pitfalls of premature formalisation.

Being well used to Tinderbox, and its strong support for incremental formalisation, this tool was used for the qualitative and coding work.

3.2.4 Ethics And Potential Harm

Although Wikipedia is owned and published by the non-profit organisation Wikimedia Foundation[8], the content of publicly accessible wikis[9] that comprise Wikipedia is edited and maintained by volunteer 'wikipedians'[10], more generally referred to as 'editors'.

Wikipedia content can be accessed and re-used via either the Creative Commons Attribution Share-Alike (CC-BY-SA) licence or the GNU Free Documentation License (GFDL) as explained in the article 'Reusing Wikipedia content' [W674]. Licence terms for content other than text may vary, e.g. for image or audio files. The article relates mainly to (re-)use of content rather than research into the content itself. The article 'Researching Wikipedia' [W672] actually lists a range of pages that might be useful to people wanting to make quantitative observations about Wikipedia as a whole but which do not relate to my form of research.

Likewise, 'Researching with Wikipedia' [W673] is more slanted towards re-use of content than study of the content, indeed the *structure of the content* that forms the site's hypertext. 'Academic use' [W427] again centres on the re-use of Wikipedia content rather than study of the hypertext.

'WikiProject Wikidemia' [W720] states "This project, Wikidemia, provides a space for articles related to academic research about Wikipedia" but actually appears to be a moribund project last updated in 2009. The 'Academic_studies_of_Wikipedia' [W426] article focuses on either tech or human, but not the interstices between which is where my research lies. The article does link to a number of pages at Wikipapers (referata.com), e.g 'List of conference papers' [W395] (last paper 2016, over 3 years ago), 'List of journal articles' [W396] and 'List_of_tools' [W397].

[8] Wikimedia Foundation, Inc. [377].

[9] Although the English language Wikipedia is the largest and most used, there are over 200 other language specific wikis. The language of each wiki refers that used to write the (visible) content of articles. However, no individual language wiki is limited to a monoglot area of interest, the scope of articles simply depends on the wikipedians contributing to that wiki; some wikipedians contribute to more than one wiki. The English wiki has articles covering the globe. Whilst articles on the same topic in different wiki are often cross-linked, the articles vary in content and may be most fulsome in the wiki of the language of that area of that topic.

[10] This is term is explained by Wikipedia itself in 'Wikipedia:Wikipedians' [W714]. That English language page is also offered (in various degrees of translation and content completion) in at least 85 other languages.

Wikimedia Foundation's 'Research' [W137] page is more about outreach than research. Wikimedia meta-wiki's offers some 'Research Resources'[11] [W123] which points to the Digital Methods wiki [W21].

Perhaps the most useful guidance comes from meta-wiki's 'Research FAQ' [W121] is this useful section:

> "**Does my project need approval?**
>
> "Most research is conducted independently, without knowledge by or approval from the Wikimedia Foundation. Rarely, the Wikimedia Foundation will provide practical support for certain research projects, such as projects that require access to non-public data. Researchers may not claim any support, approval, or special privileges from the Wikimedia Foundation unless they have a signed, written agreement with the Wikimedia Foundation that says they do.
>
> "Observational research generally does not require approval from anyone. Interventional research projects may require cooperation from the affected communities. Before beginning an interventional research project, we recommend disclosing it at a community forum, such as the local community's village pump. You should be prepared to engage in discussion with community members and, if necessary, to modify your research plan based on their feedback. Some communities require such disclosure and discussion." ('Does my project need approval?' [W122])

In summary, I found there to be no ethical issue that explicitly limited my proposed form of research.

Though my Wikipedia research methodology was to be non-interventional, this did not absolve me of a duty of care I still presumed towards those whose edits I may quote or whose activities I describe. Edits and talk page comments, where signed, are self-evidently the work of named editors. It is an editor's choice as to the extent to which they share any personal data. Some users will use their name as a username, but many prefer a non-identifying username be it for privacy or simply for use of a whimsical name. It is a nostrum of the site that editors are referred to by their usernames even if their identity is more widely known and drawing a connection between a user and their outside activities is generally considered unacceptable.

My methodology precluded contact with human actors within the research subject, so I was mindful where quoting usernames as to issues of implied critique. In discussion with my supervisors, given the scope and manner of the research, it was not judged necessary to make a formal ethics application but I have taken that apparent freedom as a reminder to be careful of how editors are treated in my writing. I considered anonymising the usernames but the source data was so messy

[11] Of the tools listed, I did investigate WikiXRay [W129] but this only results in a large SQL database that then needs further intervention. Large-scale quantitative tools, as listed on that page often hide the degree of small-scale triage that is needed to get an accurate picture of use of content transclusion.

and sizeable that, given the username (as opposed to actual identity) was a matter of public disclosure (anyone may look at Wikipedia), I judged it nugatory and more likely to add to possible errors in the data.

3.2.5 Possible Wikipedia Data Sources

Wikipedia is a large collaborative hypertext allowing editing and posting either as an identifiable logged-in user or anonymously. In the latter case, the user's Information Protocol (IP) address is logged in lieu of an username and some more advanced features are not enabled. As pages in Wikipedia are open to constant edit, an initial decision for each research task was to decide the form of dataset to use. There were a small range of options. Note that the order of listing does not imply a hierarchy or precedence amongst these choices:

- Live data (as at time of reading/sampling):
  - *Live Wikipedia website*. This uses the current online Wikipedia website. This is most useful if wanting to see the rendered page and understand the site as experienced by human users. It also allows exploratory investigation both across the corpus of articles and longitudinally on a temporal access, via edit histories. For security reasons, e.g. excluding bad actors or copyright violation, a small amount of otherwise valid material is not accessible via the web interface. Requests to access such 'hidden' data would jeopardise a non-interventional research approach
  - *Wikipedia Application Programming Interface (API)*. This allows access to a subset of Wikipedia data. It appears optimised, or designed, primarily for people wishing to (re-)use article content. The API's documentation on this is somewhat opaque, to say the least—i.e. it is hard to get a clear summary of what can *not* be called from the API. The result is the API does not provide access to all aspects of Wikipedia making it a sub-par source for open-ended research as 'unkown unknowns' are hard to triangulate. The API also has access rate limitations[12]. Higher-rate access can be achieved via pre-approval, though that would compromise a non-interventional research approach.
- Saved data:
  - *Local instance of the Wikipedia wiki(s)*. Using the larger, richer, formats from the Wikipedia data dumps it is possible to recreate a local instance of the live wiki website for a given language. The local site is essentially is a snapshot, fully working but with data only up to the time the source dump was saved, for the language specified. This gives the similar benefit of static data in not being subject to change, whilst allowing a richer and more powerful means of interrogating the data. Having such a

[12] Despite documentation [W112], there is a distinct lack of clarity as shown here: Wikimedia Developer Support forum thread, 'rate limit clarifications' [W111]

site also allows large and complex queries to be run locally, such as might otherwise not be allowed on the live Web site due to the impact on the back-end servers. It is not documented if hidden/redacted data is included in the dumps used to create a local instance of Wikipedia wiki(s).

- *Static data*. Wikipedia offers static 'dumps' of data per-wiki-language, normally on a monthly basis. These dumps vary in scope and format [W127] and are normally only accessible for a rolling 3-month window after first extraction for the source wiki. A useful aspect of this is the data is fixed and not subject to change via ongoing editing.

In all formats, there is a lack of clarity as to access to information hidden or redacted from the current version of a given page. This situation can arise due to issues such as users being banned or posting copyrighted information. Although the missing data is not a significant proportion of the readable website, the method of its removal normally breaks the overall hypertext[13] in unexpected ways, as links point to non-existent content. Not only does such arbitrary breakage harm the hypertext but it also makes it significantly more difficult to trace the relationship of editors and their edits, and thus decreases the quality of provenance.

Research task RQ1: use of transcluded content. For this, I needed to be able to discern those public articles (pages) that contain particular forms of Wikipedia mark-up in the article's source code. This specific mark-up is the clearest unambiguous indicator of use of transclusion, which is generally invisible in the rendered page unless an editor has chosen to indicate that fact. Wikipedia's mediawiki uses a form of generalised transclusion for its templating and server-side includes, thus it neglects to identify transclusion of content from other articles. The relevant articles needed to be detected, the markers extracted with contextual metadata, the data reviewed for edge cases before the true research could occur.

As pages in Wikipedia are open to constant edit, the presence or use of content transclusion is potentially ephemeral. Running a full local instance of the Wikipedia site would have involved significantly more set-up work and extraction work but without giving a clear indication of the answers to the research task. From testing, the API showed itself to be limited in scope and performance so of little value in context. For these reasons, I chose to work with a fixed snapshot of the site as this removed any ambiguity about all pages being sampled concurrently.

Research task RQ2: use of bots in sustaining the hypertext. Here, early tests showed the subject matter was not comprehensively described or documented within Wikipedia, offering no easily scoped dataset from which to form a meaningful snapshot of bot activity. The degree of disorganisation was such that it proved clearest, though not easy, to trace *all* bots from inception in 2003 through until the intended search cut off of July 2017. Clearly, static monthly data was not a viable option as it does not contain the longitudinal record, at least, not in easily accessible form (and past snapshots are only retained for circa three months). The API, was unsuitable as

[13] This would imply a mix of expediency and lack of understanding of the hypertextual medium.

an assemblage of the picture about bots would be difficult, given the incomplete state of the Wikipedia documentation and it being hard to query for target objects unknown before the fact.

This left a choice of the live site and a locally hosted copy of the site. Had a local site been used for the first task, it could have been re-used, but the data choice for that had already been taken[14] and from early exploratory testing, using the current web site proved the best option as I had the facility of regular, fast web access and no setup was needed.

The result is that although both tasks were focussed on Wikipedia as the studied hypertext, their dissimilar natures required I use a different data source and research approach in each case.

This grounded methodology was needed at a number of stages of the process:

- Understanding Wikipedia documentation as to what constituted the articles, such as read by an ordinary visitor to the Wikipedia site. An article is obvious as seen on screen, but how it is retrieved via the article source code is not immediately obvious (to a non-expert). This stage is discussed in RQ1.5, section 4.6.1.5.
- Analysing detected content transclusions, in the context of the rendered result. Rendered pages often include the output of a large number of 'templates'. The actions of the latter range from inserting a single correctly styled quote character to complex internal (to the template) transforms. As these transforms are effectively hidden, judgement must be applied as to whether the transclude seen reflects what might be expected from the source code markers detected.
- Assessing the possible causes of mediawiki source mark-up errors found in broken transclusion mark-up. After iterative tuning of the extraction scripts a small number of transclusion markers show errors: some of these inhibit correct page rendering. Whilst some examples are due to simple mistyping by the editor, others may be assessed as indicating a lack of clarity in the documentation.
- Assessing the context of Wikipedia content in which transcludes appear. In some cases, occurrence of transclusion markers was highly localised to particular subject areas.

Quantitative methods were applied to the extracted data. The primary task was to obtain and validate counts of markers extracted from the source XML by identifying appropriate mark-up. The counts allowed for detailed reporting of instances of different types of transclusion marker.

The differences between wikis in the rates of occurrence of markers, the complexity of differing sub-forms of transclusion effected, and proportionately low sample sizes of extracted markers precluded further effective analysis using more detailed statistical methods. This is discussed further in Chapter 4.

[14] The two pieces of research were not done concurrently, so the choice not to use a local instance had already been taken by the time I came to choose a data source for the second task

3.3 Qualitative Tasks

3.3.1 Transclusion Mechanisms Within Wikipedia

When studying the transclusion of article *content*, as discrete from more generalised forms of server-side includes, the construction and vernacular of Wikipedia's mediawiki software can be confusing. This is because both forms are termed 'transclusion'. Indeed, this is from documentation on 'Templates' which are essentially a form of macro for inserting known pieces of text into a page:

> "A template is a Wikipedia page created to be included in other pages. Templates usually contain repetitive material that might need to show up on any number of articles or pages. They are commonly used for boilerplate messages, standard warnings or notices, infoboxes, navigational boxes, and similar purposes.
>
> "The most common method of inclusion is called transclusion [sic], where the wikitext of the target page contains a reference to the template, using the `Template name` syntax. Another method is substitution, where the content of the template is copied into the wikitext of the target page, just once, when it is saved."
>
> ('Help: Template' [W55])

Templates can also be nested. Over time templates have been pressed into a wide range of use throughout Wikipedia, and are further extended by use of 'Modules' [W108] which contain Lua [W225] code. Both templates and modules are files within a wiki, but stored in their own namespaces (see Appendix L) discrete from normal articles.

In early discussion and review of this subject it was quickly apparent that those with greater training in, or exposure to, programming were more likely to see content transclusion and templates as the same think. Conversely, those with a humanities background were more likely to see the content (articles) and structural affordances (e.g. templates) as disjointed.

Neither view is incorrect, but the former blurs a useful boundary by emphasising a technological lens. To adapt Upton Sinclair's phrase[15]—it is difficult to get a person to see nuance when their discipline's perspective precludes their perceiving it.

A template may import a flag image, or match results, or a specially formatted description, but this is a different act to the considered use of all—or part—of another article. At the smallest of scope the two acts are similar, but at wider scope templated transclusion becomes boilerplate: text deliberately written for re-use.

[15] Original quote: "It is difficult to get a man to understand something when his salary depends upon his not understanding it". [319]

Ad hoc transclusion of article content, that is not necessarily written as boilerplate, is a more subtle act. It may occur in order to show provenance, or to cite directly or to avoid re-writing material in a way that changes its meaning. It is this looser form of transclusion that is being investigated here. Thus, transclusion of material via template and module mechanisms is ignored. By noting namespaces used in transclusion mark-up it is possible to weed out template and module use.

3.3.2 Data for Content Transclusion Analysis

3.3.2.1 Wikipedia Data Sources

Data. All data used for research tasks was sourced from Wikipedia. Source data, including the bzip2 dump files containing the actual eXtensible Markup Language (XML) data, were retained locally to allow for offline access. Although undocumented in Wikipedia itself, it is apparent that the dumps are only retained for a certain length of time[16] with newer dumps replacing older ones. For instance, by 2017 the original sample set of January 2016 dumps were no longer available. Therefore, retaining the original zip downloads provides an archive of the files no longer directly available from Wikipedia itself without a bespoke re-extraction of data from the main Wikipedia mediawiki database.

Source data and initial script-extracted working data were stored on a 2TB external disk allowing easy transfer between any of the computers. Worked data from the script output was stored in a cloud-replicated (Dropbox) section of the Mac's local hard drives which also have their own local Time Machine back-ups. Due to the size of the interim script-generated data (c.25GB) this was occasionally archived to a hard drive location. Selected key files were zip-compressed and then copied to a Mac's Dropbox for onward replication including Time Machine backup.

The latter was a pragmatic choice given that the initial working files could easily be re-generated from source. Back-up priority was given to processed data files as some of these represented significant amounts of review and processing work. The overall size of data stored on the data drive, including zipped and decompressed source files was c.325GB.

Wikipedia makes data available via two methods: the mediawiki API or as compressed dumps of XML data. Initial scoping study transclusion markers indicated the necessity of reviewing the mediawiki source 'text' of individual articles. Although the API allows persistent access to data as and when required, compared to using XML data it requires an additional phase of planning to ascertain which data needs to be interrogated and then drawn from the API. By comparison, once downloaded, the XML would contain all data needed for analysis. XML data is offered for download as 'bzip2' [74] compressed archives.

[16] From inspection of available content, this seems to be 6 months, plus the any extraction still in course of completion. For the larger language wikis, such as the English ('en') wiki, this can cover a period of days rather than hours.

XML dumps are offered in a range of depth of content. The richest sets offer pages with complete edit histories. A smaller data set offers a monthly snapshot of all pages[17] on a given date. As the Research Questions are not looking at transclusion from a longitudinal perspective, a snapshot provides a more concise dataset with less extraneous data to filter. As a result, a dataset from the dump data page described as 'Articles, templates, media/file descriptions, and primary meta-pages' was chosen as the primary source.

This dataset is available for each of the per-language wikis within Wikipedia, allowing for similar analysis methods to be applied to data from more than discrete (language) wiki. The per-wiki dataset is regenerated once a month for each sampled wiki and added to the publicly accessible set(s)[18].

On the Wikipedia page 'What's available for download' page [W127], it notes that some data is not available because it is private. This includes user data such as passwords, e-mail addresses, preferences, watchlists, etc. Likewise, deleted or suppressed content is not available for download; it may have contained spam, personally identifying information, copyright violations or other sensitive material.

Data dumps are offered in differing numbers/sizes of files depending on the source language wiki. A description of why these differences occur is given in the page 'Data dumps/FAQ' [W120].

3.3.2.2 Per-Language Wikipedia Sources

Wikipedia, though mainly known for its English-language site, actually supports discrete per-language wikis. As at December 2015 Wikipedia's 'List of Wikipedias [W103] listed 291 discrete per-language wiki, of which 280 were currently still active, as listed in the that page's 'List' section [W106]. Each language's wiki is allotted a two-letter (English: 'en') or three-letter (Cebuano: 'ceb') abbreviation which is used to address that wiki within the overall Wikipedia domain. Thus, the English-language wiki's home page is found at https://**en**.wikipedia.org/wiki/Main_-Page [W398]. These 'short codes' are also used elsewhere in this document to identify the source language for analysed material.

Although different language wiki employ 'interwiki' links between pages on the same subject in different languages, these are not simple translations. In many cases the non-English article for a topic may be longer, more informative—and differ in viewpoint—from the 'en' wiki's article. Articles on per-language history and geographical features (including town and cities) certainly fall into this category.

[17] Although the dump includes pages from outside the main namespace, these can easily be filtered out.

[18] The exact day of generation within the month may vary between wikis, presumably due to the effort involved in the process (this date difference is not formally explained).

3.3.2.3 Choice of Per-Language Sources

The English-language Wikipedia was a natural starting point for initial investigation of data structures though helpfully, every per-language wiki uses the same media wiki software—albeit with some localisation of terminology (see Appendices L, and M). Given RQ1 involved looking for mark-up, i.e. patterns, in the article source code this made possible analysis of wikis whose language, or indeed alphabet, I could not read. As a result, I chose to use 10 different languages, two with non-Roman alphabets, and covering both 'Western' and non-Western countries, as listed alphabetically in Table 3.1.

Despite per-wiki content differences on the same topic, as noted above, my analysis does not look at those difference but simply at the transclusion patterns within and between in the overall articles. At outset it was not assumed that all linguistic/cultural group would simply follow English-speaking patterns. It should also be noted that the English wiki, as the primary wiki within Wikipedia, has many contributors who are non first-language English speakers[19].

Language	Short code
Cebuano	ceb
Dutch	nl
English	en
German	de
French	fr
Italian	it
Japanese	ja
Russian	ru
Spanish	es
Swedish	sv

TABLE 3.1: Wikipedia languages selected for analysis

Of these, nine were in the top ten were in the top 10 wikis by article count as per the 'List of Wikipedias: Detailed list' section [W105], the outlier being Japanese (at rank number 13). Two wikis, Japanese and Russian, use a non-Roman alphabet. Three wikis, Cebuano, Japanese and Russian, lie outside the Western European and North American cultural area. By using this more diverse spread of sources, my intent was to investigate if there was a basic pattern of transclusional behaviour transcending language or culture (reviewed at Section 4.6.1.3).

The Cebuano wiki is also of interest because it is constructed almost entirely by a single automated 'bot' account called 'Lsjbot'[20] [W107]. The bot is operated by a Swedish Wikipedia user whose wife is a Cebuano speaker. In fact, 'LsjBot' has contributed significantly to the Swedish, Cebuano and Waray language wikis/footnoteBoth Cebuano and Waray are both languages whose use is almost exclusive to the Philippines..

[19] This difference in first language of 'en' contributors does not appear to have been formally tabulated but is readily apparent from perusal of back-of-house matter such as talk pages.

[20] Also see Wall Street Journal article 'For This Author, 10,000 Wikipedia Articles Is a Good Day's Work'[189]

The Wikipedia dataset used for the study at section 4.3 was drawn from the January 2016 data dumps, as being the most recent complete output at the time of sampling. Within this dataset, the English language wiki dump was dated 13 January 2016, whilst dumps of the other large sampled wikis were dated 11 January 2016.

Subsequently, datasets for the same wiki for January in 2017, 2018 and 2019 were also downloaded and archived for possible later longitudinal comparison. In the 2018 and 2019 downloads, the Chinese language ('zh') language was added to the dataset for an additional non-Western non-Roman alphabet resource. These additional samples were stored as Wikipedia either does not permanently archive monthly dumps or ceases to make them available for download. It is entirely likely that all past monthly dumps are not retained in the long-term for programmatic reasons of storage vs. likely subsequent use.

3.3.2.4 Data Dump Extraction

For 2016, bzip2-compressed XML files from the 'Articles, templates, media/file descriptions, and primary meta-pages' version of the data dumps[21] (as described in Section 3.3.2.1) were downloaded for the 10 discrete languages listed at Table 3.2.

Language	Short code	Zip Files	XML (GB)	Dump date
Cebuano	ceb	1	7.53	11-Jan-2016
Dutch	nl	4	6.16	11-Jan-2016
English	en	27	56.01	13-Jan-2016
French	fr	4	14.66	11-Jan-2016
German	de	4	10.03	11-Jan-2016
Italian	it	4	9.34	11-Jan-2016
Japanese	ja	4	9.35	11-Jan-2016
Russian	ru	4	16.20	11-Jan-2016
Spanish	es	4	10.24	11-Jan-2016
Swedish	sv	1	10.85	11-Jan-2016

TABLE 3.2: Retrieved January 2016 Wikipedia Data for dump type '*Articles, templates, media/file descriptions, and primary meta-pages*'(data size is for decompressed XML)

For provenance purposes and to allow re-use/review of the data by others data files were stored, per wiki, in discrete folders. This modular approach easily allows adding data for additional languages to the analysis set and for access to data at single-wiki scope. The source bzip2 files and their unpacked XML files were each stored in a separate folder tree and aliased back to the wiki folder-set. The segregation of source and output arrangement allowed for smaller back-up files for the extracted data.

The Python data-analysis scripts were constructed so as to allow a single file, all files from one or more wikis or all XML files to be traversed and data extracted. The scripts generated reports

[21] The Wikipedia source pages for the dump files has no semantic markers in the source HTML as would allow a URL with an anchor suffix to reference precisely where in the page the link to the dump file(s) is shown.

of progress, elapsed times and counts read from the command line the Mac's Terminal program. The scripts also generated a (per-language) saved run data file placed in the root folder of each discrete wiki's data output storage. The report file also includes a tab-delimited table of counts for easy transfer to other analysis contexts. Four scripts are used to extract data:

- Retrieve article, redirections, and transclude target marker data.
- Retrieve all transclude calls.
- Retrieve category data.
- Retrieve lists of discrete namespaces and article counts for each one

Apart from increasing the time to iterate any given source file(s), there is no reason the existing scripts may not be aggregated to a single script[22]. Up to this point everything may be easily reproduced: data is still available from Wikipedia[23] and the Python scripts allow complete reproduction of all the data extracts from the XML.

The text files generated were then tested via regular expression for errors such as transclusion of non-main-namespace objects (or cross-wiki inclusions) as well as literal mark-up errors. During this iterative development process, checking a number of Wikipedia articles still showed these errors remain in the current live article. It should be noted that mediawiki code is fairly permissive of variation in exact use of mark-up, reflecting the fact many users have little experience of such a technique.

Aside from the text files generated via the Python scripts, other data created are an Excel spreadsheet (counts, analyses and graphs) and a number of Tinderbox[24] data files. The latter hold research notes and were used for mapping of transclusion networks and patterns. Tinderbox was also used for analysis and coding of transclude networks in the 'en' wiki data. Lists of the 'en' wiki's transcluding and transcluded articles were imported and codified with observational metadata, such as the form of mark-up used at each side of the transclusion. Tinderbox scripting was then used to add Uniform Resource Locators (URLs) to the per-article notes to allow easy inspection of the article's edit page in a live context[25,26].

[22] Each script takes between 2-3 hours to run across all 11 datasets, using a desktop Mac. Consideration was given to initially parsing and re-writing the XML to exclude non-main namespace <page> elements, but as the scripts ran in reasonable time and only need to be run once (after development was complete), there seemed little gain in such intermediate extraction.

[23] But note dump access limitations described Section 3.3.2.1.

[24] Tinderbox [43] is a hypertext tool with strong support for incremental formalisation and emergent structure. The program stores its data (TBX) files as well-formed XML including sections of Rich Text Format (RTF) text encoded in Rich Text Format Directory (RTFD) format.

[25] By appending the Unique Identifier (UID) of a particular revision the article can be viewed as at the time of the dump as opposed to in current form at the date/time of access. Note, however, that this can fail if significant article renaming and re-direction has occurred as the redirects stop the desired page loading.

[26] It is also worth noting that as the two pieces of research were not done concurrently, the choice not to use a local instance had already been taken by the time I came to chose a data source for the second task.

3.3.3 Data for Bot Analysis

The longitudinal framing of the research requires use of a complete system so the live 'en' site was used. The complexity of this task (due to broken/incomplete Wikipedia records) precluded using other languages at this stage. The processes used to identify data for RQ2 are described in Chapter 5.

3.4 Summary

In this chapter I started by considering my scoping of knowledge within a hypertext context. I continued by expanding my overall research question into two discrete questions. Each of these each requiring a different data collection approach to the overall corpus of Wikipedia data.

Prior to looking at the data collection methods, I explained my overall methodology, including the non-interventional approach so as not to affect the behaviour of the subject being studied. I described a fluid mixed-methods approach that allowed me address issues arising from unfamiliar data sets. Given that I was not using a known standardised, methodology, I reported on the issues of ethics and possible harms to the research subjects, not assuming that the lack of a formalised methodological structure in any way freed me from normal ethical constraints.

The chapter then turned to choices of the data sources for each of the two RQs, as informed by my open methodology. Each question required a different approach to the source data and thus a different data collection technique, and thus was considered separately. RQ1 involved a fair amount of selection and pre-collection of data which I have described here. By comparison, for RQ2 only one language's wiki was used (English), as the work required close reading and interpretation of the source material as resolving a large number of broken or missing link. Also, RQ2 not could use the pattern extraction from source be used to identify terms of interest though this did mean no requirement to write data extraction scripts as were needed for RQ1

The two RQs are discussed this the succeeding chapters. RQ1 about use of transclusion in Wikipedia is reported in Chapter 4. Following that, Chapter 5 covers the work investigating the contribution of bots.

Chapter 4

Content Transclusion in Wikipedia

> ‘ *“Transclusion” is the term I use now, a word coined since the book was originally written. Transclusion will be a fundamental service of tomorrow’s literary computers, and a property of the documents they will supply. Transclusion means that part of a document may be in several places—in other documents beside the original—without actually being copied there. This book is concerned in large part with the power this new kind of writing—with transclusive quotation—will bring.*’
>
> Theodor H. Nelson ‘*Literary Machines*’ [268, (ed.93-1) preface, p.5]

> ‘*Because something is not what it is said to be, Ma’am, does not mean it is a fake. It may just have been wrongly attributed.*’
>
> Anthony Blunt: Alan Bennett’s play *A Question of Attribution* (1991)

Discussion of the selection of transclusion in Wikipedia as the subject matter for this Research Question is given at Section 3.1. The over-arching idea here is to measure the incidence of transclusion of article content within the ‘main’ namespace of a number of discrete language wikis within Wikipedia. Before turning to the formal RQ1 (see Section 4.4), I will give an overview of the concept of the research before first describing detail how transclusion is used in Wikipedia and how its use may be detected. This initial exposition will help explain the rationale for the decomposition of RQ1 in Section 4.4.

4.1 What Can Transclusion Use Tell Us About A Hypertext?

Transclusion allows one hypertext note to include content drawn from another node. In a Wikipedia context this means one article can include in its rendered web page content stored in another article. The most obvious affordance this offers is that of storing canonical information once

within the overall hypertext. In theory, the encyclopædia has a collection of discrete pages on topics, but inevitably there is some overlap in the scope of articles either due to common contingent information or due to the level at which an article covers its subject; thus high-level articles will likely include the more salient aspects of articles covering parts of that subject in greater detail. Ideally this means keys facts can be recorded and verified once, and then re-used more widely. This avoids duplication of data creation (and attendant referencing) as well as the scope for mistakes when (re-)entering information into the system. Such structures are of most use when they might change over time and are mentioned in many places.

Re-use of canonical data also avoids drift in meaning that can occur when the same basic facts are being described by several different authors. Even 'fixed' historical facts are open to re-interpretation or correction from new discoveries. Where the information is that such as regulation, which may either change at a statutory period or as needed, there may be a need for many articles to reflect this change directly or indirectly. A change in a key (reference) fact should be able to cascade though the hypertext with as little manual editing as possible. A different aspect of re-using (text) content in multiple contexts poses challenges of style and boundary coherence where different texts 'join'. These are likely to show if the editors have failed to consider the styling issues inherent in transclusive use.

For data at the most granular level, Wikipedia also offers the structured data stored within Wikidata. Using this ought to offer, even if not yet, semantic markers in articles to help software agents better traverse and interrogate the corpus (beyond keyword and pattern matching).

A high level of content transclusion detection would indicate the hypertext is being constructed and updated with long-term stability in mind, as the more a change in one fact requires manual editing intervention in many articles the greater the chance some such actions will be overlooked. In theory the 'many eyes' approach, i.e. with enough editors *some* editor will see and fix the issue [296, p.30], can avoid this but that is as much hope as fact. Considered use of transclusion makes the whole more robust to such degradation.

As discrete groups of editors may cover different subject areas or types of edit task (e.g. formatting), analysis of transclusion offers the prospects of seeing communities of practice. The breadth of these and similarity of practice will give an indication to the quality and coherence of training and documentation available. In examining wikis for more than one language, this holds the prospect of revealing differing per-wiki practices due to separation or perhaps cultural differences of style. Communities of practice may be both stylistic in terms of textual content and also use of transclusional patterns.

Detection of transclusion use is as useful to Wikipedia editors as it is for my research. Without suitable markers (warnings) of transclusion editors may misunderstand what should the added or editing in which article. In counterpoint, being able to easily see or detect transclusion should enable human—and bot—editors to find areas to improve transclusion use.

Conversely, a general absence of transclusion use may point to organisational problems in editorial practice that may hinder the long-term health of the hypertext. Thus in terms of organisational knowledge stored and maintained in a hypertext, the health of the whole hypertext is as important as any single article within it. This is because the knowledge therein resides not only in the sum of the individual articles but the manner of their interlinking from which the reader may find insights not explicitly recorded in an single node of the hypertext.

4.2 Types of Transclusion in Wikipedia

The origins of the general concept of transclusion has already been described in Section 2.9 (see also Appendix C). Wikipedia's notion of transclusion is complex and blends both the concept of a software include and Nelson's transcluded hypermedia content. The primary departure from Nelson's idea of transclusion is that the Wikipedia transclusion process does not hold a version state for the source media. Therefore, the transcluded media is the current version of that source at time of transclusion whereas in Nelson's 'Xanadu', the transcluded content would come from a specific edit of that source.

This difference is not a dogmatic departure from Nelson's method. Although there is no clear documentation on how the difference first arose, it likely reflects the fact that the design of the mediawiki[1] on which Wikipedia runs was informed by existing general Web behaviours at the time mediawiki forked as a discrete software offering (q.v. mediawiki's genesis.). The HTML `<img>` tag for inserting images in web pages uses this same un-version source method of transclusion seen in mediawiki.

Wikipedia's concept of transclusion is illustrated in Figure 4.1[2]:

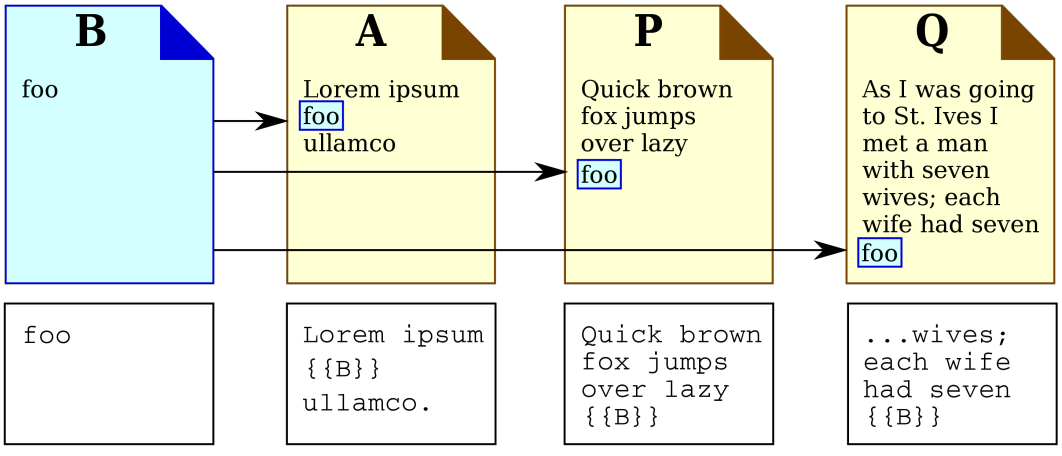

FIGURE 4.1: Wikipedia's conceptualisation of transclusion (Image source [W25]).

[1] The Wikipedia and mediawiki organisations seem to employ varying cases for the name of 'mediawiki' software, though all-lowercase would appear to be preferred. I have used that convention except for capitalisation at the beginning of sentences or in quoted contexts.

[2] Image as seen in Wikipedia's 'Transclusion/How Transclusion Works' [W57].

Wikipedia's use of transclusion is most easily understood by looking at the mark-up used to employ the process. Thus next Section of this Chapter will explain in more detail how transclusion is effected in mediawiki. I will discuss the mark-up and the varying effects this has on the process of transclusion. For instance, the type of tagging used can control the amount of a source article that is transcluded or it may rely on scoping mark-up inserted into the code of the source document. It is also apparent that various parallel initiatives, some seemingly now moribund, mean that there are duplicate mechanisms (as illustrated in Figure 4.2).

4.3 Mediawiki Transclusion Mark-up

Wikipedia's mediawiki software offers a surprising number of variants for transclusion markers, which hint at incremental formalisation over time. The canonical mark-up for transclusion of the page *somepage* is `{{somepage}}`, but within this method, the following basic variations occur [W691]:

- `{{somepage}}` transcludes the page *somepage* from the `Template` [sic] namespace.
- `{{:somepage}}` transcludes the page *somepage* from the main namespace, i.e. an article.
- `{{namespace:somepage}}` transcludes the page *somepage* from the given `namespace`.

Note that the first, unmodified, method indicates a template-based transclusion. Mediawiki makes heavy use of transclusion templates for a variety of macro work. Thus instances of the {{somepage}} code form occur in most articles' source and certainly far more frequently than for non-template-based transclusion of article contents. The variable aspects of templates are generally serviced by code 'modules' in the wiki's 'Module' namespace. Both by evidence of general use and in documentation, Wikipedia appears to make no effort to differentiate deliberate transclusion of (article) content, as being studied here, from more general template-based macro-type work.

Of the three transclusion forms listed above, only the second is of direct interest to the analysis of direct content transclusion. As the Research Questions refer to direct transclusion of article content in the main namespace of a given wiki, the presence of the string {{: in a page's mediawiki source gives an unambiguous indication of the presence of a transclusion of article information[3]. The disambiguation task is made more complex by the fact that the screen title of an article can include a colon. Indeed, this appears commonly in pages listing episodic media, e.g. in the general form 'showname: seasons 1-4'[4]. Although forward slashes in URLs originally denoted sub-pages, this character too is allowed in an article name (albeit infrequently so).

[3] Apart from literal source code errors by the article's editor, noting editors have no training and the source is manually typed (bots as a notable exception).

[4] For example: 'Songs of Anarchy: Music from Sons of Anarchy Seasons 1-4' [W141].

Various tag forms based on the above interact to form a confusing and incompletely documented set of alternative transclusion pathways as shown at Figure 4.2. The purpose of the diagram is to illustrate the complexity of Wikipedia's transclusion system.

Not shown in Figure 4.2 is the fact the transclusion process can be nested inside the opaque wrapper of a Wikipedia template. It is also possible to nest transcludes within one another although such behaviour is not encouraged or observed in analysis of the English language Wikipedia. Discussion of such an approach has not been seen in talk pages (though might yet exist).

An example of self-transclusional effect within a template is discussed in section 4.6.1.4[5], in relation to limiting output of input data via internal use of `<onlyinclude>`.

Nelson's concept of transclusion was to link source and referring passages in different documents and he makes no comment on the scope of either document outwith the area of transclusion, though it is to be assumed that only the linked passage is transcluded. Mediawiki does not support visible identification of transcluded data, although optional mark-up is available to help towards this.

(Full) Transclusion. A {{: string in the calling article will transclude the entire target article content unless further code in the transcluded page modified the scope of returned content. Thus a transclusion call alone is insufficient to adjudge whether a transclusion is full or partial. It should be noted that when a 'full' transclude is employed only the content of the target page is transcluded, i.e. the header, footer and sidebar are omitted.

Partial Transclusion. An alternative 'partial' [W693] transclude approach is offered, whereby only a portion of the target page's content is transcluded. This is useful because unless an article is written with transclusion in mind it may likely contain more information than is desirable in the transcluding context (either with too much data, or simply not expressed suitably for the alternative context of presentation).

Partial transclusion is effected using mark-up within the source of the *transcluded* page. That allows arbitrarily delimited part(s) of the article, i.e. ad hoc selections, to be returned as the transcluded data; this can include multiple non-contiguous sections of content. The process employs pairs of one or more of the following tags: `<noinclude>`, `<onlyinclude>` and `<includeonly>`. Their effect on scope of transclusion is described in the table below:

Wikitext	**What is rendered here (source page)**	**What is transcluded there (destination page)**
`<noinclude>text1</noinclude>text2`	text1 text2	text2
`<onlyinclude>text1</onlyinclude>text2`	text1 text2	text1
`<includeonly>text1</includeonly>text2`	text2	text1 text2

TABLE 4.1: Partial Include Options (for target page)[6]

[5] See section on 'Hidden transclusion effects within episode listings'.

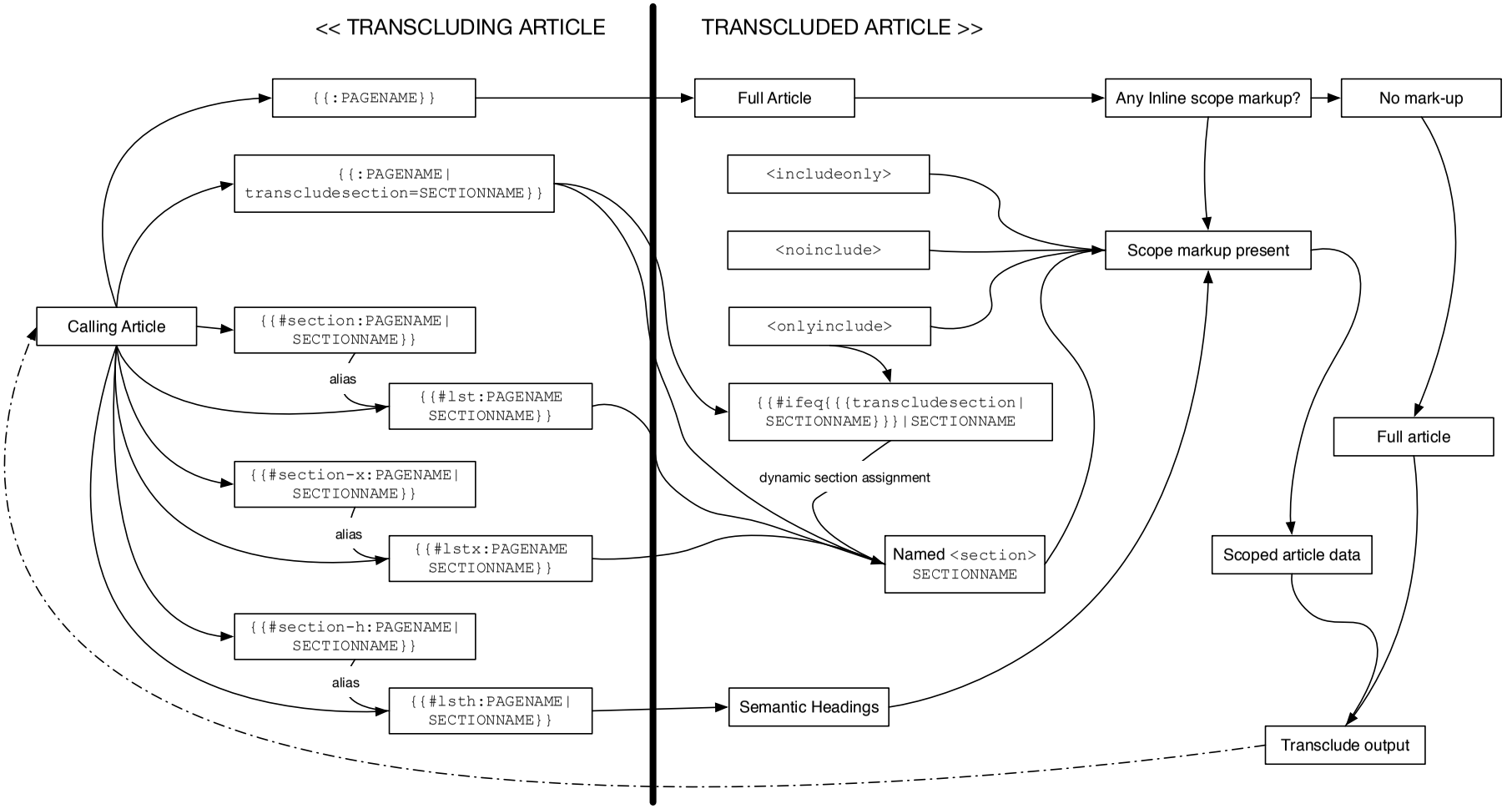

FIGURE 4.2: Transclusion pathways in Wikipedia
(excluding the Wikidata pathway, described in Section 4.3.4).

Selective Transclusion. A more nuanced partial transclusion is implemented by the 'labeled section method'[7] (LST) [W52]. This requires the target article either having appropriate parts of the source enclosed in mark-up denoting a 'section' or in some cases simply using semantic headings. Confusingly, the initial manner of using this method involved ad-hoc sectional mark-up. The calling page addressed a section transclude with code in the form:

```
{{:pagename|transcludesection=
```

whilst in the transcluded page this mark-up is added:

```
<onlyinclude>{{#ifeq:{{{transcludesection|SECTIONNAME}}}|SECTIONNAME|
...transcluded content...
}}</onlyinclude>
```

The target section name is given as the input parameter to the mark-up macro at both ends of the transclude. This method allows transclusion of sections of content not reliant on explicit semantic mark-up, such as headings.

The LST method was subsequently[8] enhanced by a series of mediawiki code functions allowing editors to mark up a `<section />` of an article and easily target it for transclusion. Thus a part of a page, to be transcluded as section 'chapter 1', would be marked `<section begin=chapter1 />this is a chapter<section end=chapter1 />`.

As well as the original `transcludesection=` mark-up, the newer sectional use offered a set of `#lst` mark-up (see Table 4.2 below). As some Wikipedians adjudged the 'lst' acronym name too obscure, more readable aliases were subsequently added and offered. These aliases can also be localised [W117], although there is no full listing of existing localisations or even cross-linking of locations where they might be recorded. The original `transcludesection=` calling method can target either an `#ifeq{{{}}}` macro or an explicitly coded `<section />`. The latter is much simpler as the `#ifeq` form also relies on it being enclosed with an `<onlyinclude>` tag pair.

Function	**Alias**	**Three arguments: 1) pagename, 2) the source section or a range-start,) 3) source-section range-end**
#lsth	#section-h	Heading/section-title oriented. Ignores `<section />` tags if any. Range contents included but not range headings. If no section heading given, maps to section 0 (i.e. page's lead). Ignores `<onlyinclude>` tags.
#lst	#section	`<section/>` tags required. (tags honoured in source.) Range begins at first heading ends at tag. No range given is normal transclusion.
#lstx	#section-x	With `<section/>` tags only. One section only. #section-x is the complement of #section.

TABLE 4.2: Labelled Section options[9]

[6] This table is copied from the 'Markup' [W692] section of the Wikipedia documentation of transclusion.
[7] The American English spelling with one 'l' is used.
[8] There is no clear timeline, in public documentation, as to when the new method was added.

The partial method for Transclusion first appears in documentation in April 2008 [W684]. Selective transclusion was first mentioned in July 2011 [W685]. It was not until February 2014 [W686] that explicit cross-reference was made within the 'Wikipedia:Transclusion' article to the LST method. A Help page for LST was first added in May 2103 [W51], though the original documentation for the mediawiki Extension dates back to 2006 [W115]. Mediawiki use of the `<section />` (non-HTML) tag thus pre-dates HTML5[10].

4.3.1 Tags Marking Transclusion

Incremental improvements to mediawiki's codebase have resulted in multiple forms of transclusion mark-up. The method termed LST is a recent addition and allows partial transclusion of specified 'sections' of a called page. A 'section' is either text following a defined heading or ad hoc ranges of text as defined by mark-up in the called page using the `<section>` tag[11]. For transcluding articles there are five pertinent source code tags, some with aliases[12]:

- `{{:somepage}}`. Marker: '`{{:`'. Denotes full article transclusion from `somepage` unless modified by scope-resisting mark-up within the transcluded article.
- `{{:pagename|transcludesection=section_name}}`.
  Marker: '`|transcludesection=`'. Partial transclusion via '`<section>`' tags or ad hoc section definition in target article.
- `{{#lst:pagename|section_name}}`. Marker: '`#lst`'. LST partial transclude. Only `section_name` is transcluded from `pagename`. Aliased as `#section:`.
- `{{#lstx:pagename|section_name}}`. Marker: '`#lstx`'.
  This is an LST exclude. All of `pagename` is transcluded except `section_name`. Aliased as `#section-x:`.
- `{{#lsth:pagename|heading_name}}`. Marker: '`#lsth`'.
  This LST targets headings. Only the content below `heading_name` up to the next heading of same depth is transcluded from `pagename`. Aliased as `#section-h:`.

[9] This table is copied from the 'Function names' section of the Wikipedia [W53].

[10] HTML5 became a formal W3C recommendation as of 28 October 2014 [385]. Note: initial drafts of the HTML5 specification date back to early 2008 [384].

[11] See: `https://www.mediawiki.org/wiki/Extension_talk:Labeled_Section_Transclusion` [W23]. The introduction date is not documented. The `<section>` tag is a Wikipedia innovation that predates the HTML 5 `<section>` tag and there is no functional connection between the two same-named tags, although the Mediawiki tag's mandatory attributes make disambiguation easier. Wikipedia's documentation is ambiguous as to whether '`#lst`' and'`|transcludesection=`' are or are not (by design intent) full functional equivalents.

[12] LST Aliases were added to make the mark-up's intent less confusing for inexperienced editors. Wikis may optionally localise for their language.

4.3.2 Tags Controlling Scope of Transclusion

Unless an LST or a '`transcludesection`' call is used in the transcluding mark-up, it is otherwise not possible to tell whether the article is wholly or partially transcluded. For transcluded articles there are 5 pertinent source code tags:

- `<noinclude></noinclude>`. Marks ad hoc sections of the called article which are not to be transcluded (but is still rendered for the article itself).
- `<onlyinclude></onlyinclude>`. Marks ad hoc sections of the called article which are the only parts of an article to transclude (and rendered for the article itself).
- `<includeonly></includeonly>`. Marks ad hoc sections of the called article which are not rendered *except when transcluded*.
- `<section begin="section_name">`
  `<section end="section_name">`. Marks ad hoc sections of the called article transcluded by '`transcludesection`' calls or via LST.
- `<onlyinclude>{{#ifeq:{{{transcludesection|`
  `SECTIONNAME}}}|SECTIONNAME| }}</onlyinclude>`.
  Marks an ad hoc section of the called article to be treated, dynamically, as a named transcludable section by 'transcludesection' calls (bullet #2 in list above). The `#ifeq` method predates LST and the `<section>` tag.

These tags can be used in (nested) combination nested. Nested use is described in Table 4.7.

4.3.3 Visible Indication of Transclusion

All editors, both human and automated, are allowed to use transclusion although it is not self-annotating. It is thus an editor's optional task. English Wikipedia documents allow template-based mark-up to indicate whole[13] or partial[14] transclusion, including a link to the source article.

4.3.4 Transclusion from WikiData

The 'Phase 2' of the Wikidata[15] project was implemented for all Wikipedia languages in 2013[16], via the `{{#property:P_number}}` mark-up. However, the method is not yet documented in Wikipedia nor publicly recommended. This subject is discussed further in Section 4.6.1.6.

[13] See: 'Template:Transcluding article' [W192].
[14] See: 'Template:Transcluded section' [W191].
[15] See: 'Wikidata.org' [W390].
[16] See: 'Wikidata blog: Wikidata all around the world' [W391].

4.3.5 Substitution

Mediawiki also supports substitution [W680], which equates to a once-only transclusion where the transcluded content is used to replacing the substitution code[17]. The substitution method is described as a side case to transclusional use but is not pertinent for this study and is mentioned here for completeness only.

4.4 The Research Question - RQ1

Having explained the nature of Wikipedia transclusion, I now turn to RQ1, which defines my first major section of research for this thesis. RQ1 was defined in Section 3.1 and the sub-questions were framed to draw out different issues arising from the use—or not—of transclusional data, well as possible patterns of use and different employment in the various discrete language wikis sampled.:

- RQ1. *Does Wikipedia show signs of editors making deliberate use of content transclusion?* This was split into six sub-questions:
    - RQ1.1: *Does Wikipedia show evidence use of deliberate use of transcluded article content?*. this is the main indicator of (pre wiki) incidence of content transclusion.
    - RQ1.2: *Is use of transclusion only found within discrete areas within Wikipedia?*. Within the main findings, can it be shown that particular (subject) areas show greater use of transclusion.
    - RQ1.3: *Does use of transclusion vary noticeably between different language versions of Wikipedia?*. As more than one discrete language wiki is being sampled, are there differences between the wikis, that might indicate different countries/language use transclusion in differing ways.
    - RQ1.4: *Does use of transclusion content show distinct patterns of transclusion?*. The presence of definable patterns can, if found, help with either the detection of other repetitions of the pattern or to help locate parts of the wiki where the same pattern might usefully be applied.
    - RQ1.5: *Can an editor easily identify and trace transclusion?*. As it is generally human editors who implement transclusion, does the wider body of editors know or understand the affordances of transclusion or how to use it in their own editing.
    - RQ1.6: *How does Wikipedia appear to view use of transclusion?*. Is it possible to gain a sense of whether, regardless of the detail of use, editors feel positively about the capabilities of transclusion.

[17] After this action the transcluded content's code is now a literal inclusion within the calling page and further transclusion no longer occurs, nor does Wikipedia record the previous act of substitution and implied linkage: the reference to the transclusion source is lost post substitution. The mark-up employed for substitution takes the form: `{{subst:pagename}}`.

The chapter continues by describing the initial investigation of the XML archives downloaded from Wikipedia.

4.5 Initial Investigation

Inspection of the downloaded XML files (see Section 3.3.2.4) showed that they each contained a variable amount of <page> elements (articles) from a variety of wiki namespaces. The dumps include no manifest or descriptive data. Some general information as to structure and content of the <page> items gleaned from Mediawiki.org's Help:Export page [W47]. Otherwise structure had to be reverse engineered for inspection of the downloaded XML. For non-English wikis (most) namespace names were localised. In line with the incomplete state of documentation seen when investigating transclusion, Wikipedia appears to have no look-up table of per-language namespace customisations.

As filtering on namespace is necessary to isolate content articles, it was necessary to research and compile the tables of XML namespace numbering and name localisation (Appendix L) and terminology localisation (Appendix M) using information from a variety of sources. Documentation pages in Wikipedia, Help[18] and mediawiki namespaces plus the XML <namespace> element of the XML dump files yielded information. No single source was complete and anomalies exist between the sources as to localised name strings. The non-namespace keyword for 'disambiguation'—a common meta-task in content—also had to be sourced by comparing different language wikis' content, both live page source and the XML dumps.

To give initial reference counts for the number of 'live' main namespace articles, estimated total active article counts were taken from the 'List of Wikipedias'[19]. These counts helped address the fact that the dumps include an unspecified amount of pages from other namespaces, with no metadata as to the degree of mix.

The row sort order in Table 4.3 is on the count in 'Articles' column, whose figures should equate roughly to content pages in the main namespace. The 'Depth' data is described in the article 'Wikipedia article depth': [W128] as '*a rough indicator of a Wikipedia's quality, showing how frequently its articles are updated. However, it does not judge the academic quality of articles.*' whereby a higher rating implies better 'quality' resource although this is not regarded as an absolute measure of quality. The low depth rating for the Cebuano ('ceb') language is consistent with the fact that despite its high article count, it is constructed almost entirely by a single automated 'bot' account called 'Lsjbot (as previously described in Section 3.3.2.3).

[18] For instance, the English Wikipedia's Help article on renaming namespaces [W118] and Mediawiki.org web domains as well as the XML headers in the XML dumps. Similar pages exist for some other languages, but coverage is incomplete.

[19] For Wikipedia: 'List of Wikipedias', section 'Detailed list' [W104]. As these Wikipedia reports are not cached, data for Table 4.3 were retrieved from the Internet Archive's 'Wikipedia (10 January 2016)' [W105]. This was used as archive versions of these listing pages are not available

Languages	Wiki	Articles	Total	Admins	Users	Active Users	Depth
English	en	5,052,329	38,205,583	1,327	27,197,820	114,783	910.61
Swedish	sv	2,251,084	5,480,784	74	473,197	2,438	8.13
German	de	1,895,152	5,409,315	245	2,328,136	18,773	98.56
Dutch	nl	1,850,537	3,524,909	51	740,979	3,996	10.58
Cebuano	ceb	1,794,820	3,704,746	5	24,704	96	2.21
French	fr	1,713,373	7,899,714	165	2,417,196	15,375	205.36
Russian	ru	1,280,921	4,658,001	90	1,825,472	10,211	131.24
Italian	it	1,246,229	4,293,398	104	1,312,947	7,775	115.19
Spanish	es	1,224,703	5,344,187	76	4,058,125	14,243	198.25
Japanese	ja	998,670	2,871,196	49	1,049,941	11,527	72.47

TABLE 4.3: Per-language website data via Wikipedia 'List of Wikipedias' page, rows ordered by count of 'Articles'.

Initial comparison of grep (grep)-based specimen counts with the overall article counts (in Table 4.3) still indicated an apparent over-population of main namespace articles, most likely caused by re-direction stubs[20]. Inspection of XML revealed use of a `<redirect>` tag that is unique to redirection article stubs. The tag thus acts as a unique type identifier for such articles within the main namespace, allowing for appropriate filtering in the data extraction scripts.

In all languages, the XML in the downloaded dump files was checked and found to use consistent mapping of (localised) namespace names[21] to numbers used for the `<ns>` sub-element values of all `<page>` elements (i.e. per-article data); main namespace articles take value of 0 (zero).

The first discrepancy is shown in Table 4.4. The count of `<page>` items extracted from XML dump files is in each case lower the total page count gives in the 'List of Wikipedias'—the 'Total' in Table 4.3. In all cases the XML dumps yield a lower total item count. The difference lies in the pages in some of the other, non-article namespaces, such as the 'User talk' (namespace #3). This difference was validated by comparing dumps from the full 'pages-meta-current' series with the files, covering the same record UID range, from 'pages-articles' used for the research. This confirmed the missing content was not pertinent to the planned study. A notable point is the significantly degree of this additional content in the 'en' wiki compared to the other samples.

The second discrepancy is between the listing of 'Articles' in Table 4.3 and the count of extracted `<page>` items in namespace(0), as in the 'All ns(0)' column in 4.5. The latter appears to show a significant over-count again the earlier totals. Further inspection showed that XML namespace(0) contains a mix of article and re-redirection stubs. Wikipedia's What is an article? [W705] notes that disambiguation and re-direct pages are '*Pages in mainspace that are not usually considered*

[20] Redirection stubs are non-loading pages used to automatically redirect incoming page requests and serve the directed-to articles from their true URL. These redirections arise as Wikipedia content matures and is re-factored over time. Wikipedia does not automatically make retroactive changes to links in article source code when article names are changed. Although later edits my correct links to old titles, wiki style is to retain all edits and add re-direction as necessary.

[21] See Appendices L & M for addition detail on localisation details uncovered during this process)

Languages	Wiki	'Total' Pages	All XML Pages
English	en	38,205,583	16,236,105
Swedish	sv	5,480,784	4,508,011
German	de	5,409,315	3,743,044
Dutch	nl	3,524,909	2,684,231
Cebuano	ceb	3,704,746	3,113,297
French	fr	7,899,714	3,115,027
Russian	ru	4,658,001	3,526,962
Italian	it	4,293,398	2,567,431
Spanish	es	5,344,187	3,230,662
Japanese	ja	2,871,196	2,016,021

TABLE 4.4: Per-language data, overall 'Total' vs. <page> items extracted from XML dumps. Rows sorted as per Table 4.3.

articles-proper'. By triaging the <page> items in namespace(0) into actual articles as discrete from redirect-stubs, the true article counts align. There is small divergence owing to slightly different sample dates, with Wikisource remaining an outlier.

Languages	Wiki	Articles	All ns(0)	Articles	Redirects	Re-dir %
English	en	5,052,329	12,188,486	5,055,811	7,132,675	58.52%
Swedish	sv	2,251,084	4,140,672	2,533,120	1,607,552	38.82%
German	de	1,895,152	3,200,021	1,895,965	1,304,056	40.75%
Dutch	nl	1,850,537	2,516,924	1,850,771	666,153	26.47%
Cebuano	ceb	1,794,820	2,963,362	1,811,648	1,151,714	38.87%
French	fr	1,713,373	3,115,027	1,713,868	1,401,159	44.98%
Russian	ru	1,280,921	2,807,922	1,281,320	1,526,602	54.37%
Italian	it	1,246,229	1,873,355	1,246,493	626,862	33.46%
Spanish	es	1,224,703	2,807,709	1,184,099	1,623,610	57.83%
Japanese	ja	998,670	1,602,047	998,908	603,139	37.65%

TABLE 4.5: Per-language data, total article pages from 'List of Wikipedias' vs. namespace(0) <page> items from XML Dumps, rows sorted as per Table 4.3.

Of note is the high incidence of redirection stubs in the 'en' wiki. Much of this reflects two cycles of improvement. Originally, article names were in 'CamelCase' style with no spaces used. The first large-scale use of redirection arose when naming conventions relaxed to allow spaces via underscores, thus 'CamelCase' could become 'Camel_Case', needing a redirect from the former to the latter. The second change was the allowance of dropping underscores in favour of true spaces, thus 'Camel_Case' could become 'Camel Case' requiring a further set of redirects. Combined with actual, non-stylistic name changes, this has resulted in a large number of redirects in the 'en' wiki (which is also the oldest wiki in Wikipedia). Whilst the re-directs function well, generally unseen by those reading the wiki pages, this does represent a burgeoning technical debt which will only grow in size with time.

4.5.1 Transclude Marker Extraction

Using the findings of the initial data interrogation, scripts were developed in Python 3.4.4 to extract relevant transclude marker strings (see Section 4.3.2) for data processing. The Python scripts employed the `lxml` module's `etree.iterparse` function that allows iterative object parsing of large (multi-GB) XML data files without excessive memory use.

As test counts of articles with transclude markers were low compared to overall article counts, it was feasible to save extracted data into sets of UTF-8 plaintext files matching the number of their XML source. These per-source-file extracts allowed for easy review during fine-tuning of the scripts and were easily aggregated in to single composite text files for later detailed analysis. As article mark-up is human-written code, script parsing errors were to be expected and visual inspection of results was desired to help gauge whether editors understood the mark-up required for their task.

The section-based transclude methods made it necessary to collect instances of `<section />` mark-up[22]. The category assignments for each article were also recorded to assist analysis of patterns of use, if any. Overall, 19 discrete listings were extracted for each XML file, with additional counts, giving the following extracted dataset:

- Count of all pages (`<page>` elements) within the XML data.
- Within the main namespace for each XML file:
  - Count and listing of titles of all main namespace articles (pages).
  - Count and listing of titles of all re-direct articles, with target of re-direct.
  - Count and listing of titles of all non-re-direct articles, i.e. 'live' articles.
- Within live articles:
  - Count and listing of titles of all occurrences of `<section />` use. The listing includes article title, count of in-article instances of the tag plus separate counts of 'begin' and 'end' markers (which should be equal for normal use and free use human error) and the actual instances of the tag.
  - Count and listing of all articles with categories assigned. The listing includes article title and each discrete category applied to the article.
  - Count and listing of all articles using any form of transclude. Count of articles and aggregate of discrete occurrences of transclude mark-up. The listing records each transcluding article and each discrete transclusion controlling mark-up instance.
- Transclusion mark-up in live articles calling transcludes. For each of the following a count was recorded of discrete articles using transclude and a separate count of each discrete

[22] Any defined section is a possible section transclusion target, as are semantic headings (H1, H2, etc.).

instance of transclude mark-up. The listing records each discrete use of the transclude mark-up within the article (including the target article name):

- `{{: -- i.e. 'basic' transclude.`
- `{{:pagename|transcludesection=targetsection`
- `#lst`
- `#lsth`
- `#lstx`
- `#section`
- `#section-h`
- `#section-x`

- Transclusion mark-up in transcluded live articles. For each of the following tags a count was recorded of discrete articles using transclude and a separate count of each discrete instance of transclude mark-up. Listing includes article title, count of in-article instances of the tag plus separate counts of 'begin' and 'end' markers (which should be equal for normal use) and the actual instances of the tags.
  - `<includeonly></includeonly>`
  - `<onlyinclude></onlyinclude>`
  - `<noinclude></noinclude>`
  - `<onlyinclude>{{#ifeq:{{{transcludesection|input1}}}|input...`
    `...</onlyinclude>`[23]

After initial review of script results, the same scripts were used to investigate the Template namespace[24].

4.6 Evidence of Content Transclusion

[23] In analysis, the `#ifeq{{{}}}` mark-up was always encountered 'paired' with `{{:pagename|transcludesection=targetsection` mark-up in the transcluding article. As such, this effective double-marking has been allowed for in analysis counts.

[24] The 'en' wiki sampled has 645,865 templates which were ignored by the initial analysis scripts. The same extraction techniques as used in the main analysis were used, although further inspection of template code methods would be warranted to ascertain if template 'code' might differ from the known manner of placing the content transclusion markers in content articles.

4.6.1 Results

4.6.1.1 Evidence of Deliberate Use of Content Transclusion

Using data extracted from ten different language Wikipedias definite evidence was found of transclusion mark-up (as described in section 3.3.2.4). However, this activity occurs at extremely low levels.

The data in Figure 4.3 is shown in percentage terms to give some degree of normalisation, reflecting the varying size of article counts across the languages sampled (see also: Table 4.3 for language codes and article counts). Figure 4.3 below shows the percentage of content articles with transcluding code, i.e. those pulling in content from other articles:

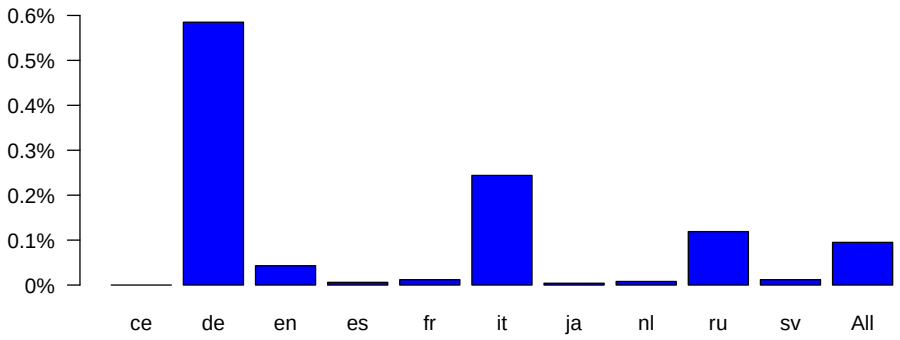

FIGURE 4.3: Percentage Occurrence of Articles with Any Transclusion Mark-up (per source language plus as a global average of all sampled languages; main namespace articles only)

At best, it is less than 1.0% presence, with 'de' reaching 0.58% after which only 'it' exceeds 0.2%. Although largest by sample count, levels in 'en' are only 0.02% (2 in 10,000). The 'ceb' languages shows a complete absence of includes.

The complete absence of transclusion in 'ceb' language Wikipedia, the second largest by article count, is likely explained by the fact that a very large proportion of the articles are bot-created stubs. These are almost entirely the work of a single a bot account 'Lsjbot'[25] that generally creates stub articles on non-controversial subjects such as plants and animals[26]. Transclusional activity would require additional complexity in the bot design and is more pertinent to use in more complete content articles. 'Lsjbot' is also active in article creation in the Swedish 'sv' Wikipedia, which also shows low levels of transclusion activity.

Articles containing transcluded content, i.e. articles holding content transcluded into articles charted in Figure 4.3 above are charted at Figure 4.4. The data is tabulated per-source language

[25] see Section 4.5

[26] For example, an article on the insect *Xanthopimpla punctate*[W14]

wiki, discretely for each of the 3 main transclusion scope-limited tags.and as occurrences of each of the 3 main scope-modifying transclusion tags. The percentage values represent each discrete article containing one (or more instances) of the tabulated tag type as a percentage of main namespace articles

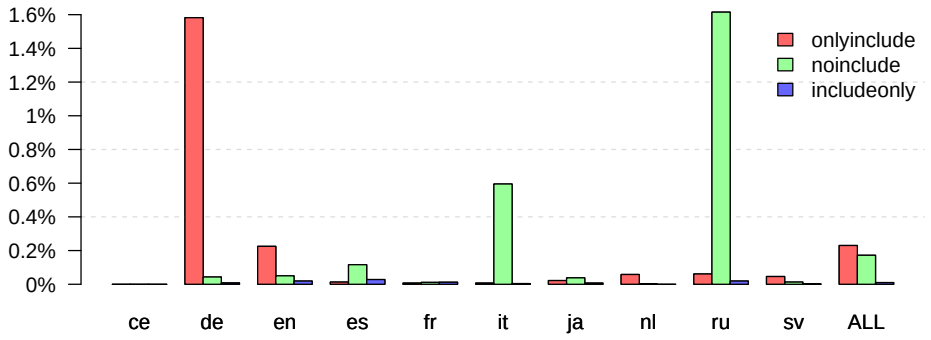

FIGURE 4.4: Percentage Occurrence of Articles with Specific Transclusion Mark-up Tags (per source language plus as a global average of all sampled languages; main namespace articles only)

LST markers (see section 4.3) were omitted in Figure 4.4 due to their much lower occurrence levels. These differences are explored, further below (see Table 4.6), in the context of the data for the English 'en' Wikipedia. Differences in per-language use of transclusion mark-up is discussed in Section 4.6.1.3

The greater occurrence of transclusion-scoping makers is consistent with a single article transcluding more than one other article. There are 8 discrete mark-up forms to call a transclude, noting that the '#section' styled forms are simply aliases[27] for the '#lst' based equivalents. However, few Wikipedia editors may be aware of that distinction given the fragmented nature of the documentation. The breakdown of the four main forms categories of transclusion mark-up for just the 'en' Wikipedia is shown in Figure 4.5.

Within the 'en' Wikipedia sampled, there are 2,184 articles which transclude 7,890 discrete articles' data. Source code of all 10,074 articles was reviewed to validate the use of transclusion code and observe the style of use. This stage of work was used to spot errors such as transclusion of non-existent pages.

Filtering transcludes of non-main-namespace articles, and errors such as calls to non-existent pages, the total of valid main-namespace transcludes is 2,127. Of these, 917 (43.14%) articles transclude more than one discrete transclude target, including 7 articles that use more than one

[27] 'Internally [sic], the parser functions all use the #lst prefix, for consistency with the name of the extension. Since this acronym may be confusing to non-developers, readable English variants have been introduced, so the functions can currently be called from either name.' (q.v.) [W53]

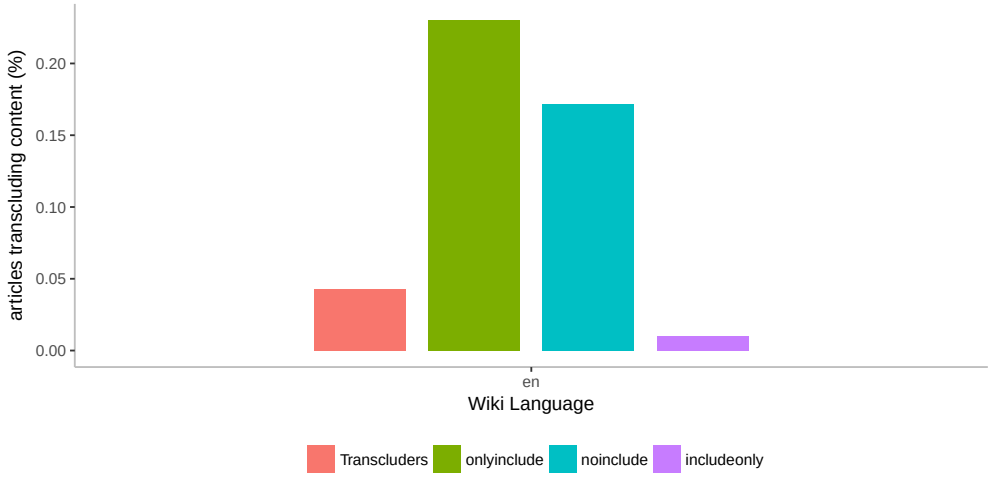

FIGURE 4.5: Occurrence of Transclude mark-up in English Wikipedia (as % of all main namespace articles)

form of transclusion mark-up. Occurrence of the 8 transclusion calling methods are used as shown below:

Mark-up	**Occurrence**	**% of Transcluding Articles**[28]
`{{: (i.e. whole article)`	1,978	92.73%
`transcludesection`	29	1.36%
`#lst`	73	3.42%
`#lsth`	3	0.14%
`#section-h`	0	0%
`#lstx`	0	0%
`#section-x`	0	0%

TABLE 4.6: Use of different transclusion-invoking code forms in 'en' Wikipedia

From Table 4.6 it would appear that whole article transclusion is the predominant usage. But, because transclusion is configurable in both transcluding *and* transcluded article, it was necessary also to consider scoping code in the transcluded context.

The above 2,127 pages transcluded a total of 8,243 articles representing 7,888 discrete main-namespace articles. Analysing the latter, the following degree-of-transclusion mark-up was found:

The Table 4.6 shows that transclusion of entire articles is actually extremely rare. Partial transclusion is the overriding form of use, though note that this cannot be discerned with ease by inspecting the rendered article. It involves either comparing rendered (screen) text with the source

[28] The 7 articles using 2 discrete forms raise the effective overall occurrence count from 2,127 to 2,134.

[29] In 2 of the 7 instances the transcludes is an effective duplicate as one is to a redirect between accented and unaccented page titles for the same article.

[30] The → denotes nested tagging.

[31] A deliberate pattern where `<onlyinclude>` tags bracket a show's episode count such that the effect of the transclude is a 'macro' to return the show's current episode count—see section 4.6.1.4.

[32] All 3 inclusion markers found in no particular nesting/arrangement.

[33] LST: Labelled Section Transclusion. Here, called via #lst (155), #section (100) and #lsth (2) forms.

Mark-up	Occurrence	% of Transcluded Articles
Fulll article[29]	8	0.10%
`<includeonly>`	4	0.05%
`<includeonly>` →[30] `<noinclude>`	4	0.05%
`<includeonly>` → `<onlyinclude>`	1	0.01%
`<noinclude>`	220	2.79%
`<noinclude>` → `<includeonly>`	60	0.76%
`<noinclude>` → `<onlyinclude>`	11	0.14%
`<onlyinclude>`	6,347	80.46%
`<onlyinclude>` → `<noinclude>`	585	7.42%
`<onlyinclude>` → `<includeonly>`	308	3.90%
`<onlyinclude>` 'count' `<onlyinclude>`[31]	50	0.63%
`<onlyinclude>` `<includeonly>` `<noinclude>`[32]	5	0.06%
`<onlyinclude>#ifeq` dynamic section	21	0.27%
`<section>` via `transcludesection`	20	0.25%
`<section>` via LST[33]	244	3.09%

TABLE 4.7: Use of different transclusion-invoking code forms in 'en' Wikipedia

code of source and target transclusion pages. Wikipedia does have templates for indicating the presence of transclusion [W189] and section transclusion [W190] that insert boiler-plate text such as 'Part of this section is transcluded from Trace amine. (edit | history)' (q.v.) [W6]. This is akin to the templates for articles pointing to the main articles on a subject. However, these templates did not appear to be widely used compared to the occurrence of actual transclusional use. The lack of clear documentation and worked examples must be considered as a factor in the lessened use of the templates.

In the 'en' wiki, the `<onlyinclude>` mark-up (see section 4.3)) is the predominant means of controlling transclusion scope. In fact, the <onlyinclude>-based methods form 92.42% of use. As the `<noinclude>` form is the functional opposite[34] of `<onlyinclude>` their combined share amounts to affecting 96.11% of all transcludes.

The data in Table 4.6 shows seven different nesting combinations totalling 974 instances, 12.34% of the 7,888 discrete pages with scope -controlling mark-up. Nested usage is inferred to be a means to control inclusion of page elements like the title and lead paragraph. The is no clear sign, albeit in a very small sample, of 'defensive' coding whereby one page's editor(s) endeavour to stop re-use of the page via transclusion. That said, there is nothing, technically, to stop such behaviour occurring. Were it to occur there is no sign of any automatic mechanism to detect such behaviour other than by other editors viewing the page source. This is tentative evidence that the effective managers of the site (mainly the most active editors) have few if any tools *at present* to police such eventualities.

[34] Mark up what should be omitted from transclusion ('noinclude') vs. only what should be transcluded ('onlyinclude').

During the transclude analysis in Tinderbox, it was noted—via dates used in article names—that articles about recent events, especially sports-related events, were starting to use section-based LST. It is unclear whether this reflects the habits of a few editors or a more general move consensus. In context, informal analysis suggested sports pages, especially for team sports, use template-based data rather than transclusional article elements. This may be an echo of the 'en' wiki's move away from the use of sub-pages: templates are effectively more hidden. No discussion of this subject was found.

Wikipedia talk page comment [W169] indicates a possible socio-technical aspect affecting the generally low use of content transcludes. That page showed concerns about oversight of page content and of transcludes being a vector for inclusion of unevaluated work or just outright vandalism. Thus, although transcludes appear to be used, the degree of their use is much lower than might be expected if the technique were employed as a deliberate approach to writing effective hypertexts.

Instances of non-transcluded duplication can be found, though not with ease or in volume. The article on President Gerald Ford [W33] has a section on the pardon of President Nixon [W34]. The source code in that section is the identical, bar some source code white space differences to the opening paragraphs of the article on the 'Pardon of Richard Nixon' [W131][35]. Note that the latter page was only created on 1 February 2016 (since the date of sampling) and may still be under active edit. Indeed, re-sampling both pages on 27 July 2016 (Nixon, Ford) [W132, W35], the Ford page has gained a new third paragraph and the original 3 paragraphs have 18 textual differences from subsequent edits to both pages. This is a clear case where a sectional transclude would be appropriate for coherency in the wiki's corpus of knowledge.

Although not available in the data dump, a Special namespace page 'Most transcluded pages' [W150] lists the top 5000 such pages. The article is updated at approximately 6-month intervals; the sampled data was as at 28 April 2016. The list contains no active articles and comprises items from 6 namespaces: Template, Module, User, Wikipedia, Portal, MediaWiki. Of the first 10 items, 7 are templates and the remainder are modules (code used by templates or, less often, directly by page mark-up).

Item 5,000 in the above list has a transclude count of 2,526 pages which places it well above the counts seen for transclusion of live articles. The maximum count found in the 'en' wiki for an article transcluded into other articles is 22 for 'List of Polish gminas' [W92][36]. Wikipedia editors appear to see little distinction between templates and article transclusion. Apparently the highest ranked template attached to a single subject is 'Template:Country data United States'[37] [W175], appearing in the list at position 307 and used on 185,738 pages.

[35] On-screen reference numbering also differs. Numbering is ordered when rendering and the two pages differ in the overall reference list. Numbering thus differs due the two containing page having differing reference lists.

[36] The polish plural of *gmina* is in fact *gminy* but the article title used is as per the source.

[37] The template [W175] allows a page to transclude a flag icon for the United States of America into the calling page: such templates are heavily used in sports-related pages.

Namespace	Count in list	First listed item
Template	4,422	#7: *Template:Yesno* (used on 7,620,988 pages)
Module	283	#1: *Module:Arguments* (used on 18,149,945 pages)
Wikipedia	239	#314: *Wikipedia:WikiProject Film/Coordinators/Banner* (used on 179,275 pages)
User	46	#77: *User:ClueBot NG/Warnings/FPReport* (used on 1,390,480 pages)
Portal	8	#1060: *Portal:Slovenia/Things you can do/tasks* (used on 16,091 pages)
MediaWiki	2	#587: *MediaWiki:Toc* (used on 68,278 pages)

TABLE 4.8: Subject groupings of most transcluded articles ('en' Wikipedia).

As already shown, use of inline mark-up to scope transcluded data appears common in transcluded articles. However, there are instances of misplaced or non-paired tags which suggests editors are not clear as to correct usage for such mark-up. The trio of tags for controlling transclusion scope in the transcluded page should be used as opening and closing pairs but errors are not uncommon. The error rate is likely flattered by the presence of bot-created content which should be less prone to general mis-keying errors (assuming proper coding).

Language	Tag	Articles	Unmatched	Error %
'de'	onlyinclude	30,000	102	0.34
	noinclude	811	69	8.51
	includeonly	158	0	0
'en'	onlyinclude	11,354	550	4.84
	noinclude	2,563	1,121	43.74
	includeonly	941	7	0.74
'it'	onlyinclude	86	24	27.91
	noinclude	7,416	463	6.24
	includeonly	47	0	0
'ru'	onlyinclude	784	89	11.35
	noinclude	20,692	771	3.73
	includeonly	242	2	0.83

TABLE 4.9: Error rate for closure of transclusion-scoping tags.

The counts in Table 4.9 show all articles using the tag type compared to only those with uneven tag pair counts.

4.6.1.2 Evidence of Localised vs. Pervasive Use of Content Transclusion

Deliberate use of transclusion may be evidenced by particular areas of content showing a greater presence of transclusion, assuming the latter is not the work of the same editor.

In the 'en' wiki there are 2,184 discrete pages using transcluded material, of which 2,127 are valid non-main-namespace transclusions. Within this, 243 articles are associated with a (then) current project to reorganise listings of minor planets[38], representing 11.27% of the overall count. These transclusions were scheduled to disappear soon after the sampling date as part of the project's completed re-organisation (though not due to use of transclusions per se) and so were discarded from further analysis.

The remaining 1,884 items were coded into 24 broad (subjective) subject groups. Aggregating all grouping with less than 2% occurrence, 6 main subject groupings emerged as areas where transclusion occurs, as seen in the table below:

Subject	**% of Use**
Episodic Media	59.71
Sport	13.22
Political	8.17
Administrative	4.51
People	3.18
Climate	2.07
18 groups of <2% each	9.13

TABLE 4.10: Transcluding articles ('en' Wikipedia) grouped by general subject[39]

The biggest grouping, over four times the size of the second, is for what may be termed 'episodic media'[40]. This is comprised primarily of TV show listings, with a smaller number of (print) comic book series as well. Given the preponderance of TV listings within the transclusion counts for episodic media, further investigation was made of the degree of use of transclusion across the subject area using a list of UK and UK TV shows derived from the Wikipedia page listings: 'List of American television series' [W77] and 'List of British television programmes' [W78]. The combined US-UK listing contained 3,974 discrete items (excluding a further 159 duplicate items). The intersection of that list with the above transclusion was 174 items (4.48%), whereas on count alone the transcluding page list of 1,133 items was 28.5% of the overall US-UK listing. This is likely to be due in part to the fact that many of the transcluding articles are titled as 'List of...'. Stripping the latter from the 1,133 transcluding articles leaves 654 items which equate to 16.5% by count (not match) of the US-UK combined list.

Although episode listings formed the largest single grouping of transclusion use within the 'en' Wikipedia, it remains a small proportion of the available subject articles in that area. A note of caution is warranted, because less recent shows may not have detailed episode data still available and thus recorded on Wikipedia. Thus the existing transclusion use may be a more significant

[38] More detail can be found in the pages of the 'List of minor planets' [W89].

[39] Totals as 99.99% due to rounding.

[40] This can describe media in various formats which is created and broadcast/printed as more than a single release. Thus TV show episodes, editions of a comic or magazine. Long-running TV shows may have a multi-level episode arrangement whereby a show will comprise seasons or series, each containing a number of episodes.

proportion of contemporary episodic media than the figures presented above, but it is not considered that matches to a revised show listing would be significantly higher.

As part of preparatory analysis for RQ4, the 'episodic media' grouping above was broken out and observed in more detail. This smaller sample contained 4,683 items formed of 1,125 transcluding articles and 3,545 transcluded articles. That included duplication of 221 items which appear in both roles. In Table 4.11 the transclusion-calling code methods are restated using the same criteria as Tables 4.8 and 4.9, only showing mark-up items with non-zero values:

Mark-up	Occurrence	% of Transcluding Articles
`{{:` (i.e. whole article)	1,123	99.82
`#lst`	2	0.18

TABLE 4.11: Use of different transclusion-invoking code for episodic media articles in 'en' Wikipedia

Note the almost total absence of calling code for partial transclusion. This implies most of the scoping of transclusion is being done in the transcluded pages. Table 4.12 shows the same sample pages' data for in-page scoping:

Mark-up	Occurrence	% of Transcluded Articles
Full article	3	0.06
`<noinclude>`	4	0.11
`<noinclude>` → `<onlyinclude>`	9	0.25
`<onlyinclude>`	3,086	87.05
`<onlyinclude>` → `<noinclude>`	95	2.68
`<onlyinclude>` → `<includeonly>`	287	8.10
`<onlyinclude>` 'count' `<onlyinclude>`	50	1.41
`<onlyinclude>` `<includeonly>` `<noinclude>`	4	0.11
`<section>` via LST[41]	7	0.20

TABLE 4.12: Use of different transclusion-restraining code forms in 'en' Wikipedia (tag usage variants are described in footnotes to Table 4.7)

The sample contains all of the 50 instances of the episode 'count' form but interestingly contains little use of LST transcludes. This may be explained by the fact the latter is a newer form of transclude-scoping mark-up and editors may still simply be copying older examples from before LST was introduced.

Analysing data from the 'de' wiki, 11,165 transcluding articles were reviewed with 'en' data for broad subject area, as shown below. Note that due to language the 'de' data was classified on the titles of the calling article and not reviewed to the depth of 'en' data as for RQ1. The 'de' data subject areas were:

The 'de' Disambiguation category most closely maps the People in 'en' data, though in the 'de' wiki disambiguation pages are used much more widely than just for people's names and extends

[41] LST: Labelled Section Transclusion. Here, called via #lst (155), #section (100) and #lsth (2) forms.

Subject	% of Use
Disambiguation (‘*Begriffsklärung*’)	88.26
Music	5.39
Sport	3.67
Film/Media	0.15
All other groups	2.53

TABLE 4.13: Subject groupings of transcluded articles (‘de’ Wikipedia).

to companies, administrative bodies, brands, etc. Lack of facility with the German language precluded deeper investigation of areas like Episodic Media, to see why it might not feature as a grouping as within the ‘en’ data.

The ‘it’ use of transclusion is even narrower. Of 5,306 transcluded pages, 1 does not exist, 9 are outside the main namespace and 130 are called by dead (renamed?) pages[42] resulting in a sample of 5,166 articles which relate to the following subject types:

Subject	% of Use
Born…(‘*Nati…*’)	46.52
Died…(‘*Morti…*’)	53.41
Personalities	0.04
Computer Games	0.02
Sports	0.02

TABLE 4.14: Subject groupings of transcluded articles (‘it’ Wikipedia).[43]

Here, apart from a few outliers, all transclusion use is in a single area of recording the date of birth or death of people. Together, they form 99.92% of transclusions.

4.6.1.3 Variability of Content Transclusion Between Per-Language Wikis

Comparing forms of use across different languages sampled, three (‘de’, ‘it’ and ‘ru’) wikis are noticeable for having a higher percentage of articles using transclusion. Lack of understanding of Cyrillic precluded deeper study of the ‘ru’ data but analysis of the other two wikis did show a significant skew in subject area. In both cases these were different from each other and from ‘en’ data.

The ‘de’ Wikipedia contains a lot of pages including *Begriffsklärung*’ (‘disambiguation’) in the title. Whilst these are explicitly disambiguation pages, the wiki also contains a significant number of unlabelled disambiguation pages (as discovered when investigating content associated with explicit disambiguation pages). Thus there may be discrete pages for forenames, family names, general uses of a word, spelling variants (for names), etc. For instance, the ‘Schmid’ [W41] page

[42] This appears to be due to a recent re-organisation of a set of pages all relating to listing computer games.
[43] Totals as 100.01% due to rounding.

lists 523 people of that family name. Whilst some entries are individually coded, the article's list also includes 57 full-page transcludes (using the '{{ ' mark-up form).

One such transclude is the Albert Schmid [W36] page, wherein the source code wraps the bulleted list within the page content in <onlyinclude></onlyinclude> tags. This pattern of disambiguation pages transcluding several partial pages would account for the large number of <onlyinclude> instances in the 'de' wiki. This profuse use of disambiguation pages does not appear to be bot-created[44], but is perhaps a positive example of '*Gründlichkeit*'[45], and shows deliberate and correct use of the transcluded-page mark-up. By comparison, the 'en' wiki analysis showed instances of similar tags incorrectly paired or positioned within the content. As is seen in the next example 9in the 'it' wiki) it is possible that individuals, or small groups of editors, may be responsible for localised occurrences of transclusion or for wide use across a given type of article. It is also worth noting that the more clearly defined the scope and context of the task the greater the opportunity for a bot to be used for the task.

The 'it' wiki also had a predominant partial transclusion pattern, though based on a different theme—that of dates. There were 3,046 transcluding pages of which 2,901 (95.24%) are pages relating to a year, or a day of a month. Of the 5,306 transcluded pages, 5,160 (97.25%) also relate to dates of birth or death. Thus, article '96 a.C.' (96 BCE) [W63] transcludes 'Nati nel 96 a.C.' (Born in 96 BCE) [W67] and 'Morti nel 96 a.C.' (Died in 96 BCE) [W66]. Although this appears to not be a completely bot-created set of articles, the bot 'Gacbot' certainly appears [W68] heavily involved in development of this strand of listing articles— both transcluding and transcluded.

Of the remaining 'it' transcludes, all but 5 are listings of video games e.g. 'Lista di videogiochi per Commodore 64' which since appears to be renamed , or is a redirect for, 'Videogiochi per Commodore 64' [W69].

The main group of transcluded articles in 'it' follow the same pattern as in the 'de' case above with the difference that whereas the 'de' articles wrap the inner 'content' section of the article in <onlyinclude> </onlyinclude> tag pairs, the 'it' articles use two separate pairs of <noinclude> </noinclude>. Both approaches work as the tags are functional opposites. However, it can be noted that whilst the underlying pattern is the same, two different language wikis use a different implementation and that one ('de') is arguably more efficient than the other—if only in needing half the number of mark-up tags to achieve its aim.

The 'en', 'de' and 'it' wikis all show content transclusions but there is little obvious similarity other than the general manner or scope of use; other languages show very little transclude use at all.

[44] From inspection of the earliest page records. In themed pages that are bot-generated, a bot or AutoWikiBrowser edit is found as the original page creation edit or early revision. Bots may subsequently revise the page, but in a different role.

[45] A German notion of 'thoroughness'. This skein of name disambiguation pages—first or family names—is carried through with a consistency not seen in other language wikis. Even so, within that frame, the scope of transclusion use is less than it could be.

Another way to observe differences was to compare articles that might be expected to be similar despite languages, such as articles for sports teams in a multi-national league and comparing the pages across different datasets. An examples is shown in Table 4.15. As the languages sampled were all present in the Northern Hemisphere, ice hockey was chosen as a sample context. The 'Boston Bruins' ice hockey team was found to have a page in 9 out of the 10 sampled languages, although in 4 cases these are just stub pages with little data.

Of the 5 languages with a non-stub page there is little consistency in how the team roster listing is implemented: the 'fr' article has featured article (FA) status. In Table 4.15 two pages used inline mark-up (i.e. no transclusion), two used a template [sic] and one used a transclude.

Wiki	Uses Transclude?	Article
de	Template: `{{Navigationsleiste Kader der Boston Bruins|Format=Tabelle}}` [W42]	'Boston Bruins' [W37]
en	Template: `{{Boston Bruins roster}}` [W9]	'Boston Bruins' [W8]
es	None: stub page only	'Boston Bruins' [W142]
fr	Article: `{{:Effectif actuel des Bruins de Boston}}` [W28]	'Bruins de Boston' [W27] [46]
it	None: stub page only	'Boston Bruins' [W64]
ja	None: stub page only	'ボストン・ブルーインズ' [W71]
nl	None: stub page only	'Boston Bruins' [W22]
ru	None: created via inline code in main article	'Бостон Брюинз' [W138]
sv	None: created via inline code in main article	'Boston Bruins' [W165]

TABLE 4.15: Differential implementation of the same data in different languages

Cross checking against other teams in the same country for Canada ('fr'), US ('en') and Sweden ('sv') the choices of article transclude vs. template transclude vs. manual code appeared consistent. This implies different counties hold to different editing styles, though the initial choice is unrecorded and harder to pin down; it is likely that as content has been built out existing article coding styles have been followed cementing an apparent consistency of style. Lack of clear style/best practice guides also mitigate against consistency to a pan-language best practice.

Another investigation used the article for 'Manchester United F.C.' [W109] as this club has a high profile globally and has a full article in all but one sampled wiki ('ceb'). The existing articles had featured article status in four (en [W109], es [W143], ru [W139], sv [W166]) of the remaining nine wikis (de [W38], fr [W29], it [W65], ja [W72]).

The 'en' page had no article transcludes, but transcluded 7 discrete tables of data—albeit via templates:

- Template:FIFA Club World Cup winners [W179]

[46] This page has 'Featured Article' status.

- Template:Football in Greater Manchester [W180]
- Template:Manchester United F.C. [W183]
- Template:Premier League [W186]
- Template:UEFA Champions League winners [W193]
- Template:UEFA Cup Winners' Cup winners [W194]
- Template:UEFA Super Cup winners [W195]

The rationale for use of a template rather than an explicit article transclude is not clear but several causes may be posited. Firstly, there has been a historical bias against transclude, as seen in the review of documentation (see section 4.6.1.6). Secondly, templates are regularly used in the creation of tables despite prescription during the early days of Wikipedia of the use of '(static) prose in templates', emergent practice has clearly drifted away from that intent.

The tables described above are all what Wikipedia terms 'navboxes' [W425], collapsible tables normally shown at the foot of an article. The 7 templates used by the English Manchester United page actually create 10 navboxes (some nest within others[47]). Of the other languages, 'es' and 'de' showed no navbox use. The remainder showed varying amounts:

Language	**# of Navboxes**
en	7(10)
de	0
es	0
fr	2
it	8
ja	8
nl	2
ru	5
sv	4

TABLE 4.16: Navbox use in 'Manchester United F.C.' articles

The navbox templates are reproduced, or localised, in each of the languages. In contrast with interwiki usage, comparing the English Manchester United with other English articles for UK Premier League clubs, there is some common use, e.g. 'Template:UEFA Cup Winners' Cup winners' for those clubs having one that cup. League club's pages tend to show common elements through use of templates (transcluded data) although they otherwise diverge in style. That is unsurprising as key editors will likely be keen supporters focused more on their own club that data about other clubs.

[47] This may be due to nested transclusion within the templates, but analysis did not cover this.

4.6.1.4 Patterns of Content Transclusion

Of the 10 sampled languages, the transclusion structures of three ('de', 'en' and 'it') were studied in greater depth. Although 'ru' also had transclusions present, lack of native understanding of Cyrillic precluded further study to date. RQ 1.4 seeks to ask if there are fundamental patterns of transclusion than can be observed in a large collaborative hypertext system.

Transcluded lists. The most common pattern of transclusion is calling information listings: for instance, lists of media episodes, people and events. The largest single example was found in the 'en' Wikipedia, in the 'List of heads of missions of the United Kingdom' [W83] article that summarises the United Kingdom's diplomatic missions around the world. The page used 205 transcludes calling 160 discrete target articles; some target articles were used more than once e.g. where an ambassador is responsible for more than one country. The article was also unusual in that the deliberate manner of the process is explicitly mentioned in the source code of the calling page [W84]:

```
<! NOTE: Where transclusion code has been inserted
(e.g. Afghanistan, Albania, etc.) this page doesn't need
to be changed for a new ambassador or high commissioner
- change the country list page only >
```

Instructions are also added in the transcluded target pages [W82] such as:

```
<! The coding in the following line is for the *current
ambassador*. On a change of ambassador, please make a new
normal line (above this comment) for the outgoing ambassador,
then put the new name (wikilinked as appropriate)
between the onlyinclude and /onlyinclude tags, and
any reference *after* the /onlyinclude tag. This takes
the name automatically to "List of heads of missions
of the United Kingdom" >
*2015-{{As of|2015|9|alt=present}}
: <onlyinclude>[[Sebastian Wood|Sir Sebastian Wood]]</onlyinclude>
```

This shows an atypical degree of thought about the editor likely to be making maintenance edits to the articles, for instance a junior Foreign Office clerk with more pressing tasks at hand: good briefing notes help.

The transcluding article's talk page [W167] showed that it was first created in 2006. By 2010 it was reported as out of date and to be considered for deletion. The current solution was implemented in 2012, involving the placement of scoped-inclusion mark-up in all the to-be-transcluded articles. Also of note was the talk page comment of 26 October 2012 suggesting use of a substitution-based template 'Template:Onlyinclude' [W185] on the target pages. It is not made clear whether this was due to editors' unfamiliarity with mediawiki mark-up or for other reasons.

In the case of 'de' and 'it' wikis, partial transclusion of lists was the pre-eminent pattern seen. The transcluding (calling) article will partially transclude one or more other listing-style articles. The transcluded articles contain inline transclusion-scoping mark-up such that only part of the article is transcluded. For instance, a listing article might be configured to only transclude the main listing part of the page omitting what may be regarded as the listing's header and footer content. The 'de' wiki opts for use of `<onlyinclude>` markers whereas the 'it' wiki prefers `<noinclude>` markers. The two forms are functional opposites though, as used, they work to the same end: the former encloses only that which *is to be transcluded*, the latter that which `should not be transcluded`. The choice of mark-up is consistent with either bot-created code or human editors copying code patterns seen in existing pages.

Lists of lists. This is seen in the 'en' project listing minor planets, such as in 'List of minor planets: 147001–148000' [W90]. In this pattern, each page is transcluding 10 sub-pages, e.g. 'List of minor planets/147001–147100', 'List of minor planets/147101–1472100', etc. This pattern cannot be traced in entirety in the current 'en' Wikipedia as the minor planets project has removed all the sub-pages since the data was sampled and the sub-page links are now re-directed. The talk page [W168] for the main 'List of minor planets' article is useful in showing some of the project's editors' discussion in finding consensus on how to improve the listing structure.

The 'ja' wiki has a single instance of `#lstht` use but also the largest in all the sampled wikis, the page '年度別映画興行成績'[48] [W75], which lists film box office data for the 55 years 1960-2014[49]. The page transcludes 148 other articles employing between 1 and 7 per-locale data[50] transcludes per year, each of which transcludes the top 10 grossing movies for that locale.

Episodic media listings. Multi-level listings inter-related via transclusion can be seen in the 'en' pages relating to the US TV crime drama *Criminal Minds*. The main article [W16] describes the show and includes a table of episodes [W19] that uses manually coded similar to that in the 'List of *Criminal Minds* episodes' [W80] article. The list article this transcludes data from 11 per-season episode listings: '*Criminal Minds* (season 1)' [W18], etc. Since the date of the dumps, the main article has been altered [W17] to transclude character detail from the 'List of *Criminal Minds* characters' [W79] but the listing of season and episode data is still manually encoded.

[48] The title translates roughly as 'Movie box office, by year'.
[49] The page was created in September 2004 [W73] but only moved to using LST in November 2015 [W74].
[50] Each transclude is per-year data for a different locale such as global sales, North America, Japan, etc.

Hidden transclusion effects within episodic listings. The transclusion by article for the 'List of the Penguins of Madagascar episodes' [W100] shows episodic listing involving transclusion-scoping 'hidden' in a template (a complication discussed in section 4.3). The main list calls three per-season episode listings (Season 1, Season 2, Season 3 [W196, W197, W198]).

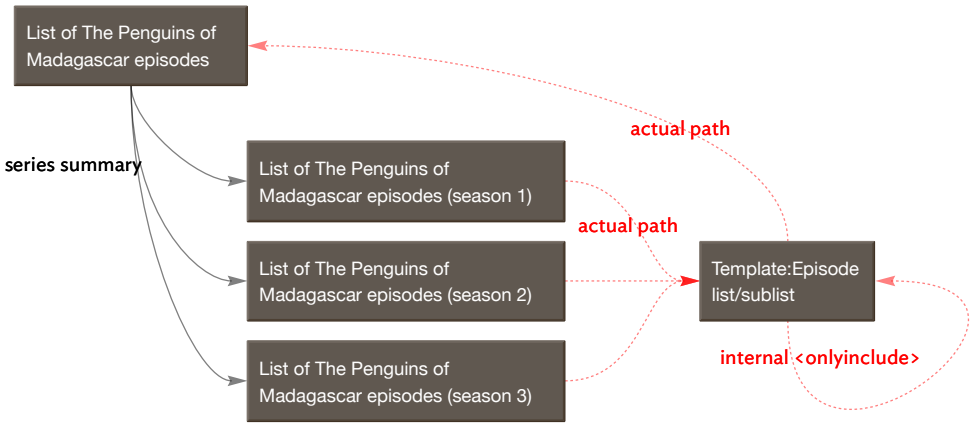

FIGURE 4.6: Hidden Transclusion Pattern

The template used above is actually a nested call from within 'Template:Episode list' [W176] to 'Template:Episode list/sublist' [W177] using Lua code from 'Module:Episode list' [W130][51]

Thus the listings pattern may involve transclusion of more than one article, e.g. of several per-season episode listings of a TV show, and each transclude may use only part of the source article although with forethought the pages could be structured so as to avoid the need for a lot of inclusion-scoping tags.

During investigation of these templates, a number of Wikimedia namespace articles relating to episodic data were found that have some discussion (also in talk pages) relating to the manner of setting up listings:

- 'Wikipedia:WikiProject Television/Episode coverage task force' [W719]
- 'Wikipedia:Naming conventions (television)' [W661]
- 'Wikipedia:Television episodes' [W681]

Although the listings pattern was observed in the 'en' wiki, no significant use was detected in other wikis. The reason for this is unclear but is consistent with the generally lower levels of transclusion use in other wikis. The 'de' wiki transcludes only 48 episodic media related articles, but 595 pages for musicians and singers transclude part of their discography page, e.g. Marianne

[51] The suppression, or non-inclusion, of the full episode list table data during transclusion is handled in the closing part of the Lua code.

Faithfull[W39] transcludes her studio album list from her 'Marianne Faithfull/Diskografie' page [W40] [52].

Episode count. A number of TV show articles transclude so as to return the show's current episode count. This is most widely seen with the 'en' article for the TV cartoon series, 'The Simpsons' [W199]. The article is transcluded 10 times by 9 different articles:

- Fox cartoons [W26].
- History of The Simpsons [W59].
- List of longest-running U.S. primetime television series [W87]. The article is transcluded twice, in the listings for '23–29 seasons' and '12 seasons and up'.
- List of longest-running U.S. broadcast network television series [W86].
- List of longest-running United States television series [W88].
- The Simpsons [W199]. The article transcludes itself to use the episode count in the article's info box.
- List of *The Simpsons* episodes [W101].
- List of *The Simpsons* guest stars [W102].
- Matt Groening [W110].

The use of a show's article to transclude the episode count appears a recent pattern. In the case of *The Simpsons* this format was only adopted in October 2015 [W200]. The rationale for the change is not discussed in the talk page, but the use of the transclude replaces previous use of transcluding the episode from a now-redundant template 'The Simpsons episode count'.

This pattern is the most easily quantified, given its narrow scope. Analysis of the 'en' wiki revealed 50 instances of the 'episode count' form of mark-up (see section 4.6.1.2). Every one occurred in articles relating to episodic media. However, this indicates only limited coverage amongst articles on other shows using the 'listings' pattern (c.1,000 others, 5%).

Circular transclusion. The 'en' wiki's main article for the '*iZombie* (TV series)' [W70] transcludes the article 'List of *iZombie* episodes' [W85]. The both use `<onlyinclude>` mark-up to limit the degree of transclusion. The main article transcludes a summary table of show episodes. In turn, the List article transcludes the main page, which uses similar mark-up to return the current episode count for the series using the episode count method already described:

There is no discussion seen as to possible side-effects of this circular transclude were an editor to use insufficiently tightly scoped mark-up. These circular links can be more complex:

[52] It is not clear why studio albums are transcluded but the separate live album listing is not. This practice is consistent across other similar pages sampled.

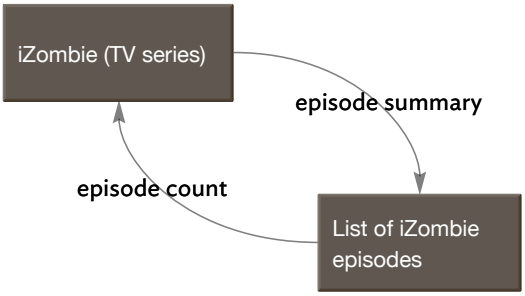

FIGURE 4.7: Circular Transclusion Pattern

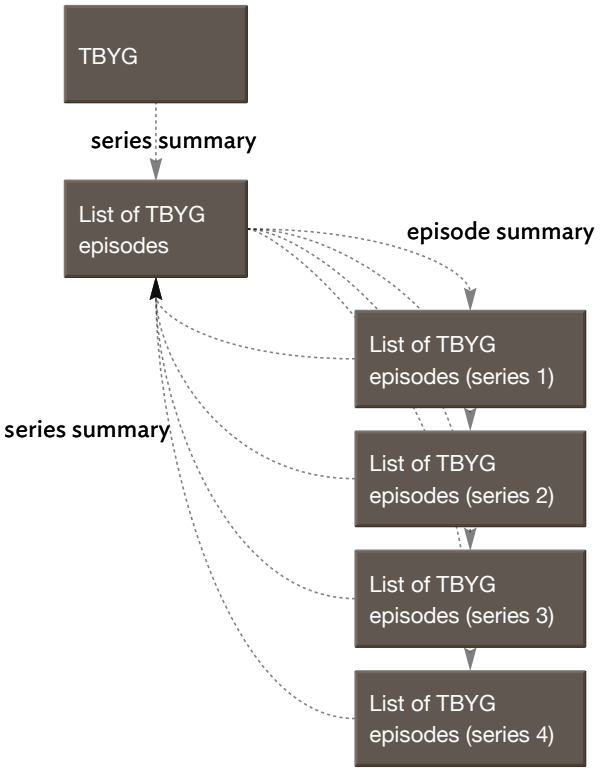

FIGURE 4.8: Circular Transclusion Pattern in 'TBYG' Show

The circular is pattern is seen on a larger scale in articles for the show 'Talkin' 'Bout Your Generation' (TBYG) [W170]. Here the article 'List of Talkin' 'Bout Your Generation episodes' [W95] is configured to return a summary table of the show's series. This is transcluded by the show's main article [W170] and by each of the four per-series 1-4 listing articles: series 1, series 2, series 3, series 4 [W96, W97, W98, W99].

Transcluded indexes. The 'en' article 'List of Polish gminas' [W92] contains an index row of letter links to sub-lists of gminas [W43] (Polish administrative areas) starting with the same letter; letters are only included if there are gmina starting with that letter (22 of 26). Thus, the index's 'A' link opens the page 'List of Polish gminas (A)' [W93], etc. Each per-letter listing repeats the main page's letter 'index' list via transclusion. The main page's inclusion control actually defines

two portions of content, the page down to and including the index and the category section. This apparent 'wrapper' for content works due to the fact Wikipedia page rendering places categories at the end of the page regardless of where they are in the articles source code. However, the editors of the listings clearly felt the need for an index at the bottom of long listings (for example the 'R' page [W94]). This foot-of-page index was effected via a wiki navbox [W425] using a template 'List of Polish gminas' [W182].

Note the pattern is slightly odd in the that main page also includes the 'A' content. Inspection of the page history shows that the list was originally a single page and was subdivided c.2008; possibly the editor felt there was insufficient content for the main list page and so retained the content listing for the letter 'A'. The main (original page) was written in March 2007, splitting out per-letter sub-pages [sic] in January 2008. In December 2009 the sub-pages were 'moved' to full articles[53]. The navbox template was subsequently created in May 2009 and added (without edit comment, or talk page notes) to all per-letter list pages.

Self-transclusion. An example of self-inclusion of content, for deliberate repetition purposes, is seen in the 'en' article 'List of devices using Mediatek tablet processors' [W81]. The header row of the per-item data table is source-labelled as the section 'header', whilst all subsequent tables on the same page transclude this via`{{#section:List of devices using Mediatek tablet processors|header}}`. Note this pattern is also a circular transclusion.

With the exception of the large numbers of listing transcludes in the 'de' and 'it' wikis, most patterns above are found in very small numbers. In most cases there is no clear record of the initiating editor's intent. The variations seen would suggest that in some cases proliferation of a pattern is by loose copying of an existing pattern with variations occurring added either deliberately or through the editor's lack of understanding of transclusion patterns.

4.6.1.5 Finding And Tracing Content Transclusion

Collaborative tools like Wikipedia allow remote access (with or without login[54]) for editing. It is thus reasonable to expect that an engaged new user should be able to find answers about the mechanics of editing and article structure, without *having no option but to* 'asking an expert'. The latter amounts to a failure of the Help and documentation on offer for all but the most arcane details. Whilst not necessarily deliberate, it is still a failure by the authors towards those who *need* the documentation by writing for themselves rather than the reader.

Tools to help trace or review transcludes are not well represented. Only the 'en' Wikipedia article edit pages have a list of transcludes but as this includes all templates and models and appears to

[53] Not cited in revision notes, but presumably due to the 'en' Wikipedia's move to remove sub-pages in general.

[54] Sensitive pages, such as areas of social interest or political disagreement are often protected and only logged in editors may alter a page. The most sensitive pages can have further restrictions requiring a user to have been registered for a specific period of time and/or have made a specified number of edits. These latter measures attempt to establish that any edits are done by experienced editors.

include items transcluded by transcluded items, this list is not helpful in determining transclusional use. Documentation is present, but as discussed in section 4.6.1.5 it lacks coherence and is often out of date.

Edit pages, within the 'en' Wikipedia only, offer a listing of transcluded items. This is shown immediately beneath the source code edit box, and links to each listed item to its current source page. Items listed appears to be grouped by namespace and within those groups are sorted alphabetically. The ordering of namespaces is not documented but transcluded main namespaces appears to be placed first with Template and Module at the end. The latter two namespaces are the primary ones found in the listing for most pages, for instance 'Ted Nelson'[W171] which transcludes 0 articles, 41 templates and 38 modules.

This transclude list is confusing as it *appears* to include templates and modules transcluded by other templates and modules; there is no (linkage to) documentation on this matter. Certainly, the listing includes items not directly cited in the current page's source code. Complex templates can use several layers of transclusion and this behaviour is also generally undocumented. This leaves the listing of little practical value in the study or evaluation of transclusion, especially the new editor. The fact that the listing of transclusions on the edit page is not implemented outside the 'en' wiki can only have a negative effect in other language wikis upon the ease of using and maintaining transclusions.

Pages in most major namespaces (in the 'en' wiki[55]) offer a left-sidebar link to a Special namespace page 'What Links Here'. As with the article's edit page listing this mixes page results from multiple namespaces. The listing appears unsorted and offers no means to re-sort the data. As with many Wikipedia Special-namespace pages, an individual page is limited to returning 500 items at a time (default is 50) and has a maximum retrievable total of 5000 items. Thus on well-linked pages, the utility of (the rendered) page to the reader is reduced as it is hard to scan the data therein for items of interest. For pages that are transcluded (i.e. the target of a transclusion, the transcluding (calling) page is listed with the notation '(transclusion)' after the listed page's name. Helpfully, the page includes filters allowing individual toggling of the 3 discrete categories of links in the list: links, redirects and transcludes. The filters act across the whole list, not just the current page items, though it is not clear what happens there the list count exceeds to 5,000 item cap). The filters, if applied, are passed as Uniform Resource Locator (URL) arguments so could be used for automated retrieval. Special pages are not cached/saved and are not replicated in the eXtensible Markup Language (XML) dumps.

For example, consider the default listing [W163] of pages that link to 'The Simpsons' article vs. a listing showing only transcludes [W162]. Unfortunately, the transclusion is not so easily similarly traceable via transcluding page. The 'en' Wikipedia (only) does include transcluded articles in the listing beneath the code box on the article's edit page, but offers no filter to isolate only true transcludes. The fact that the 'What Links Here' page has a transclude-only filter indicates that

[55] All 10 languages sampled included a link to a localised version of the Special namespace's 'What links here'. Only main namespace articles were tested in languages other than 'en'.

at least some in the Wikipedia community understand the concept of a content transclude as discrete from the more generic template-modified form of transclusion.

In general, implementation of transclusion is difficult to comprehend—or to learn to use— due to the fragmented documentation. The problem is compounded by the lack of both clear examples and any (stylistic) guidance as the Wikipedia community's views on transclusion use. Maintaining articles with transclusional links is also made difficult by the lack of tools giving a clear listing of transclusional links. Despite the proliferation of alternative methods, no public Wikipedia documentation appears to show any move to rationalise use—as might easily be done.

4.6.1.6 Attitude of Wikipedians to Content Transclusion

The fact mediawiki supports transclusion does not mean wikis running on the software are required—or encouraged—to use it. How 'en' Wikipedia documents transclusion and talk page comment about transclusion should offer some insight into editors' views on transclusion.

The documentation of transclusion is scattered across a number of web pages on both the 'wikipedia.org' and 'mediawiki.org' domains and several namespaces within each of these domains[56]. This demonstrates the strength and weakness of such collaborative structures. Anyone can add content but no-one is obliged to do so. As importantly, no one is obliged to update documentation. The resulting infrequent updates are likely due to a faith in the wisdom of crowds, e.g. 'Linus's Law' [296, p.30], and a lazy assumption that *someone else* will notice and correct errors. Talk pages do allow editors to flag concerns and suggest fixes but there is no presumption for resulting action. Hereafter in the term 'Wikipedia' refers to (English language) documentation on both websites.

Wikipedia describes the general process of transclusion in the main namespace article 'Transclusion' [W202]. This gives some general coverage of the origin of the term and the general forms of use. No description of, or linkage to, documentation of Wikipedia's own use of transclusion is given. The learner must go to ancillary namespaces such as 'Wikipedia' or 'Help', or to mediawiki documentation to find description of how transclusion is invoked.

The most information contained in any single relevant article is in 'Wikipedia:Transclusion' [W687]. This is a long[57] article, describing both how transclusion works within Wikipedia and covering most (but significantly not all) of the inline mark-up required: the LST method is omitted, although there is a summary section and link to relevant content at 'Help:Labeled section transclusion' [W52]. The main Transclusion article has a section 'Applications of transclusion' [W690] though content given is uninformative with a very small number of overview examples, mostly outdated and with only a single[58] relating to content in the main namespace of the wiki,

[56] Wikipedia namespaces names are listed as Appendix L

[57] Printed out, it runs to over 20 pages of A4 paper.

[58] The transclusion of (part of) the 'HitRecord' article [W60] into the same-named section of the 'Joseph Gordon-Levitt' article [W76].

i.e. in the public-facing articles of Wikipedia. As described in section 4.3, the article's edit history shows it was slow to add a full description of the sectional transclusion (LST) when this became available.

A second shorter article 'Transclusion costs and benefits' [W688] describes the positive and negative aspects of transclude use, with—on balance—a slightly negative stance on transclusion[59]. The Costs and Benefits article links to the Wikipedia-namespace Transclusion article [W687] but oddly the latter does not link to the former, except hidden down in the page's 'See also' tailpiece link list [W694]. More appropriately positioned cross-linking with the 'Applications of transclusion' section would seem sensible scaffolding to help the learner.

The general tenor of discussion shows a misunderstanding of the value and use of transclusion and is further complicated by transclusion being a sub-set of more general 'template' use. There is evidence of a noticeable bias against transclusion partly because it removes (perceived) control, i.e. ownership of content. The 'Transclusion costs and benefits' [W688] article's edit history shows much of its material dates from this same time, when clearly there was a mood against transclusion. In the early days of Wikipedia it seems there were also concerns as to the performance effect of transclusion-heavy pages, noting that in Wikipedia terminology this relates indistinguishably to script and template use as well as more straightforward inclusion of actual content.

One of the listed examples in 'Wikipedia:Transclusion' [W693] shows that in 2008, the English Wikipedia article 'Pathology' was transcluding sections of articles on discrete sub-areas within the subject of Pathology. The 'Pathology' page no longer uses the partial transclusion in the cited example, but the erstwhile version can be seen in past versions [W134] from that time and viewed more clearly in edit view [W135].

At the same time (2008) there is extensive discussion in the talk pages of 'Wikipedia:Summary style' [W422] and 'Wikipedia:Subpages' [W420] about the issue surrounding transclusion of 'lede'[60] paragraphs. There are signs of clear disagreement as to the validity and worth of transclusion. Previous talk page comment for the 'Wikipedia:Summary style' page from 20 November 2007 [W421], gives a good summary of issues surrounding re-use of sub-article content via transclusion.

A year later, comment on the Pathology article's talk page indicates there is no settled view on the matter of content transclusion:

> "The use of lead transclusions for section summaries has led directly to this article being delisted as a GA[61] for lack of sources. This tequnique [sic] makes it much more

[59] Last edited in September 2015, the actual last significant content edits date from 2008. Given changes in scale and practice the article is essentially out of date.

[60] A 'lede' is Wikipedia vernacular for the lead(ing) paragraph(s) of an article, which are essentially an overview or summary of the article.

[61] GA is Wikipedia vernacular for a 'Good Article' [W653].

> difficult to improve articles, as an editor would have to source the entirety of all the subarticles to ensure the lead does not summarise anything uncited, or rewrite the leads so that instead of summarising the articles, they only cover whatever is cited."
>
> (Comment by user 'YobMod' 20 July 2009, 13:26 [W169])

The editor's comment uses makes a valid point about lede aggregation but uses a reverse justification that unfamiliarity and poor assistive affordances, but it is clear that the editorial kudos of close identification with FA and GA articles far outweighs interest in better structured hypertext, including improving the means to deploy content transclusion. This discomfort with the use of transclusion, taken along with the lack of any mention in Wikipedia's style or writing guides (e.g. 'Your first article' [W722], 'Manual of Style' [W659], 'Manual of Style/Lists' [W660], 'What is an article?' [W705]), there is little information to be found by editors to encourage them in deliberate use of transclusion.

Although a Wikipedia policy disparages 'ownership of content' [W663] the comment quoted above shows a seemingly proprietorial concern by an editor about the impact upon their own edits by the work of *other editors* (via transcluded content). This exposes a conflict at the heart of the collaborative method. It is also worthy of note that whilst there are guidelines for Wikipedia 'good' (GA) [W653] and 'featured' (FA) [W646] articles, the judgment process awarding these accolades is a human decision rather than a computed one.

On the 'Transclusion costs and benefits' page's 'Social' section, the first [sic] cost/benefit listed states:

> "There is a social cost of transclusion, the total expectation over time of the risk that a transcluded template page may be vandalized." ('Social' [W689])

This observation surfaces another of Wikipedia's concerns over editors' behaviour, that of vandalism (and 'edit-warring' [W643]). The emphasis on vandalism is interesting as, for the editor seeking to improve content, it is not a ringing endorsement of the general principle of transclusion: without decent contextualisation, such description will not encourage editors to try a useful technique. Depending on the parts of Wikipedia accessed, vandalism can appear non-existent or a constant threat to Wikipedia making a consensus view on the matter hard to achieve.

The problems of ownership-by-proxy, where engaged editors for a given article presume an accrued right of ownership and become overly defensive of their [sic] work, is shown in a dispute resolution from May 2012 [W639]. In this there is frequent mention of 'FL' and 'FLC'[62] showing the angst of editors' concerns that the efforts of others might rob them of the kudos of 'their' pages making the ultimate Wikipedia accolade of Featured Article status. This may be a case of territoriality rather than ownership, though the undesirable outcome is the same either way.

[62] These refer to the list of candidate list-type articles for 'Featured Article' status [W648]. FL and FLC refer, respectively, to the 'Wikipedia:Featured lists' [W650] and 'Wikipedia:Featured list candidates' [W649] pages.

The merits, or otherwise, of transclusion being poorly explained is picked up in the Wikipedia-namespace 'Transclusion' article's talk page. The latter has remarkably little discussion but includes an un-answered/debated section from 5 July 2014, 'Please add clearer real example text' which includes the comment:

> "After some reflection, transclusion sounds like a great tool. The clearer we can make the explanation of the benefits, function, and use of the tool, the more it will be used. I was able to eek out some understanding of this article and the concepts, despite the ″foo″ examples rather than because of them. It was frustrating to be deprived of clear intuitive examples. I wanted and needed examples whose inner logic supported and illustrated the points being made.
>
> "The word ″transclusion,″ the concepts of transclusion, and code to adeptly accomplish transclusion are not general knowledge. Transclusion is a computer science concept, so little known as to be marked as a spelling error by my dictionary as I work in Wikipedia, and having only 170,000 Google hits in total. There is a real challenge to making transclusion clear and salient enough to be enthusiastically adopted by the great majority of Wikipedia editors, who are non-programmers."
>
> ('Please add clearer real example text' [W424])

The Transclusion article's 'Applications of transclusion' section [W690] links to a long discussion, albeit from 2005, on 'transcluding prose' [W423]. Pertinent to this is that Wikipedia—at least, the 'en' language wiki—had just moved to suppress sub-pages. Clarification occurs in a revision of 'Wikipedia:Subpages' of 7 June 2006, 15:32, which adds:

> "Except in "main" namespace (="article namespace"), where the subpage feature has been disabled in English Wikipedia…" ('Wikipedia:Subpages' [W679])

Although the background rationale for the change is not explicitly documented, parsing from talk page discussion it appears it was felt sub-pages added implicit hierarchy whereas some articles naturally supported a number of higher level articles, i.e. had no single obvious parent page. Wikipedia's Categories were held to be the resolution to the possible organisation disconnect of moving sub-pages onto a par with main articles. The Category[63] feature allowed editors to cross-associate articles, with implied hierarchy, on other than a one-to-one basis.

Some pre-existing sub-pages may still be found in the English Wikipedia but creation of new ones is deprecated (as above). For reasons unclear, sub-pages [W640] appear to still be valid in other language versions of Wikipedia; this may be due to their smaller size and relative lack of complexity.

[63] Not commented upon is the fact categories are effectively a folksonomy. Anyone may create a category. Subsequent category maintenance is variable in scope and regularity (high traffic subjects get better scrutiny).

Another aspect of editors attitudes to transclusion is evident in the 'transcluding prose' [W423] article, with two groups talking past each other. One group seeks to explain the point of transclusive re-use whist the other takes more absolutist view that every page is an article - or something en route to being an article. Neither view is wrong per se, but the latter shows a lack of understanding of the *hypertext* beyond it being a bundle of discrete fully formed articles)as in paper documents) which happened to be interlinked. The latter externalises, to 'somebody else' all the work of maintaining coherence of the whole. Considered transclusional use should be net positive for the overall health of the hypertext.

The Wikipedia-namespace 'Transclusion' article's section on mark-up has an easily- overlooked link to a Help-namespace article on Templates [W56]. Although this link is in relation to partial include mark-up, the overall page indicates that from a mediawiki perspective Wikipedia transclusion is a subset of more general template transforms. Whilst this is true in functional, i.e. code, terms it still does not set out the case for (content) transclusion use.

Linked from the 'Using the labelled section method' section is the Help-namespace's article on 'LST' [W52] which explains how a transcluded (i.e. target) page can control the degree of data returned by the inline mark-up tag `<section/>`. Note that this is not an HTML element and unrelated to the HTML5 `<section>` tag. Whilst the Wikipedia `<section>` pre-dates HTML5, now that it is the current HTML standard the effective naming collision offers scope for confusion.

An additional link takes the reader to the MediaWiki-namespace for another article on 'LST' [W116]. This page reveals that sectional transclude-related code can be localised. Although a few such examples are given (e.g. those for German, Hebrew and Portuguese) no general reference is offered to allow per-Wikipedia-language look up of localised codes. Indeed, nowhere in the many Wikipedia namespaces is any form of localisation table for namespaces and other commonly used terms like 'disambiguation'. Given the degree of effort in cross-linking different language versions of Wikipedia this seems a strange oversight. Despite localisation support, the canonical 'en' names can be used in any language alongside or instead of localised terms adding further confusing choice. These pages do not link to the previously mentioned 'Help:Section' [W54] and 'Help:Template' [W55] further fragmenting the available documentation[64].

All Wikipedia articles have a special-namespace page 'What links here'[65] which summaries all (or up to 5,000) articles that link to the page via 3 methods: transclusions, links and re- directs[66]. There is no visual indication of each type in the listing but each type of link can be filtered out independently. Thus hiding links and re-directs will filter to reveal actual transcludes. But for heavily linked pages, the 5,000 item cap is likely to yield false results as not all data is being viewed.

[64] Further discussion of transclusion occurs in 'Help:Template' [W55] and in 'Help:Section' [W54], though this is not well signposted from other documentation.

[65] 'Special: What Links Here: Trace Amine' [W164] is transcluded by 'Monoamine neurotransmitter' and 'Biogenic amine'.

[66] Already described in section 4.6.1.3 above.

Subsequent additional analysis of the 2016 'en' wiki XML data set, showed that 2,578 discrete articles in the 'en' wiki transcluded 8,758 calls to Wikidata. De-duped, these called 169 discrete Wikidata properties. Resolving parameters on some calls, this further resolves to 95 discrete targets being used by 2,578 Wikipedia articles. This somewhat subverts the notion that Wikidata is not being used, though Wikidata use in other sampled wikis was lower except for the 'fr' wiki where 2,720 had 17,116 transclusions of 122 discrete properties (119 discrete properties if additional parameters are ignored).

Even the present 'en' wiki Wikipedia namespace guidance article on 'Wikidata' [W708][67] concentrates mainly on Wikidata's use to replace old Interwiki links [W61]. Relegated to the bottom of the page is information about 'Inserting Wikidata values into Wikipedia articles' [W710]. This is a usage still not really encouraged, judging by the 'Consensus' section at the foot of the article 'Use of Wikidata in Wikipedia' [W697] (the article was only created in September 2017[68]).

The latter indicates indicates a 'chick and egg' problem for a community like Wikipedia. Among general editors there appears to be a lack of knowledge about, or dislike for, hypertextually positive structural processes like transclusion and use of structured (semantic)data sources. This results in a lack of use the process and of documentation about how they work, or even that their possibility exists. The minimal use means factual knowledge—rather than opinion—about the technical and editorial (user experience) is not gathered so the editor base is not informed, and the loop is complete.

4.6.2 Discussion

The research tasks were all successfully completed. Although transclusion of articles is present its use is neither widespread nor coherent. It is unclear from information contained in Wikipedia as to why this should be. However, it is clear that there is no encouragement to editors to use the technique and the documentation is fragmentary and not coherently linked. It is certainly not made accessible to an editor wishing to learn about transclusion.

4.6.2.1 Evidence of Deliberate Use of Content Transclusion

Of the 10 languages sampled, all but 'ceb' show some evidence of deliberate use of transclusion. Despite its high ranking in article count, the 'ceb' wiki is largely bot-created and bots are likely not yet programmed for more sophisticated behaviours such as transclusion. Thus the absence of transclude markers is not seen as significant, except as a reflection on the capabilities of article-creating bots. In each of the 'de' and 'it' Wikipedias, transclusion was primarily employed within a single large disambiguation task. Only 'en' shows evidence of a wider ranging use of the

[67] Edit as at 18:03, 4 November 2019.
[68] See edit as of '06:31, 28 September 2017 [W698].

available transclusional mark-up (albeit some in very limited numbers). Three sectional transclusion mark-up patterns (`#lstx`, `#section-x` and `#section-h`) were not observed at all. Overall, the occurrence of transclusion observed was much lower than might be expected given the low penetration of categories where use occurred. It cannot be stated with confidence that there is general deliberate use of transclusion. Rather, it is clear that there are some editors are doing so and with a likely halo effect as others copy their work (even though the underlying methods used may not be understood). A further inherent effect of the latter is to create different communities of practice within (and across) wikis. If localised occurrences of use, or communities of good practice, can be routinely detected and fostered this could help increase use of actions beneficial to the hypertext as a whole. Another by-product of watching for such practice will be the opportunities for harmonisation and transfer of skills but within a wiki but between different language wikis, as looked at in the next aspect of the research.

4.6.2.2 Variability of Content Transclusion Within Per-Language Wikis

The small degree of occurrence of transclusion makes it difficult to evaluate this in a meaningful manner. Of the three languages analysed (en/de/it), RQ1.2 it was found to be true insofar as there were definite groupings of transclusion within certain subject areas. However, the degree of penetration within subject areas was still minor. This may reflect the fact that some subjects lend themselves to hierarchical organisation in a manner so as be natural targets for transclusion. The patterns that are seen could certainly be applied to a wider range of subject areas, were there the interest so to do.

4.6.2.3 Variability of Content Transclusion Between Per-Language Wikis

On the evidence of the above analysis, RQ1.3 is held to be true but with the same proviso as above that the degree of transclusion use is small within the overall corpus of each wiki. Different languages analysed do not repeat the same explicit patterns. However, underlying patterns such as disambiguation are repeated albeit using different transclusion- scoping tags. Only the ‘en’ Wikipedia displays any real range of use so comparisons here are necessarily conjectural.

4.6.2.4 Patterns of Content Transclusion

A number of patterns can be seen, albeit in very small numbers. Where patterns do occur they are not pervasive within the scope of their possible deployment, e.g. across a whole subject area or category. Instead, their use shows small coverage in the topic areas within which they are found, implying there is scope for greater deliberate use of transclusion.

4.6.2.5 Finding And Tracing Content Transclusion

There is no well-organised documentation, and what there is not found with any ease. Transclusion is described, as a process, but the documentation is fragmented and not linked with the learner editor or hypertext maintainer in mind. There are no references to transclusion in the Wikipedia style guides—positive or negative. There are also no good worked examples for learners. The latter are important as they allow copying of code and learning via implementation. Expecting non-technical editors to understand and implement the process of transclusion within their editing without such support seems unreasonable, nor should the logical fall-back be a need for more live (human) support; the hypertext should self-document to a degree such that having to ask another person is not required to implement process.

The lack of reporting tools for transclusional links creates difficulties for the editor trying to edit or re-factor clusters of transcluding articles as lack of coherent implementation means such linkage is affected by code in both transcluding and transcluded articles. It is also not easily possible to tell what parts of a (rendered) page are transclusion-related. This would suggest the need for some sort of editing interface for reviewing transclusions.

4.6.2.6 Attitude of Wikipedians to Content Transclusion

There is little clarity to be seen here. Whilst the articles above do combine to give a summary of how (article content) transclusion works, it takes some initiative and persistence to get a proper understanding of the details of how transclusion is implemented. The documentation is poorly interlinked from the perspective of understanding use of transclusion, and out-dated examples indicate a lack of ongoing review and maintenance of the documentation. The guidance available on how to actually employ transclusion usefully within Wikipedia is less effective and the general tone of advice balances out as slightly negative, as read on the page, towards use of transclusion.

Although this guidance may already be out of date it has not been corrected or updated as at time of sampling (January 2016). This leads to a supposition that any current transcludes are either the work of a skilled rather than a casual Wikipedia editor, or that the editor may simply be copying code patterns from other articles.

Issues of (perceived) ownership or territoriality over content may be a factor here, but with the limited overall use of transclusion, it cannot confidently be asserted as a definite causal factor. That notwithstanding, encouragement of appropriate deliberate use of transclusion should take these hidden boundaries into account lest they become a bigger issue. This aspect of editor interaction is explored further in Section 6.2.

Transclusion is not a common word nor is the concept simple at first encounter by the casual editor. More thought needs to be given to explaining the underlying concepts as the value of the concept is unlocked by the volume of its use.

4.6.3 Emergent issues from studying content transclusion in Wikipedia.

Although it is possible to view past revisions in 'live' within the current wiki, any transcluded content still uses the *current* version of the latter and not the version as at the revision (date) when the transclusion was added. This requires the reviewer to check for interim changes in the transcluded material; the general absence of revision comments or inline Hypertext Markup Language (HTML) comments is unhelpful.

- Code modifying transclusion scope within templates was hard to quantify in the analysis method used here for two reasons. Firstly, the degree of nested tasks within a template is not self-evident without deep inspection. Secondly, the degree of use of a given template requires additional iterative parsing to find which documents used templates effecting or modifying transclusion. Neither of these tasks was performed in this analysis.
- Renaming and redirects: for reasons not documented, some past revisions of pages are essentially unavailable owing to (probably unintended) effects of redirections. Whilst the latter are pertinent to the current revision, re-direction of URLs to revisions before a page was renamed and/or updated would be a better way to support maintenance.

4.6.4 Further Study

At this stage research norms might imply next step would be a user study. But in that case it would provide the appearance of diligence at the expense of rigour, as the role of non-human actors in the hypertext structure is also unrecorded. Thus, with the time available, I think it more important to investigate in non-human actors, i.e. are bots are back-filling the failure of human actors in the research context? Without properly exploring the editing environment, rushing into user interviews whilst interesting gives no firm background upon which to test the users' statements.

4.7 Summary

It would be too easy to concentrate on the perceived negatives from the analysis outcome, in terms of low levels of transclusional use. Content transclusion actually only occurs in a tiny proportion of articles (<0.1%) so any failings have minimal effect on the encyclopædia as a whole. Yet, despite the weaknesses described in the conclusions above, most of Wikipedia functions adequately nonetheless even if it does not offer a good model for other collaborative hypertexts.

However, what the research analysis does reveal is that there is little evidence of *deliberate* attempts to maintain usable, up-to-date guidance documents and which reflect current best practice and style guides. Co-ordination tends to be apparent at 'project' (subject area) level, though projects themselves are essentially ad hoc. Wikipedia roles (see Appendix D), regardless of their

names are actually defined by the scope of change they may effect on the hypertext. These roles thus represent a technocratic approach aligned more to the technical means of producing articles than supporting a hypertext as a whole. There are no roles that look *across* the hypertext as collection and engage with it as the medium containing the collective knowledge.

Chapter 5

Bots

> '*Can the maker repair what he makes?*'
>
> Roy Batty, to Eldon Tyrell in *Blade Runner* (film, 1982)

Here, I will explore the use of bots specifically within Wikipedia English-language wiki, with the wiki acting as a publicly-accessible example of Web hypertext. This also allows the findings to be considered in context with my investigation of transclusion use in Wikipedia (Chapter 4). The general concept of 'bots' has already been explained in Chapter 2, Section 2.10.

Within a socio-technical system it is pertinent to consider how the technological element may be supporting and indeed increasing the activities of human actors. When looking for evidence of deliberate human effort in structuring hypertext (in the previous Chapter 4), it was noted that some editing activity in Wikipedia involves bots. If deliberate use of transcluded content by human editors gives insight into consideration of hypertextual structure in the Wikipedia community, analysis of the use of software agents and automation may show a more indirect form of socio-technical co-creation in support of the hypertext.

Put another way, could it be that bots are doing these tasks so humans do not have to? A lack of human effort maintaining the hypertext's structure *might* simply hide the fact that non-human actors are carrying out the role. In terms of the knowledge residing within the hypertext, this research follows the approach described in Chapter 4.1 in that it looks to see how the overall structural health of the hypertext is being stained by bot activities. The intent is to ascertain if bots are filling gaps in this role left by the work of human editors.

Thus, the research question (RQ2) investigated in this chapter is:

> *Is there evidence of automated 'bot' processes being used to support the hypertext quality of Wikipedia?*

The transclusion activity previously described in Research Question 1 (see Chapter 4) lent itself to pattern detection and thus the possibility of investigating a number of discrete language wikis. By comparison, early investigation for this task showed that much closer reading of source data would be needed, requiring good comprehension of the language in use, as otherwise nuance is easily missed. Given the incomplete nature of much of the logging and documentation relating to bots, pattern-based classification would not be possible. Therefore my investigation at this stage was conducted entirely within the 'en' wiki of Wikipedia.

Before being able to determine the contribution and role of bots in Wikipedia, it was necessary to have a clearer understanding of how bots came into use in Wikipedia, how they are regulated within the website and how bots accounts and their status may reliably be identified, as is explored in Section 5.1.

Having identified more clearly the nature of the wiki's bots, the above research question was split into two sequential parts:

- RQ2.1: *Can bot activity within Wikipedia be identified and characterised?* This looks at the sorts of tasks bots are set up to perform, and the manner of their execution of said tasks.
- RQ2.2: *How do general users of Wikipedia access bot services?* A different aspect of bot use is how the wider Wikipedia community is—or is not—able to task and employ bot services (or are they simply the private tools of the bot operators?).

The two questions were investigated sequentially as the latter drew on an understanding of the 'en' wiki's bot ecosystem arising from investigation of the first question.

As in the previous piece of research in Chapter 4, the incomplete nature of documentation and record-keeping within Wikipedia meant my work was necessarily an iterative exploration of emergent structure. A cyclic review of existing data was needed for each new discovery: as a new bot would be uncovered, I used the means of its discovery to review existing data sources and thus would often uncover yet more incorrectly documented bots. Frustrating as this may be from the researcher's standpoint, it was important to remember that the gaps in the documentation were as much a by-product of the evolving Wikipedia platform as due to any deliberate human editor's action. This is reflected in my observation of a number of false starts at applying a new process or documentation of the period of analysis. But, when a stable solution emerges, the technical debt of reviewing the false starts and deleting them or re-linking data tends to be left unserviced, doubtless in part due to expediency—there being more important calls on the efforts of those (few) editors with the skill to effect the fixes needed. This aspect of the wiki is a point to which I will return and consider in more detail in the following Chapter 6.

As a direct result of incomplete nature of the wiki's records, a significant amount of precursor research and analysis was necessary (see Section 5.1) before I was able to commence work on the actual research questions described above. Early reconnaissance of the evidence of bot activities

indicated that simply taking a sample at a single date would provide an inaccurate picture of bots' contribution in the round[1].

As bots have varying roles, patterns of activity, and (active) lifespans, I therefore felt it appropriate to undertake the more complex task of a longitudinal view to identify all bots from first occurrence (in 2002) up to the then current time of mid-2017. But, without this underlying rich listing of bots, it simply would not have been possible to get a clear picture of the range of tasks to which bots were being applied (RQ2.1) or whether the bots' efforts were helping to do what the community wanted done (RQ2.2).

With that overarching narrative in mind, I will now describe the broad structure of this Chapter. Therefore, I start in Section 5.1 by describing what a 'bot' is within Wikipedia. This is expanded by a description of early bots and the evolutionary process leading to current approval and regulatory structures for bots. Building on this I give some overview of the primary documentation resources located that relate to bots (Section 5.1.6.1).

Following on from that, Section 5.1.7 I will describe some of the issues relating to provenance and research that I uncovered whilst tracing bot accounts within the 'en' wiki', from first occurrence through to a cut-off date of 30 June 2017. If such detailed annotation appears tangential to the actual research questions, it should be noted that not only was such detail necessary not only to locate more than the easily finable bots but also to understand the nature and degree of inconsistencies in Wikipedia's records of editors work—both human and bot. This was particularly pertinent to work on RQ2.2.

Having identified all bots accounts, in Section 5.2 I will address RQ2.1 and describe the range of tasks undertaken by bots within the 'en' wiki. Building upon this, in Section 5.3 I report on RQ2.2 through the lens of user requests for bot support: the range of tasks and the degree to which requests were taken up by bots, or rather by their operators. As with the detailed diagnostic work described at Section 5.1.7, deconstructing the picture of requests for bot support was also somewhat complex as is mapped out in Section 5.3.1. There was temptation—even if only for concision in the thesis' main text—to hide this in an appendix or simply not include the detail. However, without it, the reader will gain a false picture as to the (lack of) coherence of Wikipedia data, an aspect I will return to in the next Chapter.

The chapter closes with a brief summary as a deeper synthesis of the findings here and in RQ1 will be undertaken in the next chapter (Chapter 6).

[1] Reading the 'list of Users with a 'bot' flag' [W147] on 02 August 2017, it listed 296 discrete user accounts with bot flags. In contrast, further research showed over a 1,000 bot accounts have been active for periods prior to this date.

5.1 Bots within Wikipedia

Within Franklin and Graesser's taxonomy for autonomous agents [127], bots used in Wikipedia are 'software agents' of the 'task-specific' type [127, Fig.1, p.31] although there is much variation in both the scope of their activity and their degree of autonomy.

At their most basic, bots are scripts made available for re-use, and run by adding the script to a user's wiki account. At their most complex, there are bots with multiple maintainers running on clusters of computers, such as are used anti-vandalism work. Notwithstanding this, Wikipedia is certainly less than clear as to roles and types.

The nature of the open contributive structure of Wikipedia means the site has many partly complete, and likely abandoned, articles such a Wikipedia:Types of bots [W695]. Thus I found that in almost every aspect of this research task there was no complete or canonical source of information.

The simplest description of a bot, in terms of its role as a non-human editor of the wiki, is that a bot account is a user account that is used set up to do automated tasks within Wikipedia. Such a bot might find pages that meet a certain query, or read items from a pre-made list of Wikipedia pages[2] that it will then process according to predefined code rules. The pinnacle of bot use within Wikipedia, as described above, is the sophisticated bots that fight vandalism in various forms (defacement, content deletion, 'sockpuppetry', etc.). Conversely, at the lowest level, bot activity may involve no more than the automated invocation of a single small script stored at a single central location.

Wikipedia's Meta-wiki describes a 'Bot' [W119] as an "automated software program used to perform certain repetitive tasks on a wiki". It further notes that bots "are occasionally used by vandals to cause extensive damage to a wiki's content in a short period of time", which seems to portray a certain degree of bias towards the role of bots. By comparison, the 'en' wiki's article 'Bots' [W580] uses the description of "an automated tool that carries out repetitive and mundane tasks to maintain the …pages of the English Wikipedia". Its rider is more accommodating than the Meta-wiki's: "Bots are able to make edits very rapidly and can disrupt Wikipedia if they are incorrectly designed or operated.".

To give a more fulsome picture I quote from Halfaker and Riedl's 'Bots and Cyborgs: Wikipedia's Immune System' who describe bots thus (I have added the inline references):

> Robots perform a wide range of activities on Wikipedia. These include injecting public domain data, monitoring and curating content, augmenting the MediaWiki software, and protecting the encyclopedia from malicious activity.

[2] The list itself may well have been generated by another bot or a script.

> The first bots injected data into Wikipedia content from public databases. 'Rambot' [W333], widely accepted to be the encyclopedia's first sanctioned robot, inserted census data into articles about countries and cities. Rambot and its cousins act as "force multipliers" by performing repetitive activities hundreds or thousands of times in minutes.
>
> Many other bots monitor and curate Wikipedia content. For example, 'SpellCheckerBot' [W362] checks recent changes for common spelling mistakes using an international dictionary to prevent accidental "fixes" to correctly spelled foreign words. Similarly, 'Helpful Pixie Bot' [W264] corrects ISBNs and other structural features of articles such as section capitalization. The largest class of content curators is interlanguage bots, which use graph models of links between different languages of Wikipedia to identify missing links between articles covering the same topic in a different language. As of this writing, the English Wikipedia has more than 60 active interlanguage bots.
>
> Some bots extend Wikipedia functionality by implementing features that the MediaWiki software doesn't support. For example, 'HBC AIV Helperbot' [W262] turns a simple page into a dynamic priority-based discussion queue to support administrators in their work of identifying and blocking vandals. Similarly, 'SineBot' [W350]SineBot ensures that every posted comment is signed and dated.
>
> Finally, a series of bots protects the encyclopedia from malicious activity. For example, 'ClueBot NG' [W240] uses state-of-the-art machine learning techniques to review all contributions to articles and to revert vandalism, while 'XLinkBot' [W383] reverts contributions that create links to blacklisted domains as a way of quickly and permanently dealing with spammers.
>
> (Halfaker & Riedl, 'Bots and Cyborgs: Wikipedia's Immune System' [159, p.80])

Whilst a 'bot' is a user account in its own right (bar in a few edge cases) each bot account is required to have a human 'bot operator', with a separate account, who is responsible to the Wikipedia community for the orderly operation of the bot. Both bot and human users have personal user pages in the same 'User' namespace (see Appendix L). This may make it seem like the bot operator has two accounts, but the operator is charged with not using the bot as an alternative account outside the approved tasks for the bot: this restriction is backed by potential sanctions again both bot and operator accounts. In a few special cases there may be multiple operators of a bot account[3] but this is rare and generally occurs where bots have large and complex tasks requiring considerable resource.

A bot account can be censured for its editing behaviour, being blocked or banned, but depending on the nature of the offence the bot operator's personal account is not necessarily affected. Particularly severe cases my warrant both bot and operator being disciplined.

[3] There are only six instances in over a thousand recorded bot accounts.

5.1.1 Determining Bot Status

So what makes a user account into a bot account? At a technical level, once approved, a bot account is given an extra permission 'flag, within the wiki, otherwise described as a 'role'. Roles, and their interrelation with privilege levels within Wikipedia are detailed in Appendix D. Once flagged in this manner, the account is identified to the system as being a bot. This status allows it extra freedom of action compared to an ordinary user and exempts it from certain controls. Primarily, these relate to the rate of making edits and the manner in which the bot's edits are, or are not, logged in an article's edit history.

As so often with system-wide rules, there are exceptions. In this context, some accounts are authorised to exhibit bot-like behaviour yet *not* be given a bot flag. There is no formal list of such exemptions, but a typical example is a bot doing data extraction from Wikipedia thus reading data *out* of the site at a high rate but pertinently, not making any *edits*. Compared to the activities of a more general user, a bots high-speed queries or edits can place disproportionate strain on Wikipedia's servers. I found no documentation, but it is assumed such approved-yet-un-flagged accounts are still identified on some (hidden) approved-access list so as to not be affected by normal defensive controls against high-speed access.

Though this 'bot' flag method of denoting bot status dates from the time of the earliest bot use in the wiki, the genesis of the bot flag is not documented, having arisen to deal with the emergent issue of bots and the need to differentiate them from actual human users and from being mistaken as a form of vandalism. The lack of a defined process has had the result that early records of bot authorisation within Wikipedia are fragmentary or missing altogether. Only more recently, since circa 2008 did the process become mature enough to log bot approval—with or without a flag—in a consistent manner.

Wikipedia's lack of attention to detail in logging significant system events—like the granting of flags—has proved a significant impediment in trying to identify all bot accounts. The cause of these failures is hard to uncover at this distance in time. However, it but appears to be a human failing as much as a system failing, even if only through a failure to design a documented process whereby significant system changes were automatically logged rather than relying on (fallible) humans actors to manually record such matters.

Initially, the mechanism of creating bot accounts was quite informal but, as I will describe below, it subsequently evolved into the current Bot Approvals Group (BAG) control group and the Bots: Requests for Approval (BRFA) process for pre-approving a bot before it is granted a flag.

From an interview with a human bot-operator in Geiger's 2014 paper [136, p.347], it might appear that it is simple to discern if a user page is a bot or not:

> "If you want to know about the life of AfDStatBot, you might start with the bot's user page on Wikipedia. If we think of AfDStatBot as a user, this user profile best represents the bot. Profiles take many different forms, but they have long been the

> spaces in which software platforms represent users. In Wikipedia, these 'user pages' are technically little more than special wiki pages, meaning that anyone can edit anyone else's profile—as I typically did for AfDStatBot. Like most bot userpages, there are several curious elements that indicate this profile is different from that of most human Wikipedia editors. There is a large, red power button that administrators can use to stop the bot from operating if it begins to malfunction. That's not a formal requirement, but it is standard practice in Wikipedia bot development. There are also warning templates at the top that tell the reader that the account is a bot operated by me, and that it is a 'legitimate alternative account'—a notice that is somewhat of a legacy from time when bots were far more contested and controversial in Wikipedia." ('Vignette two: a legitimate alternative account' [136, p.347])

Indeed, AfDStatBot [W218], the bot described in the quote above, is well-defined in terms of most bot page attributes. However, this paints a rather false picture as will be seen by the discovery process I describe below, as it transpires that such well-defined bot user pages are in the minority. Whilst it is generally the case that looking at a user page will allow an educated guess if the user is a bot, a bot's status is far less easy ascertained.

5.1.2 The first Wikipedia bots

Most current bots may generally be located by a valid bot flag, but locating earlier bots requires some sleuthing, especially if investigating some 15 years of activity within a wiki sustaining constant evolution and growth.

So, from whence did Wikipedia's bots arise? A slim article on the history of bots [W655] indicates the earliest bot-like activity occurred at 05:18, 16 February 2002, with the scripted uploading by user 'The Anome' [W374] of data relating to 'Federal Standard 1037C' [W24]. Though automated in part, the upload occurred via the user's account and had significant human assistance during and after the process. It would therefore appear the first use of a bot in the 'en' wiki was for automated addition of article data, i.e. creating and populating stub articles.

The first true bot *user*, i.e. a bot with its own discrete user account, is generally held to be 'Rambot'[4] [W333] which started operating at the end of 2002. The account was created by user 'Ram-man' [W332]. Although the code used by Rambot may have been run previously under the account of Ram-man, on 04:21, 9 December 2002 Rambot made its first edit to the wiki's main namespace, creating an article on the town of 'Autaugaville' (in Alabama, USA) [W4]. Initially, adding pre-made content for US counties and towns not yet in Wikipedia, the script could also re-edit articles to update certain metrics. The account made 131,309 mainspace edits [W728] (94.9% of all its edits), ending 28 December 2004 [W159].

[4] Wikipedia user account names are always used with a capitalised first letter in queries and links, regardless of the actual user's preferred choice of case. Otherwise, Wikipedia usernames are case-sensitive *except* for the first character.

The award of a 'bot' flag (or 'bot' group membership) for Rambot was never recorded, but the flag's removal *is* logged in 2014 [W152], just under 10 years after the bot last edited. This missing data is an example of the aforementioned lack of proper record keeping within Wikipedia at the time.

The mediawiki page on 'User rights' [W113] *implies* the 'bot' rights flag, and user group, were added to MediaWiki software in Version 1.5 [W114] June 2005. However, an earlier original of the page 'Wikipedia:User access levels' [W700] from 19 November 2004 already lists a 'bot' flag.

Wikimedia's Metawiki, which handles global permissions, first shows a bot flag request in April 2004 [W125]. It is therefore unclear as to what the term 'flag' truly describes in this context, though it should be bourne in mind that a formal notion of bot accounts on Wikipedia was still a rather new idea at the time. It is also the case that a custom flag had been added to the wiki that was later formally adopted by the wiki's source mediawiki software.

In an open, collaborative, hypertext like Wikipedia the genesis of a newly formalised process can often precede any useful documentation. This is because at outset, the new process is still in flux and probably only known to the most expert or most frequent users. Despite that, as already noted, in a well-rendered hypertext, consideration should be given to not avoiding the technical debt thereby arising.

5.1.3 The Evolution of The Bot Request for Approval (BRFA) Process

The BRFA process is the core of the current bot regulation and requires operators of new bots (or new tasks for existing bots) to provide information about their bot as part of their request for approval. A small amount of leeway is given for initial testing ('Valid operations without approval' [W453]), but bot operators are now expected to make a BRFA submission before undertaking work of scale or significance.

Before formalisation of the BRFA process and the BAG, aspirant bot operators would simply announce their intent to run and ask for consensus approval. There was little more formalised control. Examples of bots with no traceable BRFA start in 2002 with 'Rambot' [W334] . It is not until 18 April 2004, that there is the first logged evidence of granting of a bot flag at the 'Wikipedia stewards requests' page [W677] for an interwiki[5] bot 'Bgbot' [W236], though that user page's history has no edit to reflect this fact. The bot was created on 7 April 2004, had task information added on 17 April 2004 (a day before flag grant) and then has no user page edit until 5 September 2005 when newly added text marks the bot as inactive. The 'en' wiki logs do not record the flag grant but do record flag removal in March 2008 [W399].

This shows that in the earliest years of bot use, up to 2006, bot approvals were seemingly given (and occasionally logged) by stewards [W652] before being deferred—at least in larger wikis

[5] See Interwiki links [W61].

like ‘en’—to wiki administrators and eventually to the BAG (see Section 5.1.4 below). It was during March 2006 that the BRFA process began in the ‘en’ wiki and its original BRFA page was created on 13 March 2006 at 18:27 [W582]. The first recorded bot request on that page occurred shortly thereafter at 21:35 on the same day, for approval of ‘ShinmaBot’ [W592].

The first actual recorded approval of a BRFA is noted just over a week later at 00:03, 23 March 2006 for ‘Pegasusbot’ [W595], albeit with the bot flag grant still being made centrally [W124]. Thus it is not until 19:36 on 22 April 2006 that a bot ‘Zbot370’ [W386] is both approved and flagged completely within the auspices of ‘en’ wiki:

> “19:36, 22 April 2006 Linuxbeak (talk | contribs) changed group membership for Zbot370 from (none) to bot (Congrats, you are the first bot promoted on the English Wikipedia)” (‘Zbot370’ [W149])

Since this event, with the exception of global bots [W146], management of ‘en’ wiki bot approvals has occurred within the auspices of the wiki.

Early BRFA submissions were free-form text, but within 6 months, a structured set of headings for new requests evolved. These made clearer to BAG reviewers the nature of a proposed bot and its task(s). The first BRFA to use the current structured layout of BRFA headings is a proposal for ‘Cronbot’:

> “14:29, 6 September 2006 Omicronpersei8 (talk | contribs) . . (2,969 bytes) +49 . . (→Current requests for approval: Cronbot).” (‘Cronbot’ (edit at 14:29, 6 September 2006) [W596])

Subsequently, these structured headings for new requests were formalised further into a set of templates. New BRFA requests are now initiated using a form [W593] on the BRFA page, ensuring a degree of common structure to the data provided for new submissions. Although the bot operator creates the new BRFA, it is created in the name of the proposed bot’s account. The latter is the primary reason why there are bot-sounding account names for accounts that are not bots.

An overview of the BRFA requirements is also given in the Wikipedia article on the ‘Approvals process’ [W451]. The majority of bots have one (successful) BRFA but a bot may have many discrete BRFAs, especially if the bot is carrying out diverse tasks as proposed by general users of the wiki. The most extreme case is ‘AnomieBOT’ [W225], which has been active since September 2008 and was still active as the sample cut-off and which has had 73 discrete successful BRFAs (as well as two more that were withdrawn by the operator without approval).

5.1.4 The Bot Approvals Group

Currently, approval of bot status is granted by the 'Bot Approvals Group' [W442], more generally referred to via its acronym 'BAG'. The group was not formed until 2006, being announced in March 2006 [W444] alongside the new BRFA process as already described above.

Essentially, the BAG has delegated authority from the 'en' wiki[6,7] to regulate the use of bots on the that wiki. For a BRFA to be approved at least one current BAG member must assent to the proposal—or give approval modified as necessary to reflect editors' discussion of the BRFA. For this reason, all BRFAs are (or should be) archived regardless of outcome, as the BRFA then becomes a primary reference if there are subsequent issues over the operation of the bot.

There are four primary outcomes of a BRFA:

- *Approved*. The bot is approved to run. The approval process may authorise and require evidence of successful completion of requested preliminary trials for which the BAG will state explicit authorisation on the BRFA page. In addition, approval usually results in the bot account receiving its bot flag. In a few special cases the bot may be approved to run un-flagged at BAG discretion. In either case, approval may be made with specific listed restrictions to which the bot must adhere. For example, bot 'AAlertBot' [W597] was approved for a clerking task reporting page changes to interested parties.
    - *Revoked*. If an approved bot operates in an unacceptable manner—including the operator's activities *as relating only to the bot*—the BAG may revoke the BRFA. This amounts to an immediate cancellation of the existing BRFA approval *for that task—and that task only, if more than one—unless stated otherwise*. for example, approval for 'Andrea105Bot' [W598] was revoked after discovery that the operator was a "sockpuppet of a banned user".
- *Denied*. BAG members have not given their approval for the proposal. If the idea holds merit but cannot be approved in current form, if may be suggested that the operator revise the proposal and submit a *new* BRFA. Denial of a single proposal does not preclude subsequent proposals for the same bot, or different bots run my the same operator. For example: 'Helpful Pixie Bot' [W608] task request #46 was denied for "problems with scope, and a lack of civility in response to feedback that is utterly unacceptable for a bot operator.".
- *Withdrawn*. In this instance, the cancellation is at the behest of the *proposer* and not the BAG. Re-submission at a later date would require a fresh BRFA. Example: the proposed 'DeltaQuadBot' [W603] was withdrawn by the operator.

[6] English language wiki within Wikipedia.

[7] This also appears to hold true for most wikis of size within Wikipedia, whereby that wiki's users administer bot approvals. Note though, that there appears no pan-Wikipedia regulation on this matter.

- *Expired*. The BAG may so mark a proposal if it is adjudged to have become stale, e.g. if the operator has not responded to questions on the BRFA page, or not carried out trials as requested by BAG staff. As above, re-submission at a later date would require a fresh BRFA. for example: 'BHGbot' [W600] task request #3 expired as the operator has not respond to BRFA discussion points.

Any Wikipedia editor with an 'en' wiki' account may apply to join the BAG. Prospective members need to show a requisite understanding of bots and be active on the 'en' wiki. This is important, given that BAG approval is the last control measure before a bot is allowed to run flagged. A failure to understand the implications of a proposed bot, or ignorance of past relative information (e.g. the idea has tried and failed before), might otherwise lead to bots causing significant damage before corrective action could be taken. BAG members are auto-retired if they are inactive for over 2 years.

Thus, over time, the application for bot status has been regulated by the BAG and systematised through use of the BRFA (see Section 5.1.3). But, even with BRFA and BAG approval, the bot flag is not *carte blanche* for the bot operator who is expected to abide within the scope of task approval they have given. The latter article has this quote from a BAG member:

> "You can more or less think of a BRFA as being a driving license. If you drive without one, you can get in trouble even if your driving skills are fine. And if you have a license, it still doesn't give you the right to run over people or drive 200 km/h in a 50 km/h zone. BAG gives out these driving licenses. If they are abused, they can be (and are) revoked, and bots are blocked accordingly."
>
> (User 'Headbomb', quoted in 'Approvals process' [W451])

Though bot accounts are run by a human operator, the BAG has no direct disciplinary authority beyond policing the orderly use of bot approval and privileges. Misbehaviour by a bot's operator is subject to normal Wikipedia administrative procedures.

5.1.5 Discovering Bot Accounts And Their Operators

It might be assumed that having now clarified the means of denoting bot status, it would be a simple to retrieve a complete list of bots. However, after initial investigation I found further problems with Wikipedia's record-keeping. Yet more diagnostic work was required before the research dataset could be drawn from the overall Wikipedia corpus.

The primary reference for current bots is the report 'List of active bots' [W147] that is a built-in query in the 'Special' namespace and which is not archived or saved in XML dumps. Therefore it can only be read for the current time and cannot be used to study previous bots that are no longer active. The listing supposedly derives from user accounts that have the bot flag set but there

are inconsistencies. The cause of such differences cannot be resolved as the actual underlying query used to create the list is itself undocumented[8]. A further limitation of listings generated in the Special namespace is that they are limited to maximum of 5000 returned items and cannot contain more than 500 items per returned web page, as in 'Special:ListUsers/bot' (500 items per page) [W148]. Such reports neglect to list the total number of results and the reader must cycle through the pages and count manually. These constraints are likely pragmatic defence against overloading Wikipedia's back-end processes, but it would be useful to the users of the system if such design intents were recorded and appropriately linked.

Although the 'List of active bots' looks—at first sight—like a complete listing, it is actually incomplete data and a trap for the inattentive researcher. Viewed with the hindsight of having completed my research on Wikipedia bots this is unfortunately typical of such pages in the Special and Wikipedia namespaces. In addition, whilst listing pages outside the main namespace doubtless attract fewer eyeballs they still get vandalised [W696] for no apparent reason.

Thus the initial task of identifying bots became an iterative hypertextual search through the Wikipedia namespace pages, bot user accounts, talk pages. Essentially, each new bot that emerged was added to my existing listing and the corpus was re-investigated for more of the same type of exception. In part, this involved mapping out the Wikipedia documentation relating to bots and their activities.

5.1.6 Wikipedia's Bot Documentation

As part of understanding Wikipedia's judgment of which constitutes a bot account, it was also necessary to identify the main documentation of the subject within the 'en' wiki. This too, was less simple than might be assumed.

The 'en' wiki has a main namespace article on the general subject of bots 'Bots' [W10] but its description of bot operations upon Wikipedia itself is held in the 'Wikipedia' namespace at 'Wikipedia:Bots' [W580]. The latter page is the main page of a cluster of articles all relating to bot use and governance, described below.

5.1.6.1 Major Wikipedia bot pages

Many of these pages have an abbreviated (shortcut) name that can be used to quickly identify and/or link to that page. The shortcuts take the form of a prefix 'WP:' followed by several capital letters reflecting an acronym of the subject title. Where present these shortcuts are shown in bold at the start of the paragraphs below.

[8] This does not mean the query details could not be obtained, but doing so would involve direct contact with Wikipedia users which fell without my research methodology. In a well ordered hypertext, such information would we available because those maintaining [sic] the content would understand it might need to be viewed by others lacking the technical expertise to write their own API queries.

The totals of pages and threads given below in this section are recorded as at 31 July 2017, the end of the sample period. Most pages will have generated more (archive) content since that time. As will be seen, some of these pages have generated a significant amount of discussion and information over their life.

WP:BOTS. Wikipedia:Bots [W580], the primary page about bots. The associated talk page [W581] had 23 archive pages containing 492 discrete threads.

WP:BOTDICT. Wikipedia:Bots/Dictionary [W586]. This is a glossary of bot terminology used within Wikipedia.

WP:BOTPOL Wikipedia:Bot policy [W448]. A reference about bot-related policies. It was started at 02:49, 26 October 2002 by user 'Ram-man' [W332]. Its talk page [W408] has an archive of 167 discrete threads across 6 pages. The article defines activity requirements [W454], though the requirement was only added as recently as 15:13, 3 January 2017 [W449][9]. Although the section has moved around the page the policy is still as originally added. It states:

> "Bot accounts that have had no logged actions or edits for two years, where the listed operator has also had no logged actions or edits for two years, will be deauthorized. Following a one week notification period on the bot owners' noticeboard, and the operator's talk page, prior task approvals will be considered expired and bot flags will be removed. Should the operator return and wish to reactivate the bot, a new request for approval must be completed." (Activity requirements [W454])

Prior to the activity requirement being formally stated, there appears to have been an assumption of 2 years being a reasonable cut-off for judging moribund bot accounts. But, as will be seen from my analysis (see Section 5.1.7), monitoring of activity was not pursued with much vigour until 2011. Since then their appears to be a single annual weeding of defunct bot accounts - as opposed to any sort of automated process when the 2-year deadline falls due.

WP:BAG. Wikipedia:Bot Approvals Group [W442] is the core page of the BAG. Its talk page [W404] has an archive of 268 discrete threads across 8 pages. A sub-page of the BAG page handles BAG nominations [W446]. Prior to 2008, the BAG nominations were informal and never recorded. The current page lists 47 nominations of which 38 (80.8%) were successful. It also has an archive [W447]. First created 02:23, 26 March 2006, the archive holds 29 discrete discussions of potential group appointees. The earliest thread is for user 'Cobi' [W241] in April 2008.

WP:BAGG. Bot Approvals Group/Guide [W445]. Created as recently 00:50, 12 February 2017, this is starting to formalise the role and tasks of the BAG.

WP:BOTN. Wikipedia:Bots/Noticeboard [W588]. A "message board for coordinating and discussing bot-related issues". Its talk page [W418] archives 830 discrete threads across 11 pages.

[9] This occurred as a result of *unlinked* talk page discussion [W400] in December 2016. This is a typical example of significant formalisation not being linked to the considerations that when into its creation.

WP:BOTNEWS. Bots/News [W587]. Started 23:30, 8 August 2016, there were 4 newsletters in the study period. Past editions are archived to discrete sub-pages.

WP:BOTREQ. Wikipedia:Bot requests [W455]. Started 14:24, 19 August 2004. There were 4,276 discrete threads across 75 archive pages. The talk page [W409] archives 57 threads across 2 pages.

In fact, the Bot requests total thread tally (threads starting pre-August 2017) was 4,433, of which 157 (3.54%) were duplicate or broken content errors from (presumed manual) archiving. The errors were unevenly spread, with Archive #22 [W488] from July–October 2008 having 53 bad threads in 209 (25.36%). The nature of the errors meant that automated scanning of the data tended to miss, or misinterpret the errors. As this resource was the core dataset for RQ 2.2, I will discuss this further in that Section 5.3.

WP:FDB. Wikipedia:Bot requests/Frequently denied bots [W577]. Started 03:40 5 April 2007, it has under 100 page edits and has essentially been in the current form since April 2013. It gives a useful list of rejection reasons and is cross-referenced from the bot requests page.

WP:BRFA. Wikipedia:Bots/Requests for approval [W594]. Started 18:27, 13 March 2006, see section 5.1.3 on BRFA formalisation. The talk page [W419] archives 339 discrete threads across 15 pages.

WP:CAB. Help:Creating a bot [W45]. Started 21:06, 9 December 2006, it gives fairly comprehensive guidance on bot creation.

Wikipedia:History of Wikipedia bots [W655]: Started 03:27, 28 October 2002. This page is appears unfinished and sadly is not a particularly useful reference, due to the omissions of relevant bots, lack of explanatory text and missing cross-referencing.

Special:ListUsers/bot [W147]: (Special Report). Like all 'Special' namespace queries, it gives no total and will only return up to the first 5000 matches. It defaults to returning 50 items per page, but allows a maximum 500 items per page, e.g. [W148]. In theory, this alone *should* list all the active bots, but sadly it does not. It is also not possible to read the listing for anything but the current date via this page.

Global accounts list: Global bots [W146] This lists those user accounts that have a 'bot' flag granted globally rather than per-language-wiki as is the norm. The list is another 'Special' namespace query with the same limitations as described above.

Template:BAG Tools [W172]. Started 14:13, 16 October 2006 and added to over time this provides a set of icons to use inline to show the status of requests, tasks, etc. It has been useful both as an indicator of ongoing formalisation and to help to understand some of the more laconic posts of some users involved in the bot environment.

5.1.7 Precursor Work: Understanding The Bot User Environment

Having mapped out manner in which Wikipedia describes, authorises and oversees bots, the next task was to identify all bot accounts up to a sampling cut-off date of 30 June 2017. That date was chosen as being a month-end halfway through the year and occurred close to the time the sampling was being undertaken, so was essentially capturing all available data.

Yet again, the discontinuities in Wikipedia's records proved problematic. In the next section I will describe a series of discrete problems found during my attempt to extract an unambiguous full listing of bot accounts that have been used on the 'en' wiki. Some issues clearly belong only to the early use of bots, but others still remain. The fact that the presence of a 'bot' flag is held to denote a bot is undermined by the fact that bots may be approved to run without such a flag and yet there is no easily retrievable list of such anomalies.

Doubtless many would point to the BAG's stewardship of the process and the fact they monitor bot behaviour and function. But the issue here is not the work of row BAG but the opacity of information about simple high-level questions such as there being a list of *all* approved bots regardless of how they are technically identified within the system. There also seem to be weak areas where reliance on human action leads to loss of ongoing record, e.g. changing of flag status.

In this section, bold headings describe 19 different informations strands encountered in analysing the individual bot accounts. A few are useful and informative, such as giving examples of code used by a bot or giving fulsome and well-linked descriptions of a bot. In most cases though, the issues have detrimental effects for research and understanding the nature of a bot's work.

Naming Rules & Renaming Bots

A simple method of signalling that a user account is a bot is to include the word 'bot' in the account's username. The most common forms are all/part of the bot operator's name with 'bot' appended in some form. Thus, we see names like 'MartinBot', 'CbmBOT', 'Josvebot', 'WP 1.0 bot' and 'Amalthea (bot)'. Another approach is to indicate the purpose of the bot such as 'CitationCleanerBot', 'MediationBot', 'DYKHousekeepingBot' and 'Tom's Tagging Bot'. Note in the latter, the use of an apostrophe; this is unusual but not unknown.

A current request made of new bot operators [W579] is that "The account's name should identify the bot function (e.g. <*Task*>*Bot*), or the operator's main account (e.g. <*Username*>*Bot*).".

However, in an open system it can be hard to enforce community rules. An example is shown in a comment on the talk page [W204] of bot 'CanisRufus' whose operator 'RedWolf' [W337] states:

> "This account name was in place years before any such convention was proposed or established. Why should I be forced to change the name of this account to satisfy some silly (IMHO) new convention? I can see it for new accounts but accounts in

> existence before any new naming *restriction* [sic] was put in place should have the right to keep their original account name."
>
> (Comment by user 'RedWolf' [W337])

This example shows territoriality within a system where participants supposedly agree to not assert ownership. If renaming the account makes its purpose clear to the general user of the wiki, one might question why the bot's operator would assert the right to retain a non-standard name. Interestingly, despite no activity since 2012, according to an API query the 'Canis Rufus' bot still retained a bot flag at the 2017 sampling date. This would appear to run counter to stipulated guidance about moribund accounts. It is easy to see a bot owner might like to retain a flag in case of wanting to do bot work without the encumbrance of a new BRFA process, yet such behaviour undermines the BAG's role in oversight of bot work.

Non-standard names, of both bots and their operators, proved problematic to researching bot activity. As an example, a significant bot operator 'Betacommand' changed their username to a Greek Capital Delta 'Δ' (Unicode +0394) but some users referring that name mistakenly use the Unicode 'Increment' character, '∆', (Unicode +2206). Neither character can be easily typed directly from a normal keyboard, and the character will be an alien symbol to many unfamiliar with the Greek alphabet. Both characters are visually similar if used, but they fail to match each other in a search/regular expression context. There are also many single letter usernames employing most of the available characters in the normal Roman alphabet, e.g 'O' [W324] and 'R' [W330]. Practically, a single character name offers increased scope for error when inputting or analysing user names. As there are a finite number of such single-character names they are likely seen as a badge marking an early participant in the wiki.

My eventual listing of bot and bot-op names also included usernames in Arabic, Hebrew, Cyrillic, Devanagari and Japanese scripts, as well as less common European accents for Roman characters, and Nordic alphabetic characters. As Hebrew and Devangari scripts are written right-to-left this gives giving additional scope for text processing problems for the unwary when searching data.

Further to original names, consideration needs to be given to the renaming of usernames, i.e. account names. Renaming causes subsequent problems for administration or research because links in archives and logs may no longer resolve. This amounts to unpaid technical debt. Most efforts to repair the hypertext that I encountered appear to be made towards ensuring the renamed user retains a history of—or at least the attribution of—pre-existing edits by the old-named account. That choice reflects the close linking of a user's kudos and standing within the wiki community and their overall edit count. By contrast, names in past edits of page are often untouched. Ostensibly this is to preserve the original record and, to an extent, to avoid the extra work of retaining provenance. However, the result seems ill-considered given the functional damage caused to the hypertext. More problematic for a reader of Wikipedia is that archive pages, e.g. completed BRFAs, talk pages and such, are not corrected so as to link to the renamed account. As a result, nearly one in twelve approved bots cannot be traced with any ease from their primary source references within the wiki.

For example, on 1 April 2009, user 'Carnildo' [W237] created a joke BFRA for a 'Skynet' [W625] bot, which was 'speedily approved' (though no flag was actually granted). The BRFA originally linked to a User:Skynet page but the archived BRFA now shows a red link for the account indicating no such user. Following that link [W352], the resulting page indicates the username was changed under the Single User Login (SUL) process to 'Skynet~enwiki' on 22 April 2015. Yet following a link to the new username [W353] actually results in a redirect to a user 'SkynetAI' [W354] with no user information. The latter is a result of a further renaming on 13 November 2017 [W356]. However, the eventual resolved user account does have edits though the totals differ between the SUL listing (71 edits) [W355] and the account's contribution logs (69 edits) [W357]. Despite this being an example of a trivial editing account it nonetheless displays a series of poorly handled name changes such as can affect any user account that is renamed. The are problematic in terms of both the bot accounts and those of their operators.

Renaming also breaks various other forms of bot documentation. On 14 May 2018 'Luuvabot' [W301] was renamed to 'A-lú-mih-bot' [W215], with the original user page directing to the newer one. But this now breaks the user page's info box link to the bot's BRFA (the only such link on the page link). This occurs as the makers of the 'Template:Infobox_Wikipedia_bot' [W181] failed to consider the effects of renaming, as the auto-generated BRFA link assumes use of the new username, whereas the archived approved BRFA uses the bot username as at time of approval.

A bot operator may change their name several times, as in the case of 'HighInBC' who operated several bots. 'HighInBC' was renamed to 'H' (2006/02–2007/06), then 'Until(1==2)' (2007/06–2008/07), then 'Chillum' (2008/07–2015/09) before reverting to 'HighInBC'. Though all past edits by the user were merged into the current username's record, references to the user in archive pages break and unless the links are followed and any redirects noted, can lead to erroneous identification of users. For bot operators to cause this is egregious breakage of the hypertext as a whole in a wiki focused on its public face. It implies the edits and editor kudos outweigh consideration of the health of the overall hypertext. Allied with incomplete/incorrect logging these additional breakages make investigating the wiki's data both hard and potentially inaccurate as 'unkown unknowns' are hard to diagnose.

This ambiguity of reference data, and the lack of transparency of basic user data is surprising but proved to be a common occurrence when researching bot accounts. The was no indication this hypertextual breakage was due to ill-intent and it might best be explained as unseen and unresolved technical debt, allied with over-emphasis on user standing as evinced by edit count. It is only a matter of good fortune that most of the affected accounts are not significant bots, or it may be that only higher profile breakages get fixed. This does run counter to the 'many eyes' notion [296, p.30] of a robust commons where errors are rapidly seen and fixed.

Bot Chains

Bots may have tasks subsumed by other bots or be totally replaced by other bots, which further confuses the picture as to bot activity. A prime example is 'Shadowbot' [W349] which was

subsequently renamed [W158] 01:17, 11 July 2007 to 'AntiSpamBot' [W229] which was then replaced[10] [W230] by 'SquelchBot' [W363] which in turn became 'XLinkBot' [W383].

Retirement/Deactivation

A further impediment to easily tracking bot user accounts is the behaviour seen when a bot user is retired. Often, but inconsistently, the 'Retired' template [W187] is added to the top of the user page. The template notes states intended use: "Do not use this template unless you plan to completely and permanently stop editing". The template adds a large banner to the top of the page (or where the template is inserted) indicating the account is retired "RETIRED This user is no longer active on Wikipedia.". An example is 'HiW-Bot' [W268].

Correctly configured, template also includes specific code to show the user account is a retired bot, which changes the banner label from 'retired' to 'deactivated' and adds an appropriate different explanatory caption under the banner: "DEACTIVATED This bot is no longer active on Wikipedia.". An example of this correct usage is seen with 'MessageDeliveryBot' [W313]. The bot 'One bot' [W326] had a 'deactivated' banner set but this is never seen as the page was then made a re-direct to 'Legobot' [W295]. 'NVS(bot)' [W322] had a description of the bot's work but this was removed on retirement [W323].

In some cases when a bot is retired, no status banner is added and the bot's existing information is merely deleted. Typical examples are 'HerculeBot' (compare the page after retirement [W265] and before [W266]) and 'SpebiBot' (compare after [W360] with before [W361]).

A bot may also be retired in favour of a newer bot, which again can also cause confusion. For 'STBotD' [W365] there was information about the bot until the bot was deprecated [W366] in favour of another bot 'STBot' [W364]. This also links with the issue above of bot chains.

Replacement by A Non-Bot

A different form of retirement can occur where the bot is replaced by a different function, such as in the case of 'ENewsBot' [W253] which was replaced by a simple script rather than by another bot.

Right To Vanish

The Wikipedia notion of Right To Vanish (RTV) allows a leaving user to request, subject to the community's assent, removal of information [W635] relating to the user. This mainly involves the user page, talk page, any of the users own-space pages and information relating to the user. However, whilst allowing the user to vanish from Wikipedia, it essentially damages the integrity

[10] The nature of this replacement cannot be traced, though AntiSpambot was marked as 'shut down' 03:35, 2 January 2008 and not until 12:16, 19 January 2008 is there mention of Squelchbot. The latter's BRFA [W627] was first posted 04:09, 11 January 2008 and after the equivalent of *27 A4 pages* of discussion it finally was approved 16:59, 8 February 2008. The BRFA does state: "The code is a copy of the most recent code used by User:Shadow1 for User:AntiSpamBot, with some minor enhancements made…". This is a typical example of the difficulties caused by poor record-keeping within Wikipedia.

of the hypertext as many links now end up broken as the target page(s) are removed. An example of this is 'Redrocketbot' whose RTV makes it hard to ascertain the purpose of the bot and its behaviour, as the account's user page [W335] and talk page (including edit history) have been removed from public access. The contribution log remains and the stub user page has a link to it [W336]. The bot's BRFA also remains available [W620], which thus identifies the operator.

Whilst specialist direct querying og the wiki's database *might* find the edits and allow inference of the true role, in a well-ordered hypertext allowing such changes as RTV merit some reflection. For instance, it would be feasible for the RTV-related changes to the user page to record what the bot did (or was approved to do) whilst removing reference to the (human) operator and their Wikipedia activities. Such actions that give consideration of the hypertext as a whole could reduce (hypertextual?) technical debt as the hypertext matures.

Revoked BRFAs

Revocations of currently approved BRFAs are actually rare events and only 17 are listed [W13], relating to 6 discrete bot accounts. An example is 'Kumi-Taskbot' [W290], which initially made approved [W615] category edits as seen at 'Category talk:Artists from New Hampshire' [W12]. However, as a result of negative feedback about its work, all six of the bot's approval tasks were revoked. Despite this there was some [W443] 'discussion' [W405]by the BAG as to how to deal with ongoing tasks left orphaned by the revocation of the bot's tasks; these ended inconclusively. Laudable as it may be to stop bad actions by a bot, the community—working as it does by volunteer effort—has proved less organised in clearing up that which it breaks.

Another case is 'Anybot' [W232] which initially received BRFA approval [W599]. After initial bot edits there were complaints [W438]. Discussion[11] [W584] resulted in Anybot's BRFA being revoked. The discussion show clear tension between those who revel in process and those who want to effect progress—albeit without entirely without ignoring process).

Page Deletion or Block

Page deletion means that a bot's purpose can be hard to ascertain. Whilst complete page deletion (including the page history) might be assumed to be the result of disciplinary action, it is also allowed at the owner's request (as per Wikipedia protocol 'Criteria for speedy deletion: U1' [W636]). An example of this is '13bot' [W368]. The bot's user page content is now blank and cannot be read, nor can the page history [W369]. The bot's edits ('contributions') [W370] may still be accessed and fortunately edit comments refer to a now-orphaned sub-page [W371] with some information as to the bot's task including a link to an approved BRFA [W628]. This shows the ad hoc nature of some of Wikipedia's processes and the lack of attention to record keeping. Ironically, in the above, neither is the ex-users activity totally lost (as it is assumed they hoped) nor is it impossible to get at the users edits, albeit with difficulty.

[11] The discussion is originally pointed at Archive #4 [W583], rather than Archive 7, showing how easily inattentive human archival processes break the hypertext.

In some cases the user page is simply blocked. An example is 'Mobius Bot' [W316] where the user page only describes the user as 'permanently blocked'. The page history is still available and the previous user page edit [W317] (09:30, 19 May 2010) shows user information including an approved BRFA [W616]. The odd nature of the block here seems related to this bot being seemingly abandoned by its operator ('Mobius clock' [W319]) and subsequently being hijacked to undertake page content deletions (described [W208]; edit diff [W318]).

In the case of 'John Bot' [W274] the user page is blocked for being a ' (sockpuppet)'. As with 'Mobius Bot', the user page edit history remains accessible [W279] as are its user contributions [W280]. A previous edit [W275], before the content was removed gives some clue as to its use. It is one of a cluster of four bots ('John Bot II' [W276], 'John Bot III' [W277], 'John Bot IV' [W278]) all banned for the same offence. Of these, only 'John Bot' ever achieved a BRFA [W613], albeit at its third attempt.

A different sockpuppetry-related case is 'MelsaranAWB' [W310]. Here the bot's user page (and history) have been deleted due to its operator ('Melsaran' [W309]) being blocked as a sockpuppet. Despite this, the bot's user contributions [W311] have edit comments that link through to approved tasking [W667].

By comparison, 'KonstableBot' [W289] has a deleted user page but with no statement as to the reason. Its user contributions [W145] are not redacted and appear normal. Guessing the operator's name to be 'Konstable' [W288] it would seem that the bot was not sanctioned but rather it was de-flagged due to its inactivity and that of its (now-retired) operator [W207].

Once again there is confusing inconsistency in the way cessation of bot function is handled across a variety of contexts with direct or indirect causation.

Bot Account Never Created

Occasionally, although the bot's task was approved, the actual bot account was never created. An example of this 'ZscoutBot' [W387]. It had a BRFA approved [W630] but there is no evidence the bot account nominated in the BRFA was ever created. Whilst the convention appears to be that a bot account is set up before a BRFA is submitted, there is something to be said for not making an account in advance if the operator does not have faith in the proposal being accepted.

Infoboxes

An 'Infobox' [W48] is a small panel rendered in the right margin of the page and holding useful information about the bot (*if* correctly configured). The code for the box is inserted by use of the template 'Template:Infobox Wikipedia bot' [W181]. The template uses as an example a real bot, 'Polbot' [W328]. This enables the example to display an infobox using almost all of the template's features. It should be noted there is no requirement to use infoboxes, despite their utility for the general user.

In reality, only few bots make full or proper use of the infobox template. Of the bots that have gained BRFA approval, a mere 194 (8.9%) use infoboxes. One such example is 'HersfoldArbClerkBot' [W267] with an infobox that has most items complete including a link to its BRFA [W611]. The username both reflects the operator and the purpose of the bot (automated clerking for Wikipedia's Arbitration Committee). This is of considerable use to a general user trying to understand more about the bot, its purpose and its operator. It is regrettable that few bot operators avail themselves of this free affordance and that the BAG appears to do little to encourage uptake.

Compare that example with 'Josvebot' [W282]. This has an infobox but it lacks useful information. Or, 'MenoBot' [W312] whose infobox has no task data in it. It is worth noting that all these three bots still held a bot flag as at January 2019 (i.e. were *approved* active bots), despite which there are significant differences in the degree of information offered to the visitor as to the bot's purpose and regulation. This shows there is little pressure from the BAG, or the bot community in general, to formalise the information offered to the general community nor is there interest in achieving a conformity of presentation.

Little or No User Page Information

In far too many instances, the bot user page shows a lack of information about the bot, be it the task(s) undertaken, the operator, BRFAs or other detail. This is to the detriment of the rest of the Wikipedia community not least because, whilst run by a human operator, the bot itself cannot interact with other users. Certainly, the generally poor quality of bot user pages made this research into Wikipedia bots far harder than it might have been.

'ListeriaBot' [W298] is an example of bot with minimal on-page information, although the more experienced wiki user can follow links to (some) relevant pages. 'SunCreatorBot' [W367] is an example of bot that has no descriptive copy. 'VIAFbot' [W378] is a comparatively recent bot (from 2012) that does not even use the 'Bot' template [W173], such as provides a standard page head block. There is some information on the page about the bot's function, but only in diagrammatic form.

'JYBot' [W284] links to a BRFA in the header block but lacks general user page information. 'KarlsenBot' [W286] lists approved tasks which do link to the relevant BRFA, but this is unhelpful to the user. For instance, the BRFA for task number 4 states: "Function overview: template maintenance resulting from TFD". Only further searching in Wikipedia reveals that 'Templates for Discussion (TfD)' is actually a descriptor for a page relating to consideration of templates nominated for deletion or merging [W682].

Omissions relevant information range from those like 'KamikazeBot' [W285], offering no useful idea as to its purpose, to 'Xenobot Mk V' [W381] which describes everything *except* its purpose. 'LaraBot' [W293] has a page with no status icon—normally the most common element in a bot page. The status icon, as to whether the bot is approved, disapproved, or testing is a useful

indicator for the human user and an egregious omission. That said, the icon is not always correctly configured and thus indicates the wrong approval status., this despite the overview of the BAG.

Less common is for user page information to be supplied, but in a non-standard format, e.g. 'Lucasbfrbot' [W300]. Or, cut-off controls for the bot's operation may be supplied without information about the bot itself—i.e. *why* such a control is needed, e.g. 'Seedbot' [W348]. The bot's page may lack any information, e.g. 'Jayden54Bot' [W269] or may lack the usual header block, e.g. 'Le Pied-bot~enwiki' [W294]. There can be information on the page but no useful links, e.g. 'JerryBot' [W270].

Ideally, bots link to their BRFAs but occasionally the link is unhelpfully to an overall BRFA link, e.g. 'Lightbot' [W296]. Bot 'MessedRobot II' [W314] does link to BRFAs and give task descriptions but some are too minimal to be of use to the human reader.

User page information may be missing due to inconsiderate marking of retirement, as already discussed above. Compare the current [W304] page for 'MarshBot' with a pre-retirement version [W305]. A more considerate marking of retirement is catered for using Template:Retired [W187] as can be seen with 'MessageDeliveryBot' [W313].

Information about the bot is sometimes not kept on the user page. 'Werdnabot' [W380] chooses to use its talk page to describe its purpose. In some cases a bot may link instead to its operator's account, e.g. 'Jumbuck' [W283] whose user page is a soft[12] redirect to the operator's page on the Icelandic wiki.

Interwiki bots—a large sub-group of Wikipedia bots—such as 'Ptbotgourou' [W329] usually link to the interwiki article Interwiki links [W61] (which explains the interwiki link process). However, many interwiki bots simply include a link to the interwiki page but little other information, for example 'SassoBot' [W346], 'TARBOT' [W372] and 'Rubinbot' [W344].

No Operator Identified

A bot operator may create a bot account but not identify themselves as the operator. If the operator uses the bot account to make the bot's user page, there is no means to identify the user. Dake~enwiki [W243] is one such bot. The same dead-end occurs if the user is banned, as the user page becomes unavailable as with 'Lancbot' [W292]: its log page [W156] offers no clue as to the operator.

Poor Log Data

Another source of confusion is that standard Wikipedia logs, which ought to record the granting and removal of special account flags like 'bot', but which fail to do so. An example of inconsistent log data is the interwiki bot 'Lt-wiki-bot' [W299]. The bot has no record of initial grant of its 'bot' flag, causing this [W157] edit comment when the flag was removed:

[12] A 'soft' redirect is where the called page is loaded but renders a message pointing to the intended redirection page. Thus the user is aware of the redirection. A normal redirect involves the user ending up on the redirected page, i.e. a different URL to the one called.

> "Remove bot flag of inactive bot, feel free to re request. (Marked as Inactive Interwiki bot), Yay another bot with no log adding the flag :D"
>
> (Edit comment by user 'Addshore' 20:20, 24 July 2015 [W157])

In the case of the very early bots accounts, such as 'UgenBot' [W377] it is likely flags were granted before such matters were logged in the current manner. Despite this, there are still many cases of incomplete logs for reasons that cannot be easily explained. The comment above would imply these incomplete logs are a known issue for administrators.

Unlisted Bots

Another anomaly of poor early record-keeping is that bots such as 'Soulbot' [W359]—likely an interwiki bot—can be found that were never listed in any of the early 'Bot Policy', 'Bots/Registered bots', or 'Bots/Status pages'.

Cloned Bots

Early anti-vandalism bots were sometimes cloned, with the same code being used in multiple bot accounts run by a different operator. One such is 'Xenophon (bot)' [W382] which reports itself as a clone of 'Tawkerbot 2' [W373], whose user page is currently a re-direct to 'AntiVandalBot' [W231]. The latter also has clones running the same code, such as 'MartinBot' [W306]. It is also an example of a bot with no BRFA relying instead upon being a clone of an existing (approved) bot.

Re-Flagged Bots

It is possible for a bot to regain its flag. In the case of 'DarafshBot' [W244], the logs [W155] show it was originally flagged, after a successful 'BRFA'[13] [W601], as an interwiki bot. It was subsequently de-flagged as an inactive bot. The operator has since gained a fresh BRFA for a different task. In some cases the BAG may, at their discretion, return a flag without a new BRFA.

Name Protection

Occasionally, an operator will adopt additional accounts to avoid other editors using names which might be confused with the bot. Thus, the account 'DeltaBot' [W247] is a deliberate alternative account to avoid naming collision/confusion with 'Δbot' [W154] whose name would pronounced, when spoken, as "Delta Bot".

Multiple Operators

In rare cases a bot may genuinely have multiple operators as is the case with 'ClueBot NG' [W240] which is run by 'Cobi' [W241], 'Cripsy1989' [W242], and 'Rich Smith' [W340]. As in this case, the most common reason for such an arrangement is the complexity of the bot requiring round-the-clock operator supervision, as well as overall complexity. It is worth noting

[13] At third attempt, the first two being denied owing to lack of expertise.

that ClueBot NG is one of the most complex bots in use in Wikipedia and is at the heart of its anti-vandalism response. The bot has complex infrastructure and uses AI techniques, in which context multiple 'ownership' does not seem egregious and is a valid use of the WP:IAR rule [W656].

Name Clash

Wikipedia account names are case-sensitive (the first letter is *always* capitalised) offering scope for name clashes where names differ only in case—which might confuse an occasional visitor, especially if one or both users are bots. Thus 'MedCabBot' [W325] was renamed 'OldMedCabBot' in order to avoid confusion with 'MedcabBot' [W308].

Examples of Well Documented Bots

Despite all the difficulties above, some bots do provide the community with useful information. 'FastilyBot' [W254] is an example whereall tasks described and linked as well as reporting their status. 'JJMC89 bot' [W271] and 'Legobot' [W295] provide a good set of data for visitor. 'EarwigBot' [W252], 'MetsBot' [W315] and 'Ganeshbot' [W255] are further examples of offering the viewer a clear indication of tasks, links to BRFAs and general informative content.

Offering Example Bot Code

In a few cases, a bot's page will also give examples for users who wish to use the bot's output on their own pages. 'AlexNewArtBot' [W219], a bot for finding new articles, offers script examples for those who wish to use the bot's output on their own pages. 'WdefconBot' [W379] goes further, showing the code used by the bot to do its task.

5.2 RQ2.1 Work Undertaken by Bots

After the preceding groundwork, it is useful to re-state RQ2.1:

> *Can bot activity within Wikipedia be identified and characterised?*

To investigate this, analysis was undertaken to identify all user accounts that are, or have been, members of the 'bot' user group within the English language ('en') wiki within Wikipedia. This covered the period from their first occurrence within the 'en' wiki of Wikipedia, up until 30 June 2017, and it involved close analysis of over 2,000 discrete (possible) bot accounts. In accordance with the methodological approach adopted, data was drawn only from Wikipedia itself (including logs and the mediawiki API) and without any direct contact with Wikipedia editors. As has been discussed above, Wikipedia's back-of-house records are incomplete, even allowing for a system that has grown over time. As a result, the eventual list was of 'likely' accounts. this is because for a small number of cases the state of available data makes accurate judgment of bot status impossible for every account.

The earliest bot user account traced is user 'Rambot' [W333] which started operating at the end of 2002 (see Section 5.1.2 for more detail) and thereafter all bots *approved* before the sampling cut-off were recorded.

To avoid double counts in relation to BRFA status, inclusion in a BRFA category was counted in a hierarchy of priority: approved/denied/withdrawn/expired. Bots were counted by their highest priority of occurrence. Thus, a bot with an approved and withdrawn BRFA would be counted within the 'approved' group, one with denied and expired BRFAs as 'denied', and so forth. For bots pre-BRFA (Spring 2006) a BRFA-equivalent approval had to be assessed from available data drawn from across Wikipedia, such as bot flag grants on the Steward requests 'Permissions' [W678] page.

Tracking and recording bot activity proved anything but simple. Existing Wikipedia categories [W44] relating to bots were mostly abandoned, reflecting the fact that categories are essentially an user folksonomy with partial coverage, rather than a deliberate index. Despite the 'bot' flag (i.e. set of account privileges) being the primary marker of a bot, for many bots the event of granting of such a flag was never logged. This is especially true for early bots, when the control systems for bot administration was still being created. Thus, due to the absence of proper records, the grant of a flag to such bot accounts could only be inferred from edit activity. Creating an encyclopaedic list of bots therefore required close reading of the available data in an iterative analysis of multiple incomplete sources.

Being programmed by humans, bots are also subject to their creators' frailties. Wikipedia includes in its history of 'edit wars' [W642] an entry called '(Ro)bot wars' [W657] reminding us that sometimes computer automation merely magnifies the speed and scope of human error.

The result of this research was that 1,111 user accounts were adjudged to be approved bot use. 1,060 accounts had one or more BRFAs approved and the remaining 51 accounts might reasonably be assumed to be approved. In some cases, this was because the account predated the current formal BRFA process and approval had to be pieced together from a variety of incomplete sources. Until c.2006[14], bot approvals and granting of the bot flag where made on the Metawiki wiki but, even there, there are no (Web-accessible) logs retained from before 2004 removing early approvals from the public record. The 1,111 accounts, when de-duped for replacement/supersession by other bots doing the preceding tasks results in a resolved count of 1,086 bots.

Within a strictly BRFA context, i.e. for March 2006 up to 30 June 2017, 3,230 BRFAs were raised for 1,831 discrete (potential) bots. Bearing in mind that each BRFA is discrete and a bot may make more than one BRFA proposal, the following were the outcomes recorded, using the category triage described above:

- Approved: 2,141 BRFAs for 933 discrete bots

[14] There is no formal note of any change in policy which can only be inferred from (incomplete) Metawiki archive pages: the pages are complete but contemporaneous flag grants are not recorded.

    - …of which, Revoked: 13 BRFAs for 5 discrete bots (revoked by BAG or as arising from other administrative sanctions)
- Denied: 394 BRFAs for 346 discrete bots
- Withdrawn: 416 BRFAs for 302 discrete bots
- Expired: 279 BRFAs for 250 discrete bots

A simplistic analysis of the above is that 933 of 1,831 (50.96%) bots were successful, which would suggest the BAG is filtering out a fair amount of ill-considered or inappropriate proposals. Overall, including per-bot multiples, 2,141 of 3,230 (66.28%) of BRFA proposals were successful. The latter reflects that some successful multi-taking bots use discrete proposals for each task rather than a blanket approval (see Section 5.1.3).

From the start of 'en' Wikipedia until the end of June 2017 a total of 2,215 bot—and possible bot—accounts were identified. The earliest occurrence of a bot account is the inferred[15] bot flag for 'Rambot' [W333] in December 2002. The oldest still-approved bot still making active edits in June 2017 was 'RussBot' [W345]. First flagged in June 2005, it represents over 10 years of service to Wikipedia.

As noted, renaming of bot accounts caused issues of traceability. At least 44 (4.0%) of approved bots have been renamed at least once, the number rising to 79 (7.1%) if renaming under the SUL process is also included.

The bot accounts identified may be summarised into these broad groups:

- *Bots that have been approved* (1,111 accounts). For bots from 2006 onwards, this implies at least one BRFA has been approved or that the BAG has given assent for the bot to run unflagged. For the earliest bots it means a record has been found of the granting of approval, or the grant of a flag. In a few cases the status of the bot was noted from subsequent history and the start of approved activity was inferred from available data (often the first bot-like edit in the article namespace[16]). For these approved bots, it has been possible to find a resolvable BRFA (or reliable flag grant reference) URL for 1,060 (95.40%), though as stated for earlier accounts express approval was not recorded. In some cases, the same task has been undertaken by a series of bots or a bot has continued work but with a renamed user account. There were 25 such bots that were renamed or replaced by other bot accounts. This implies there have actually been 1,086 discrete bots.

[15] Recording of the granting of flags is inconsistent prior to 2005, and recording of flag grant in 'en' wiki's logs only starts in 2006.

[16] These dates cannot be entirely accurate due to the incomplete records of the time. Bots usually run for a short test/review period before being awarded a 'bot' flag. Any inferred flagging dates are, in most cases, likely within a month of the true flagging/approval date.

- *Bots with no approved BRFAs* (621 accounts). There are bots that have submitted BRFAs but which have failed to meet approval for one of three reasons. Some bots have multiple BRFAs coving a number of these categories. To avoid double counts within these categories, a bot's inclusion in a BRFA category was counted only once in a priority of a match within the order approved-denied-withdrawn-expired. Thus bots with BRFAs, but none of which have been approved, fall into 3 sub-categories.
    - *BRFA denied* (291 accounts). An unapproved bot account with at least one denied BRFA. Approval might be denied if the bot proposal were for an unsuitable task or the bot operator had have insufficient Wikipedia expertise.
    - *BRFA withdrawn* (154 accounts). An unapproved bot account with at least one withdrawn BRFA, but no denied BRFAs. A BRFA withdrawal is an act by the bot operator. This could happen, for instance, if the bot operator discovered issues during the trail phase that they were unable to resolve rendering the originally proposed bot task impossible or unsuitable. Or, they may have become busy/unavailable and were no longer in a position to operate the bot for the requested task.
    - *BRFA expired* (176 accounts). An unapproved bot account with at least one expired BRFA, but no denied or withdrawn BRFAs. BAG staff might adjudge a BRFA time-expired due to lack of engagement by the bot operator. For instance, a trial might have been approved by the BAG but the bot operator failed to run the trial or to report results.
- *Bots with insufficient information* (451 accounts). These are bot accounts where there is insufficient documentary evidence to assess their true status, other than that they fall in none of the above groups yet might actually be bots.
    - *Bad conduct* (97 accounts). These are bot accounts that were documented as blocked or banned (*before* any BRFA submission) after very limited numbers of edits due to a variety of failures to abide by Wikipedia's rules.
    - *Zero edits* (145 accounts). These are accounts created for bot use but never put into operation (this excludes bots with BRFAs, which are described above).
    - *Only editing outside main article namespace* (192 accounts). Bot accounts with a variety of purpose but not editing within the article namespace nor editing at high speed. This likely includes bots that would be approved un-flagged were they to seek formal approval.
    - *Un-authorised: no approval or BRFA submission* (44 accounts). These are bots making article edits but which have managed to do so without seeking approval or the censure of the bot-watching community.
    - *Minor unauthorised edits* (12 accounts). As per the previous group but accounts making under 10 main namespace edits.
    - *Blocked as unauthorised* (22 accounts). Accounts noted as being blocked as unauthorised bot activity and which never went on to seek approval.

- *No flag needed* (28 accounts). Accounts assessed as unlikely to have required flagging had they sought approval.

- *Not a bot* (25 accounts). Accounts mentioned in some contexts as bots but which do not appear to be so. Mainly this is due to the use of the use of the phrase 'bot' in the username. Such confusing naming is less common since the implementation of the BRFA/BAG process in March 2006. In part this is because now such names can be flagged up on creation and blocked, pending review of the account's purpose.

- *Bots Unapproved At Cut-off* (7 accounts). These were bots undergoing approval subsequently approved but not flagged within the sample period.

With the last group in mind, of the 1,111 user accounts adjudged to be bots with some form of approval, only 47 (4.3%) do not contain the string 'bot'. The presence of a '~enwiki' suffix in 49 (4.5%) bot usernames is a result of a Wikipedia-wide renaming of accounts to support a unified cross-wiki SUL. The suffix avoids naming collisions between same-named users in different language wikis.

5.2.1 Data Collected Data for Bot Accounts

As part of the process of identifying and classifying bots a number of pieces of information were collected (as listed fully in Appendix G) including:

- Names. Bot account name and URL-encoded name, and name of the bot's operator(s). For most bots the two are the same, but where non-alphanumeric characters or characters from non-Roman alphabets were involved the URL-encoded name proved important to have at hand to triage broken or wrongly-encoded links. Names of operators who renamed their accounts were normalised to the name as was current at the end of the sample period.

- URLs. The bot's user page, the operator's user page, the bot's SUL account summary page, the bot's logs summary, the bot data via the Wikipedia API, and the URL of the (first) approved BRFA for the bot (if any).

- Dates. First approval, last edit, bot flag grant and bot flag removal, and 'unapproval' date. Easiest to annotate were bots with clear bot flag information logging as these gave the effective lifespan of the bot. The last edit date was useful in figuring out bots that were moribund but still active. The first approval was essentially the first bot edit that was not a self-edit, i.e. to the operator's or bot's own user space. Where no BRFA or log of flag grant could be found, this date proved the most reasonable proxy for flag start date. The unapproval date was used to store the dates where the bot was de-authorised even though there may have been no log of approval and/or flag grant. For bots with BRFAs and logged flagging, the bot flag removal sufficed as the unapproval date.

- BRFA counts. Discrete counts for each of BRFAs approved, denied, withdrawn, expired, revoked.
- Edit counts. Edits (up to cut-off if still active) on the 'en' wiki and the bot's global edit count (via SUL reports).
- Status data. A large number of criteria were recorded such as the type of bot (approval state)) icon set, the state of approval, use of an infobox, etc.

Though some of the data may appear redundant in hindsight, it was still necessary interim data during analysis in the many instances where there was minimal trace of the expected logging of primary bot-related events, such a the grant or removal of their bot flag. The status-related data was useful at the collation stage when looking at the degree of occurrence of certain features across the overall set of bots.

5.2.2 Numbers of Concurrently Active Bots

As a result of the metadata collected when tracing and identifying bot accounts it was possible to calculate the lifespan of 1,017 (91.54%) of the 1,111 approved bot accounts (also see Appendix G, Table G.1). The missing 98 bots are accounted for by the manner of the bot's work. For example, those bots whose work makes no recorded edits such as 'Joe's Null Bot' [W273] or through missing accurate source data. This data is presented in Figure 5.1.

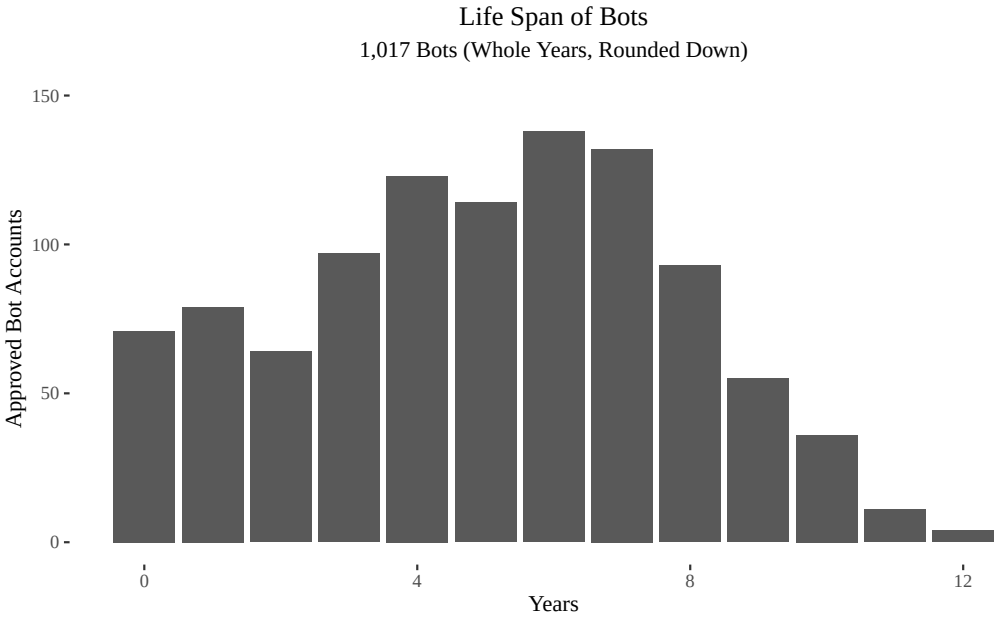

FIGURE 5.1: Life Span of Approved Bots

Given the degree of estimation of dates and the range of bot tasks, such a chart needs to be read with caution. The main observation is it shows no striking discontinuities and that the largest number of bots have a 'lifespan' of 4–7 years, though given the criteria for the weeding process

for moribund bots (see Section 5.1.6.1, 'WP:BOTPOL') arguably many of these lifespans are at least 2 years shorter than tabulated (i.e. a 2–5 year range).

In reality, a bot's lifespan is a matter of chance. Short-lived bots may simply be those created for a single or short-term task, the bot operator might have retired or been banned. Nor does the chart truly show the degree of activity of the bots, but simply that they were approved for a period of time. Though the periodicity and scale of the bots' work is missing here, such information is not necessary a helpful lens given that the data may as easily reflect a jobbing bot, or one that does a particular but infrequently recurring task, or an inactive bot that has been overlooked for removal of bot status.

It is simply not sensible to try and infer a 'value' for a bot based on its longevity, or edit rate. Volume of work, i.e number of edits, is not directly related to importance said edits deliver to the wiki. This shows the danger in strictly empirical approach. Without the significant amount of close reading of bot and operator user pages during the precursor work, number might otherwise give some false perspectives.

Using the above lifespan data it was then possible to ascertain the number of bots that were 'live', i.e. approved, any one time. As above, the same caveats apply as to the fact that some dates have had to be inferred due to incomplete Wikipedia record-keeping. Figure 5.2 below shows the number of concurrently approved bot accounts on a rolling monthly basis. Importantly, this does not imply those accounts were working actively throughout any given this period.

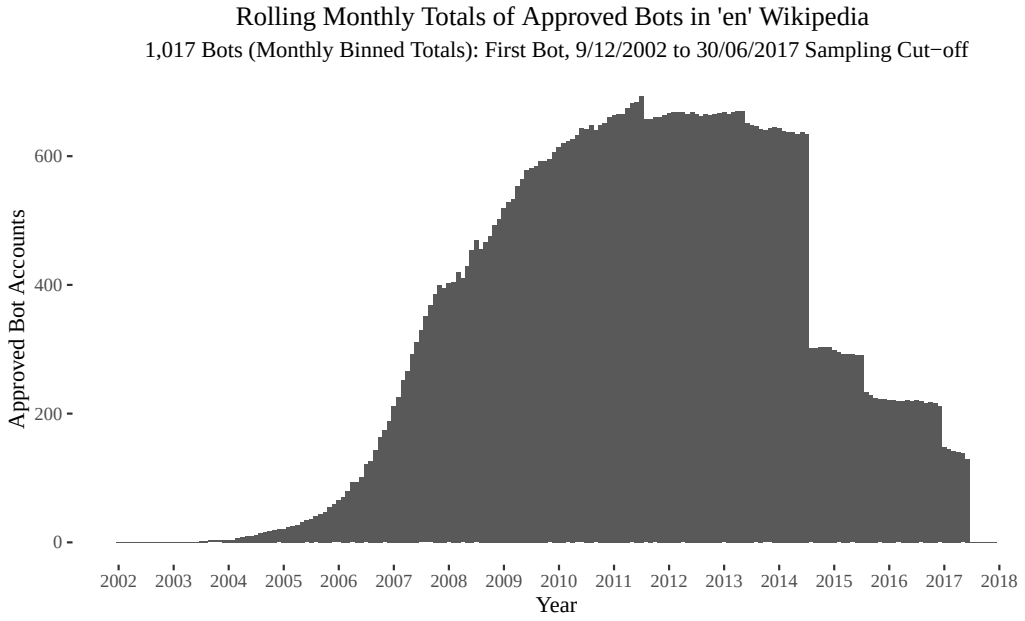

FIGURE 5.2: Rolling Monthly Totals of Approved Bots in 'en' Wikipedia

From 2002–2006 the number of bots remains tiny. But from 2006, the number rises rapidly until 2011 when the first fall back is seen. There is then a much shallower upward trend before a another small fall in 2013. This is followed by a slow decline in the total number of active bots before mass 'extinction event' in 2014 that halves the total number of bots. There is then a further gentle

decline punctuated by annual larger falls in 2015 and 2016. The graph may be oddly shaped but with the full back history of the bot population its divination is quite straightforward.

It is more easily explained working back along the timeline. The once-yearly dips are the result of the BAG doing an annual cull of moribund accounts. This review of bot account appears to have started in 2011 and has become more organised and effective since then. Operators of bots identified as inactive are informed that they will be de-flagged [W591]. This means the 'bot' permission is removed from the bot's user account, which effectively retires the bot and demotes it back to being an ordinary user account. Bot-like names for non-bot accounts are discouraged so a moribund bot account has no useful general editing role. De-flagged accounts are generally left in the system, if only so their past editing record is available to others. As with Figure 5.1, there may be a lag effect of two years or so as a quiescent period for that duration is a criterion for automatic de-flagging.

So what explains the much bigger fall in 2013? Around this time, the interwiki [W61] role was closed off and passed onto Wikidata making the pure interwiki bots surplus to requirement. One can clearly see that there were many interwiki bots, representing a big part of the bot population. The fact that not all interwiki bots were weeded on first pass shows how comparatively disorganised the bot management records were at the time (it has since improved and continues to do so).

However, on checking pertinent interwiki bot-related pages, none of them link to the fact that the interwiki role has migrated. It is described, vaguely, at Bot policy#Interwiki [W452][17], demonstrating the incomplete nature of documentation associated with bots. Such failure leads to exchanges such as an occasion [W589] where Wikidata has come on-stream and unaware bot owners are pitting their bots against the updated and differently functioning Wikidata as they (unintentionally) 'fight' each other. The latter is a good example of why most bot-related data needs very careful triage and cannot be trusted, casually, without due caution.

Whilst the longevity and number of active bots feels like a relevant fact to investigate, Figure 5.1 and Figure 5.2 show that without deeper checking of the data, simple unverified quantitative measurements of Wikipedia log data are likely to be uninformative at best and, at worst, misleading.

5.2.3 Assigning Types of Bot Contributions

After the lengthy process, described above, of trying to identity bot accounts and then to piece together their activities, I was able to code (qualitatively) their activities. As with the identification of the accounts, this was troublesome. In the same way that the naming and recording of bot accounts fails to be coherent on any analytical axis, bot tasks proved similarly incoherent. This lack of clarity posed a further problem in classification due to the way tasks are defined.

[17] In fact the information needed is actually hidden in a casual cross-reference to 'Interlanguage links' [W49].

Some operators define the task in terms of the observable outcome, i.e. that of a user reading an article page. Others define the technical textual, e.g. linking or listing work needed to effect such an outcome. Sadly the two approaches do not map easily, nor is there any evident attempt to do so by the wiki's editors.

Consider bot activities relating to Wikipedia's WikiProjects [W715]. These ad hoc projects seek to focus the activity of like-minded editors within particular subject areas. The term 'ad hoc' is not a canard: any editor may start a WikiProject. Like many Wikipedia activities they are user led and subject and scope can vary, with a result that as at Summer 2019 there are over 2,000 such projects. Similarly, numbers of project participants may vary from one or two editors, to thousands. One way in which WikiProjects indicate their interest is to 'tag' an article's talk page with a banner by adding a Template to top of the source code in a relevant article's Talk page. Thus the talk page [W340] for the 'Battle of Bosworth Field' article includes the template for 'WikiProject England' [W229], as well as two other project templates. As may be imagined this overlap can lead to turf wars over project scope. In addition, projects merge and split.

Depending on the bot operator (or a user requesting the bot's help), the tagging task above might be described as 'assessment and tagging work', 'adding templates' or 'applying categories' as the template often adds the page to Wikipedia's folksonomic categorisation [W383] system. Unfortunately, in other contexts, work on Templates or Categories can be treated as discrete task types.

For RQ2.1, a coding of 'WikiProjects' was used to bind these various terminologies together when relating to a WikiProject.

Another loosely defined area is 'clerking', where the bot task entails either a multi-part process or involves acting on a number of inter-related entities. An example of a bot in a clerking role is 'AFD Bot' [W217] which describes its purpose as:

- Maintains and updates User:AllyUnion/AFD List every 30 minutes at 15 and 45 minutes on the hour, a list of all articles currently up for deletion at Wikipedia:Articles for Deletion.
- Adds the section on the new UTC AFD subpage day one hour before.
- Adds the six days ago link.
- Closes the Wikipedia:Articles for Deletion discussion, 5 minutes after a new UTC day.
- Update Wikipedia:Articles for deletion/Log/Yesterday every new UTC day.
- Update Wikipedia:Articles for deletion.

It can be seen that a small clutch of tasks are being undertaken, all relating to a common purpose of supporting the Articles for Deletion (AFD) process. The clerking role is discussed further in Section 5.3.1.1 under 'The Clerking Process'.

Some bots are essentially jobbing bots, that will pick up any general tasks around. These may be sourced via the'Bot requests' [W455] page or via other sources within the wiki. Examples are 'BetacommandBot' [W235] (10 tasks) and 'Yobot' [W384] (26 tasks). This issue here is that the tasks are generally sequential, with some overlap, but lacking a hierarchy of importance.

Classifying the bot activities was an iterative task. In the first pass, some bots had a single clearly defined task, normally on the bot's user page or in their BRFAs. Others had multiple tasks (as above) and for these, the most common or frequent task was used for classification. In some cases, and the most time consuming to resolve, there was no clear indication of task. Here, it was necessary to work through the bot's edits and look at the sort of work being undertaken and to search for clues in bot-related articles and their task pages for mention of the bot.

The next iteration started to standardise terminology and to coalesce groupings that were close in purpose. As aspect of this was resolving process-defined descriptions with goal-defined descriptions where both had the same outcome (these differences arising from the way the tasks were requested and proposed). It was also used to set a primary task for those bots with more than one clear type of task. Again the terminology for the primary tasks was resolved to a standard set of descriptions.

The data was then revisited again to look at any bots noted as needing closer inspection to confirm their type or tasks and further reduce the number of discrete task types defined.

The end result is the data that is fully listed Appendix H, table H.1, including short descriptions of the task names. Extracted in Table 5.1 are the top 20 primary tasks undertaken by bots in the 'en' wiki, based on the number of discrete bots engaged in that as their primary task (thus ensuring multi-task bots were not double counted). The top 20 items represent 991 bots (89.199%) of the 1,111 approved bots and their edits represent 88.51% of all bot edits within the sample period.

TABLE 5.1: 'en' Wikipedia Bots, Top 20 Primary Tasks, ordered by numbers of participating bots.
Table shows: Number & % of Discrete Bots; Number & % Aggregate Edits; Average Edits per Bot per Task. (Sort Order: Discrete Bots)

Task	Bots	%	**Cum. Edits**	%	**Av. Edits/Bot**
Interwiki	310	27.903	13,711,169	16.961	44,230
Clerking	103	9.271	5,448,985	6.740	52,903
Templates	91	8.191	3,733,545	4.618	41,028
WikiProjects	79	7.111	4,418,379	5.466	55,929
Maintenance	77	6.931	2,796,232	3.459	36,315
Images	52	4.680	3,448,490	4.266	66,317
Categories	44	3.960	2,433,374	3.010	55,304
Vandalism	35	3.150	8,845,214	10.942	252,720
Link Repair	23	2.070	4,069,398	5.034	176,930
List Creation	23	2.070	1,011,859	1.252	43,994

(continued on next page)

TABLE 5.1: (continued from previous page)

Task	Bots	%	**Cum. Edits**	%	**Av. Edits/Bot**
Redirects	22	1.980	873,361	1.080	39,698
Newsletter	21	1.890	841,741	1.041	40,083
Disambiguation	20	1.800	221,485	0.274	11,074
Data Input	19	1.710	144,621	0.179	7,612
Code Cleanup	18	1.620	2,844,034	3.518	158,002
Archiving	17	1.530	1,524,954	1.886	89,703
General Tasks	14	1.260	14,181,512	17.543	1,012,965
Infoboxes	13	1.170	487,708	0.603	37,516
Stub Creation	10	0.900	519,357	0.642	51,936
Data Extraction	8	0.720	1,056	0.001	132
Totals	**991**	**89.199**	**71,555,418**	**88.514**	

In fact, very few of the tasks listed above are particularly well defined despite the triage work undertaken to standardise task terminology. As discussed above, this arises because of the differing manner in which bot operators use particular and discrete terminology to indicate things about their bots and the tasks they undertake.

Two of tasks with largest 'en' wiki edit counts show and interesting distribution of effort. The first of these is the 'General Task' category, where 14 bots are listed with this as their primary task. What that hides is that 84.71% of the edits are done by just 3 bots. Ordered by descending count (from 4.5M edits to 3M) these are: 'Yobot' [W384] (flagged June 2008), 'Helpful Pixie Bot' [W264] (including edits under previous name 'Smackbot' [W358]; flagged April 2006, renamed July 2011), and 'AnomieBOT' [W225] (flagged September 2008). Of note too, is that these 3 bots are also amongst top 6 by successful BRFA count with, respectively, successful BRFA counts of 26, 32 and 73. That said, these jobbing task approvals may be for quite wide scope of action for a given pattern and differing tasks have widely variable numbers of expecting resulting edits.

In the 'en' wiki's 'List of Wikipedians by number of edits' [W658] as at 19 July 2017, 50% of the top 10 editors are also bot operators, most notably 'Rich Farmbrough' [W338] operator of 'Helpful Pixie Bot' [W264] and 'Magioladitis' [W302] operator of 'Yobot' [W384]. Note that these are personal edits (which *may* be assisted by tools such as AutoWikiBrowser) and not those of their bots but, with edit tallies of c1.4M and c.900k respectively, these are not editors merely engaged in bot matters.

The second category of note is (anti-)'Vandalism'. Here the distribution of effort is also skewed to a few bots. Whilst 35 bots are noted doing this task, and in this category working exclusively on that task, the bot 'ClueBot NG' [W240] (4.5M edits, flagged December 2010) and its precursor 'ClueBot' [W238] (1.6M edits, flagged August 2007) account for 69.18% of the category's edits. In this particular character, the arc is of an initial spread of bots working against vandalism that is gradually winnowed down one primary 'ClueBot NG' bot and several helpers on ancillary

tasks such as maintaining user account blocks (e.g. 'HBC AIV helperbot5' [W263]). This trend reflects the increasing sophistication of the logic, including AI, needed to keep on top of ongoing vandalism within the 'en' wiki.

The 'Interwiki' category, though the largest by edit count, was passed off to Wikidata, in 2013[18]. With the exception of 'Addbot' [W216] whose 2.8M edits form some 20.70% of the category, there is a general smooth downward curve from the second-place bot' 929k edits. In should be noted the edit counts cited are in the 'en' wiki and that there will likely be counter-party edits in the other wikis in which the editor works.

These findings from RQ 2.1 will be reviewed further in Section 5.4.1 below.

5.3 RQ2.2 Bots as Supporters of Human Hypertext Editors

Having now explored and described the wiki's bot ecosystem and its activities, RQ2.2 now asks:

> *How do general users of Wikipedia access bot services?*

Having looked at bot contributions from the context of the bots, this question investigates how bots' work interacts with the broader community of editors, who are not bot operators and in most cases lack deep technical knowledge of bots and coding.

Wikipedia users are allowed to request help from bot operators. Analysis for this task was performed on the 'en' wiki's 'Bot requests' [W455] page up to 31 July 2017. That page is intended to let editors with a need for bot assistance or a bot operator looking for task to be able to interact.

The page was analysed by taking local copies[19] of all archive pages and then iterating through the discrete conversations to capture metadata about the topic, who was involved and whether there was a resolution or not. As with RQ2.1 the process involved multiple passes so as to harmonise the emergent metadata codes. Unlike the previous task, this involved a closer reading of the text as intent was often not stated clearly and the usual issues of poor link triage for events like account name changes meant some requests took significant effort to analyse.

The first recorded request at 'Bot requests' started at 14:42 19 August, 2004 since when a total of 4,276 discrete request threads were added up until 31 July 2017[20] (the end of the sampling

[18] As described on the talk page of the 'Interwiki links' article, in topic 'Wikidata and Interwiki links' [W62]

[19] This was done in case edits were made to the archive pages during the analysis, which took some months. Edits to archive pages are not *expected* but may occur nonetheless, and might alter the data as at the sampling data.

[20] This is a month later than the RQ2.1 cut-off. Data collection was carried for both RQs in mid-2017. By the time the RQ2.1 data gathering was complete, there seemed no harm in extending the RQ2.2 set slightly as the two are not inter-dependent. Care was taken to ensure threads in July 2017 did not reference any bots that were not authorised before 30 June 2017 (N.B.: there were no such occurrences)

period[21]). This request thread count was arrived at after de-duplicating content both within and between archives. Poor attention to archival work meant that 157 duplicate or partial duplicate items were detected. The most egregious example is Archive #22 from Autumn 2008, of the 209 threads archived, 53 (25.36%) turned out to be incomplete copies or duplicates. In a few cases the same thread was re-archived multiple times.

5.3.1 Issues Noted from Bot Requests Analysis

Before looking at the assessments of the types of tasks requested, in this section I will give some examples, all cited to source, of themes noted during the analysis of the outcomes tabulated in Tables 5.2 and 5.3. The sections have been themed insofar as possible, though many interrelate.

5.3.1.1 General Process

The first general area is of process, which illustrates the difficulty of separating the process from the task.

The 'Clerking' Process

An interesting challenge in coding bot requests is defining the edges between a singular task, e.g. correcting a single string of text, and a task that has several elements. Fixing a set of 'tags', i.e. altering template or category parameters, might involve any or all of:

- Identifying the pages needing action. A list might be supplied, or a query written to create such a list. Alternatively, the list might be administrative requests of some sort.
- Iterating the list. This might be a further sort (for different downstream actions) or might be the end action itself.
- Reporting may be needed.

There is no observable systematisation of this pattern with the result that numerous differing ad hoc solutions exists where a bot might do one, more, or all of the various stages. Where I refer to 'clerking'[22] this reflects that a bot is doing—or has been requested to do—more than one workflow task. Of course, the ad hoc nature of task boundaries mean specific tasks and clerking roles tend to blur at their mutual boundary.

An example of a clerking task is 'For someone with way too much time on his hands' [W473]. The requester seeks help with their Classification of Users with administrator status listing [W331]

[21] The sampling cut-off was applied to the *first* post in a thread. This is because some threads can stay open for a considerable time. Thus the thread start provides a clearer cut-off.

[22] This verb, of American English usage, occurs often enough in the subject matter as to make a reasonable descriptor for these task workflows, as distinct from individual tasks.

of Administrator-flagged users and their activity status asking "…run a query over it and remove all people that (1) aren't admins, or (2) haven't edited for three months or more?".

On being directed to a listing that *might* cover the task, they comment further: "…but I'd still appreciate (and give cookies to) someone who could quickly automate it, say with AWB[23] or something.". Here we see the task being redefined as a 'clerking' type of workflow that is eventually resourced.

Insight into Wikipedia Structure

The general structure of a wiki can be opaque. In the thread 'Historical "What links here" function [W489] as user is seeking to understand the generic 'What Links Here' [W161] offered in the left sidebar on most pages (certainly in the main, Wikipedia and User namespaces). It transpires there is no log of any of the following:

- When categories are added and removed from articles. For articles that are in a category, only the most recent time they were added is available.
- When templates are transcluded or un-transcluded in a page, no dates are available.
- When links are added or removed from one page to another, no dates are available.

The user 'Yuzhong' [W385] notes "The only way to get historical data would be to download the complete set of revisions of an article and then parse each revision.". In response user 'Dispenser' [W249] suggests they download—and implicitly instantiate—the *entire* 'en' wiki or use the 'Toolserver' [W683] or the mediawiki 'API' [W3]. All these are decidedly non-trivial actions for the general editor, especially those without significant technical background and/or non-consumer-level computer equipment.

Simple Tasks

By comparison to the above, not all tasks are complex. 'Quick text replacment' [sic] [W533] shows a task proposed, accepted and reported completed all within a few hours.

Extensive Discussion of Tasks

In some cases, bot requests lead to very long discussion of the proposal. In 'Moving reviews out of Infobox Albums' [W526], 8 different editors made 102 separate posts on this topic. Ironically, after all the careful consideration and test, user 'Tim1357' [W245] (operator of 'DASHBot' [W702]), the user offering a solution, was by then no longer in active participation. The outcome of the proposal is unknown.

[23] This is the AutoWikiBrowser tool.

Archival, Renaming And Page Moves

Over time, due to both loosened technical limitation and changes in style, wiki articles can be renamed many times over resulting in large numbers of multiple re-directs for inbound links. The proposal 'One Day International cricket' [W475] deals with a cricket-related page that has at least eight different (redirected) name versions of the page which it is proposed should be bypassed by amending inbound links. This led to discussion of Wikipedia policy 'Redirects: Do not "fix" links to redirects that are not broken' ('WP:NOTBROKEN') [W666] before the links were fixed eventually by bot 'SatyrBot' [W347].

Akin to the problem of article name changes, username change also affect links in the wiki. In request 'Username Change' [W514] user 'AnmaFinotera' [W358] requested all past signatures using their pre-change name to use their current user name. This generated discussion about the collision of rights to privacy with breaking wiki functionality: the wiki policies 'Courtesy vanishing' [W635] (WP:RTV, 'Right to Vanish') and 'Changing your username' [W702] WP:UNC ('username change') were cited. Participants suggested a redirect [sic] instead of actual name edit in the affected pages. The thread reached no consensus but it was reported that a bot ('SmackBot' [W358]) was addressing the request regardless of the views stated thus far.

An ongoing problem for outbound links from the wiki to external sites is that server and domain changes often result in 'dead' links that either return no page or an incorrect one. It is especially problematic when the target resource is a being used as a citation reference. This is illustrated in the thread 'Bulk-replace URL for Handbook of Texas Online' [W490]. The request was eventually resourced by approval of a current BRFA for DeadLinkBOT [W602].

Active Thread Discussions Being Moved Elsewhere

For no consistent reason, a discussion is sometimes moved out of the request page and elsewhere in Wikipedia, though not to everyone's satisfaction as seen in 'Tag some articles' [W481]. The requester states: "where I am confused is the purpose of this noticeboard if it is faster to just ask someone on their talk page."

It may be that the discussion moves to a BRFA page, as with 'Template change in Copy to Wikimedia Commons category' [W499]. In this case, the original requester notes the discussion has moved to a BRFA page ('SoxBot' [W626]). In a similar fashion, the thread 'Tagging request' [W508] is moved to a bot's talk page Referred to 'TedderBot' [W210] by the bot operator. These explicit references to a new location for the thread are atypical.

Links to Moved Threads Not Resolving

The combination of a thread being moved and a failure to re-link with a care to the hypertext can lead to a lot of work to validate the thread. An example is seen in Archive #14 'Request bot to convert %28 and %29 to parentheses' [W466]. This is archived by user 'Lightmouse' [W297] as single entry noting an entry noting a conversation being moved to 'Convert .28 and .29 to parentheses' [W403]. In turn, that URL opens 'Wikipedia_talk:AutoWikiBrowser' [sic],

at section 'Convert .28 and .29 to parentheses' [W401], i.e. it opens at a different page with its own discrete archives which, of course, has no record of the moved conversation. This is due to an easily overlooked redirect from 'Wikipedia_talk:AutoWikiBrowser/Feature requests'. After some fighting through the maze of redirects, the original content brings the reader to 'Convert .28 and .29 to parentheses' [W402].

Lost in that poorly-considered move-and-redirect is the actual original conversation of 9 posts, started by user the same user, 'Lightmouse'. A further 3 users took part. After the move there were a further 7 posts involving the proposer 'Lightmouse' and another user before In the moved location there were a further 7 edits, involving 'Lightmouse', and on other user before another user 'MaxSem' [W307] marked the task as implemented in AutoWikiBrowser on 14 November 2007. Not all such cases were tracked to this degree, and the example is included here to show both the information thus easily lost (or not found), and complexity of a task which does not lend itself to automation (and thus offsetting issues of scale). Not only does poor archival practice break the hypertext, but it does so in a manner that works against the human editors ability to use automation to help resolve the issue. A good example of the result of unpaid technical debt.

Templates And Categories

'Tagging' of articles, or more precisely their talk pages, generally refers to use of templates for categorization purpose, as discrete from explicit 'categorization' [W631]. Here I discriminate between those templates that offer content tasks (e.g. inserting country flags) or structural tasks (e.g. inserting 'Infoboxes' [W48]) in contrast to those that exist merely to 'tag' an article as being in the purview of some grouping like a particular WikiProject. Such templates are a common form of triage, but incautious use of automation may result in excessive numbers of tags, as referenced in 'Automated identification of excessive WP:Tag-bombing' [W567]. This is discussed in more detail in 'template bombing' [W545], addressing the problems of having too many templates.

Templates allow input parameters to help configure (and populate) the content they insert in tendered pages. Such nuance can have unexpected consequences, as seen in 'Bot to create links for unlinked pages called from template subpages' [W564], where parameter information is not percolating through the wiki as intended. This illustrates the potential fragility arising with chained templates. The thread also touches on transclusional 'noinclude' mark-up (as described in Section 4.3.2), although there is no indication of an understanding of the hypertextual nature of the task. Likewise, the thread 'Removing user namespace pages from article namespace categories' [W504] touches on editing link mark-up and subtleties of namespace markers with linking paths but again without regard to the underlying hypertextual aspect of the task.

In counterpoint to excessive tagging, evidence of very narrow classification can be seen as in 'Category:Blackpool F.C. seasons' [W491].

Templates live in their own namespace, and an atypical bot task carried out in that namespace is 'Template documentation' [W483], which was correcting the documentation mechanism offered for templates.

For the proposal 'Automated anniversaries for Portals' [W486], the task from the Wikipedia 'Anarchism task force'[24] [W431] gives insight into the flaws in Wikipedia's folksonomic approach to categorisation. Ironically, this can lead to rather anarchic structures: "The category system isn't really as hierarchical as it looks" which thus makes it less useful than intended.

The 'Hatnote renaming' [W534] request touches on the problem of renaming a template, leading to extended discussion of the merits of this as a bot task and whether re-directs should be used. The task is eventually reported as completed by bot operator 'Magioladitis' [W302] using their 'Yobot' [W384] bot.

References, and the form and styling thereof, can be another minefield for bot operators as shows in the discussion for 'Converting refs to list-defined refs' [W513].

5.3.1.2 Outcomes

The Outcome Is Not Clear

All too often, the result of a request is that there is no recorded outcome. A result may be implied but there is no record of, or link to, the fix. For instance, the thread 'Recategorization bot' [W502] mentions a (BRFA) request being fulfilled by user 'AllyUnion' [W220] but the BRFA could not be found[25]. In the thread, 'Userbox and Userboxes bot' [W559] the request is reported fixed, but the user doing so has no bot. Should a manual resolution be assumed, or might the user be reporting on an identified bot? Insufficient data is available.

Lack of any way to assess progress can lead to disagreement: in 'Bot for dashes' [W572] the task is discussed as fixed but a question as to the status of the request garners the slightly dismissive riposte: "I don't think we've ever bothered to systematically mark stuff as 'done' on this board". Whilst those handing requests are busy doing work on behalf of others, such a negative answer is not good for the community in a help-seeking context.

Analysis is impaired if there is no proper resolution record, as in thread 'Placebox template series' [W494]. The task is marked as fixed, via the 'Resolved' icon (by the original proposer, but there is no indication as to the bot undertaking the fix, or how/when the fix was effected. In another example, 'Changing a foreign word to its translation' [W565], clarification is requested and feedback given but there is then no annotation as to outcome.

There Is An Outcome, But Fixed by Whom?

In at least 10 request threads a fix was reported but the user doing the work could not be identified. An example is an early request where the closing remark is:.

[24] actually a WikiProject focused on anarchism

[25] The BRFA was not found in *any* of the five result categories described in Section 5.1.4

> "This problem has already been taken care of as according to User:Grutness. – Ally-Union (talk) 19:10, 25 Feb 2005 (UTC)"
>
> (Bot requests: Archive page 1, 'Italians in the wrong country' [W457])

User 'Grutness' [W259] is actually the requesting user. So another user 'AllyUnion' [W220] is reporting, second-had, that the work has been done. But which user(s) did the work cannot be identified. As an example of the problem of resolving this ambiguity, the subject matter is not mentioned in the talk page of 'Grutness' at the time of the request. Inspection of edit histories of some of the commune pages obliquely referred to in the thread showed no consistent instance of action. Some communes' pages were edited for the fix mentioned though in some cases this occurred *before* the date of the request. Such investigation involved human inspection of pages and could not easily be automated, as would help to resolve such ambiguities at scale.

A Fix Is Offered, But May Not Occur

Despite a bot operator offering a solution, it may fail to be delivered. In the case of 'A bot to update Wikicalendar articles' [W461], a BRFA was raised [W624] but subsequently withdrawn. At that point the process goes cold. There is no outcome reported or task triage so it is not possible to tell if the task was completed.

The discussion of the latter BRFA also shows a lack of understanding and consensus about how to achieve a task, e.g. when and if templates should be used. In reviewing the thread it is hard to gauge which critiques of process are fact-based and which are opinion-based.

A Solution Already Exists

A general editor my not know what solutions are available. Such an occurrence is the request 'Fair-use in non-mainspace' [W472] where a bot was already approved for the desired work. The oversight was quickly noted and the requester was polite in accepting the fact.

A Request Receives No Reply

For over one in ten (13.40%) requests in the sample, there was no reply to the request. The yearly rate generally varied between 10% and 20%. There was a peak in of 27.00% in 2005, when the requests process was still new. Reply rates improved in 2015 (5.81%) and 2016 (3.74%) before returning to 11.94% in 2017. Typical examples of unanswered requests are 'Unsourced BLPs'[26] [W509], and request 'Redirects for Olympic articles' [W532]. There is no indication that particular strands of request are unanswered, beyond those known to be irrelevant, but thee latter normally receive a post stating that fact.

In thread 'EditorBot' [W478] from 18 January 2008, the requester bumps their original request of 8 January 2008, saying it had been archived without discussion. Although the original thread

[26] 'BLP' is a Wikipedia shorthand for the sensitive topic of ':Biographies of living persons' [W440], also accessible via the Wikipedia shortcut 'WP:BLP'.

'EditorBot' [W471] had a reply there was no meaningful resolution. That one reply was a generally negative response from a bot operator ('NicDumZ'[27]) but there was no real discussion of the subject. The follow up garners a reply from a bot operator 'Betacommand'[28] [W234]: "Apparently no bot operator/coder is interested.". The subsequent exchange shows the weakness of this commons approach. The requester seeks help and the only person to reply essentially tells them those who can help are not interested and cannot be made to be so. There is no consensus and the requester shows frustration that no bot operator will undertake the task.

So, for the requester, the lack of any reply can be confusing. But, in an open commons like Wikipedia no one is responsible for fielding unanswered questions. Nonetheless, there is some recognition of the problem with the thread 'Meta: unresolved requests' [W557] which linked to a discussion about lack of take-up of requests by bots, at Wikipedia's 'Village Pump' page in thread 'Unresolved bot requests' [W704].

Rejection Or Refusal of A Request

Requests do get rejected, but some advice is normally offered as in 'New Bot Idea - Links' [W474]. Nonetheless this does not always mean that people listen to the advice. In thread 'YoshiBot' [W556] the requester is not happy to take "No" for an answer, as their desire for a bot exceeds their wiki expertise or ability to answer questions.

By comparison, in request 'Wikify_bot?' [W477] the request is knocked back but the requester accepts this in good humour despite not getting the result for which they had hoped.

A different reason for rejecting a proposal is that a bot operator might be unwilling to undertake a task due to fear of push-back against the work as seen with operator 'Kumioko' [W291] in the thread 'Removal of deprecated template' [W543]. In this case the work was eventually resolved via the semi-automated route of using AutoWikiBrowser.

The thread 'Vague idea for a bot/technical request' [W500] is unanswered but the user page of the requester 'GoneAwayNowAndRetired' [W257] is of note for its (2009) description of Wikipedia's dysfunctional nature. See the user's article 'Wikipedia is broken and failing' [W258]: the latter will be discussed in Section 6.2.

Rationale for Request

Only occasionally does the underlying reason for the request get made plain. In the request 'Bot For Wikiproject Aircraft' [W459] the requester explicitly mentions the tedium of the work if left as human-executed task.

[27] User 'NicDumZ' is the operator of 'DumZiBoT' [W250] [W320].
[28] The user has since renamed their account 'Δ' [W213]

Unclear Requests

An unclear request generally garners a request for clarification. For instance: 'Redirect and template bot' [W468]. In some cases it is caused by the same user as in this run of three unclear requests: 'Convert Rs. to ₹' [W569], 'Jain Agamas' [W571], and 'Infobox' [W570].

Multiple Requests in One Thread

A problem faced by those trying to help with requests is the behaviour of a small number of editors to propose multiple unconnected tasks in a single post. This can be seen in 'Bad word and Rss Feed Bot' [W458], where the requester asks for four discrete bot tasks. Listed below, they met varying outcomes.

- 'Greeting Bot': this was rejected as it is a known refused type of bot.
- 'Name/Date Wikify Bot': not taken forward as the bot logic was never clarified.
- 'Archiver Bot': replies note this task as duplicating the existing 'Werdnabot' [W380].
- 'Anti-vand Bot': a concept similar to 'AntiVandalBot' [W231], which is noted as being "dead" at that time.

This drive-by list of requests causes a confused reply, made worse as the initiator never came back to the thread again. Notable are the facts that the requester never responded to the follow-ups and clearly has not read up about known disallowed bot types. Thus the outcome was that two requests were rejected outright for reasons the requester could have looked up before posting, one was ignored as an existing bot offered the feature, and the fourth was unresolved as the requester never bothered to return and answer follow up questions.

In request 'Polynesia tagging' [W476] there are multiple requests in the one original post. Although the tasks were similar in nature they were still for discrete subject areas. All the requests for tagging were fulfilled but the results were then archived discretely making for unneeded effort when trying to trace the process through to completion.

Requests That Are Edge Case Tasks

In the thread to do with English usages describing death, 'euphemism bot' [W496], the eventual outcome was an alternative using a AutoWikiBrowser-based solution. Search also offers scope for ambiguity as in 'Need a Bot for a Job' [W506] where there is extended discussion—without consensus—of whether a bot is a suitable mechanism for delivering welcome messages to newly joined editors.

In the case of 'DashBot' [W484] the debate was over the pertinence of replacing hyphens in numerical ranges with en dashes. This is because notwithstanding Wikipedia's 'Manual of Style' [W659] guidance on how number ranges are described, it is potentially in conflict with Wikipedia's Bot policy article which warns against 'Cosmetic changes' [W450]. Although not reported,

this request led to a successful BRFA for 'DyceBot task #4' [W604]. Of note too, is the opening observation by the eventual operator that the task could be achieved with code already written for the previous proposal on the requests page.

Further examples are 'Remove unnecessary 'Image requested' templates' [W560] where task clarification is sought with an added suggestion this might better be done manually (or via AWB?), and 'WikiProject talk page tagging' [W561]. The latter proposal did get an outcome (via user 'Magioladitis' [W302]) but also records issues of a different sort of edge case—misapplication of tagging by a bot due to difficulties correctly detecting context.

Manual Fixes

Often the requester does not know, or care, about the degree of work to be done. Thus, in some cases a fix is achieved by manual edit rather than via bot, as in: 'Tchaikovsky navbox' [W463]. Here, editor 'DavidRF' [W246] fixes the issue and comments "I just did this by hand. Took me five minutes". This once again indicates shows that users are not always clear as to what tasks really need a bot.

Sometimes, the task is simply easier by hand, as with 'Book for Nobel laureates' [W515]. The user completing the task, 'JLaTondre' [W272], notes: "Done. Easier (for me at least) to do with Vim rather than a bot since the page formats were not consistent.".

5.3.1.3 Consensus

Issues of Consensus

Wikipedia sets much store by consensus as a basis for pre-approving any possibly contentious edits. Any editor may take part in such discussions. A lack of consensus was common occurrence amongst the bot requests analysed and in some cases precluded an outcome (or record thereof). In proposal 'Changing of {{otheruses4}} to {{about}}' [W511] different bot-operators dissent as to validity of a change. Handling of dates could also disturb consensus 'Convert dates to template' [W512], as could signatures for user edits in talk ages and the like 'Bot to change old signature/user name' [W479].

For proposal 'Removing templates from disambiguation pages' [W531], a user's 'simple' request in relation to letter and number combination styles quickly produces a lack of consensus about approved styles and compliance with the Wikipedia 'Manual of Style' [W659]. The result is no outcome to either the request or the discussion, which was likely continued elsewhere.

In the case of 'Missing settlements' [W480], consensus fails over the quality of the source material that might be used by a bot to make stab articles for undocumented settlements.

Even if there is some consensus on the Bot request page, the outcome may be derailed elsewhere. This is seen in 'Creating a set of new articles' [W547] where a BRFA for 'Theo'27s Little Bot' [W629] was submitted but then withdrawn in the face of vociferous opposition.

Consensus Needs Contributors

Another necessary aspect of consensus-based approval is generating sufficient engagement to adjudge consensus is reached. In the proposal 'International law wikiproject template' [W517] no resolution was reached and lack of consensus was cited as a reason for the proposal essentially being rejected. However, the discussion shows a lack of a common understanding of the actual scenario playing out. This illustrates consensus is not an empirical measure.

Proposal 'Template:NFL player → Template:Infobox gridiron football person [W527], which had no outcome, again illustrates the shallow nature of the consensus argument. Here the consensus is about something being 'non-controversial' [W716] in a WikiProject context. Similarly, proposal 'Change of template name' [W505] fails over WikiProject-related consensus issues.

Another example of the confusing manner of (non) consensus forming is seen in 'Changing of {{otheruses4}} to {{about}}' [W511]. A number of bot operators are involved as is one IP-based (i.e. anonymous) editor. A form of consensus voting can be seen. Again there is no report of an outcome so the result cannot be adjudged without significant extra analysis, that would involve figuring out where (across possible multiple locations) the fix would have been made and to look for edits reflecting such an outcome. I judged such additional work was out-of-scope for my analysis due to the significantly open-ended nature of such work, i.e. searching for evidence that does not exist.

The thread 'Links to Commons' [W530] shows not only a lack of consensus but an interesting disconnect between bot operators and the lay editors. The 'proposer' [W321] states "Therefore its probably in WPs [sic] best interests to create a bot with", casually laying down an expectation that other editors will volunteer their time on an unproven contention. The thread proceeds without consensus, but of note is a (collapsed) section 'Bot's thinking' where bot operator 'Avicennasis' [W233] includes pseudo-code to show possible bot logic to resolve the task.

Despite a lack of consensus, a task may proceed, as in 'Category:Infobox person using deprecated parameters' [W524]. A long confused thread appears to lack consensus but there are notes towards the end suggesting a fix is being effected, possibly manually. Manual fixing sidesteps necessary consensus about a bot-based solution but still does not mean editors are in accord about the overall work proposed.

With combative and persistent editors, lack of consensus can persist beyond the request page, as seen in 'Bot to tag articles only sourced to National Register Information System' [W553]. The task relates to a WikiProject, one of whose members is antipathetic to the task, although the thread decides there is consensus and a BRFA [W605] was submitted as a result. The same dissenting editor followed to the BRFA page and this resulted in the BRFA being withdrawn due to lack of consensus. As the bot operator notes in their withdrawal statement:

> "Editors bringing the project cannot get a consensus in the project that would be most affected by this task. Withdrawing this request until such time that the project

> can clean it's own house or a consensus of all the editors is established as there is a no-win situation that the bot's actions will be either reverted or called as against consensus as soon as it begins the tag run." (User 'Hasteur' [W261])

5.3.1.4 The Good, The Bad And The Trolls

On a more positive note, there are examples to be found of informative use of the 'Bot requests' page. In the case of 'Announcing a bot run [W498], a bot operator announces a bot ('MandelBot' [W303]) will be running *before* the event in order to allow for comment. Such advance notice of bot activity is unusual, although note strictly pertinent to this particular page.

Another good short example is 'Paraguay Templates' [W503] where the bot operator mentions the bot used to make the fix and reports when the task is complete. A bot's name is not necessarily something a general user might know. As importantly, for the requester or later reader, it is useful knowing than a task is complete and the identity of the editor undertaking the task. In a similar vein is proposal 'BSicons' [W574]. The discussion pulls in expertise from editors who are not regular of the 'Bot requests' page. It also links to the resulting BRFA and reports the task as complete.

It also helps if there is explicit reference to the bot being used,as in this thread: 'Tag Genetics stubs with poject template' [sic] [W482]. The bot op accepting the task notes this and states–and links to—the bot being used for complete the task.

There are also useful threads relating to use of AutoWikiBrowser as a solution, such as 'Schuylkill' [W552], where using AWB rather than a bot action and completion are both noted in the thread.

Though tangential to bot use, the thread for 'Holding cell' [W516] is interesting as it shows thought as to Wikipedia's transclusion and incremental formalisation to aid non-disruptive improvements. This is echoed in Template:FA-star' [W178].

There is the occasional discussion of code, as used by bots or, as here, on the Toolserver in 'Bypassing redirects with accent characters' [W520]. User 'Dispenser' [W249] seeks feedback on the pertinence of some of their coding for tools for the general community. In reality, this would be more pertinent to the 'Bot requests' talk page [W410] but the latter sees very few postings (see Section 5.3.3).

Despite ultimately not being acted on (though not explicitly rejected) 'Search for references/citations' [W539] shows an example of extended and informative debate.

Use of Status Icons in Request Threads

In reviewing thousands of proposals, an innovation that first appears in May 2007 is the first use of (template inserted) inline icons in proposal thread to indicate the progress with a request. In 'WP:EA' can be seen the first use of these icons, in this case it was a tick mark to indicate a successful outcome.

Subsequently, and noted via the 'Bot requests' talk page, in May 2008 thread 'Template:BOTREQ' [W414] reported the codification in a template with support for and automated addition of task status icons via template input parameters.

Incivility

Antipathy rears its head in various ways. Here, an uncivil exchange: 'School IP tagging bot' [W469]. Or, it can be a user's frustration: 'American film taskforce tagging' [W501], which ended up being referred onwards to 'US FILMS' [W212].

The incivility can be toward a bot instead of a human. In 'Project tagging bot' [W538], the thread has indications of antipathy towards bots and exasperation on the part of the requesting editor. User 'Hazard-Bot' had a BRFA withdrawn as the bot operator was too busy to respond promptly. User 'Kumioko' then had a bot approved on 12 February 2012 only for it to be revoked shortly thereafter on 18 February 2012, as recorded in 'Bot Approvals Group: Kumi-Taskbot' [W406] and which is covered in fuller detail in this 'later edit' [W407]. Also see: 'Kumi-Taskbot and collapse parameters' [W590].

Differing expertise can cause a clash. Here the requester does not accept the view of those with more (relevant) experience: 'Create a simple TemplateData for all Infoboxes' [W575].

In the thread 'Adding information templates to files that don't have one' [W566], the users take against the actions of Wikimedia Foundation staff who are accused of ignoring the editors who actually create the wiki's content.

Trolling

More unhelpful behaviour can include sarcasm, as in 'Bot Creation' [W528] or outright trolling: 'Tagging articles for a WikiProject' [W562].

Trivia

In some cases the subject matter is very trivial, but does get a response and outcome, such as in thread 'Dead Playboy Playmates' [W456]. Likewise, in 'SpongeBob SquarePants task' [W568] the requester seeks to resolve the spelling and capitalisation of the name 'Spongebob'. Surprisingly, the task is done (manually) and thanks are given by the requester. Although thanks are not registered very often, this is in part because most threads cease once the a task has been taken on by an editor.

5.3.1.5 Bot-related Issues

Requesting Editors' Understanding of Bot Tasking

Those requesting help can sometimes have little understanding of the task with which they want help. For instance, in 'Need a bot to set up assessment schemes' [W462], asked "Is there a bot that can adjust a template and set up the categories?". This thread showed the lack of collaborative spirit that I observed as an undercurrent in much of Wikipedia research I undertook. The bot operators—those who engineer the bots' code—take a perspective that the need to entertain lots of process variation is effectively a waste of their time. The unsuccessful thread is closed by this telling comment "There are hundreds of WikiProjects - why waste hours of people's time doing something that can be automated?" but the comment fails to elucidate their perspective of what is constituted by 'automation'. The impact of lack of effective human interpreters at the technical/non-technical interface is writ large.

However, the expertise of the bot operators is seen in 'Bot to cleanup Template:Duplicate file info' [W541] where in reply to the proposer, bot operator 'Skier Dude' [W351], having started the task then reports: "This list is full of copyright violations - blatant, no permissions - no FuR's, - as well as incorrect licenses. I've gone through a couple hundred & think that this is something that should be reviewed by people, not a bot.". This indicates the proposer either did not review the task before asking, or simply did not care. The bot operator rightly and understandably withdraws for the task given.

The thread 'Bot to identify recreated pages' [W464] asked to track re-creations of previously deleted article. The request was dismissed over the issue of potential false positives—which would then likely fall to the responder to fix or be censured by the wider editing community.

Even a bot operator may get confused as to the boundaries between roles. The requester, a bot operator, asks for solution that already exists such as in 'Protection check bot' [W467].

Thread 'Mark a lot of pages for microformatting, round 2' [W549] raises a question of human versus machine point-of-view and this becomes an issue. A BRFA is withdrawn with no consensus or outcome, due to continued dissent and bot operator retiring. The thread is an interesting glimpse of fractious but pertinent debate on discontinuity of machine versus human perspectives. See also: 'BRFA RileyBot 9' [W622], 'Requests for comment/Start date in NRHP article' [W670], 'Template:Start date' [W188].

In 'Category:Pages using citations with accessdate and no URL' [W554] a BRFA expires due to user push-back and there is no outcome. The bot operator appears to have run out of patience with editors nitpicking over process and practice. The BRFA [W607] in question warrants reading in full.

Sometimes the requester simply does not understand the nature of the task the are proposing, as with 'Add navbox to articles in Category:Infoshops' [W485]. Despite this, in this case there is a (manual) resolution and thanks are given.

The misunderstanding can be in the context of the bot's role: 'Bot to convert from citation templates to non-citation templates' [W525]. There is no outcome and the request is declined as not a task for a bot. The discussion of the page speed problem shows lack of clear understanding by the proposer of the problems envisaged by the bot operators participating in the thread.

Withdrawal of Bot Code

Problems arise for the hypertext as whole if bots use code which is then withdrawn, as is seen with the case of 'Replacement bots' which opens:

> "In a decision completely against the spirit of wikipedia and the open source community, *[username]*, sockpuppeteer and former bot operator, has pulled the code for his 'open source' bots and refuses to send them via email. As such, we need replacement bots up." ('Replacement bots'[W465])

The result is the failure of 5 active bots: 'Android Mouse Bot 2' [W222], 'Android Mouse Bot 3' [W223], 'Android Mouse Bot 4' [W224], 'Anchor Link Bot' [W221], and 'DinoSizedBot' [W248]. There is a positive response of other operators offering to cover certain roles. As is normally the case, there is no feedback as to the outcome. This request is interesting in that some of the relevant accounts are linked (often not the case): three of the original five bots are linked to via their BRFA, another has its user page linked in a user's reply. Of the BRFAs arising, one (subsequently expired) is explicitly linked and the other implied but the bot is not identified. To the later reader, the is no indication that any of the five affected bot functions were replaced.

There is a similar case in 'Sockpuppet investigations bot' [W507] with a request to re-use the code of a 'dead' bot. This shows the problem of getting code released for re-use by others.

Loss of Bot Operator

The retirement, disappearance, ban or sanction of a bot operator can be problematic for the users of the bot's work. Even if the operator is disciplined for offences using their personal account it can still affect the bot as it falls under their control.

The retirement of a bot's operator can result in a withdrawal of the bot's pending BRFA as in thread 'Non-notable diplomat stub articles' [W550]; There, thread discussion led to operator '~riley' [W341] raising a BRFA for their bot 'RileyBot' [W342] at 'RileyBot 10' [W623] only for the operator to report themselves as retiring a short while later and thus withdrawing their BRFA.

The reality is less clear, as this withdrawal is not noted on the bot request thread nor in the requester's or bot operator's talk pages. As is often the case, when trying to trace such an activity neither the requester or operator has talk page archives marked with date ranges making selecting the right archive a matter of trial and error. In addition, archival occurs based on the date of the last edit, so he resulting archive page is not always in data order based in the topic first

post. The operator has also changed username from 'Riley Huntley' to '~riley' on 4 May 2016 (causing broken redirects from edit signatures). The bot and operator share a talk page but the relevant talk archive page [W209] has no entries from November 2012 to January 2016, and the request activity was in March–April 2013. The requesting editor 'Kleinzach' [W287] has some references to the underlying matter in their talk page Archive page 34 [W205] and 'page 35' [W206].

To add further confusion 'RileyBot' [W342] lists task #10 as active under 'Active Tasks' but in the 'All Tasks' section below that the same task shows the BRFA approved for trial (true, it was) but inactive. Yet in the same section other BRFAs are marked as withdrawn (e.g. task #2) and the bot still has active tasks approved as recently as 2016 (when the operator changed their name). A log page [W343] for the bot's task #10 indicates a small amount of editing on 4 April 2013 (the BRFA was withdrawn on 29 April 2013). The subject of the task, tagging articles with the Template 'Notability' [W184], was clearly contested as Wikipedia's article 'Notability' [W662] criteria is a subject of great sensitivity (due to being somewhat subjective whatever the rules).

Sanction against a bot's operator may cause problems, as seen in 'Unreferenced biography of living persons bot to get projects involved in referencing.' [W510]. Here the work was being done by a previously sanctioned bot-op, who was then restricted in scope of action they were allowed. The thread 'indicates' [W211] that another bot operator had taken up the task but that offer seems to have failed due to the new operator being overloaded.

A bot operator may simply become too busy, as seen in 'Archive version bot' [W546]. The bot operator has other tasks ongoing, so action is delayed. In this thread: 'Need new UAA bot' [W523] there is a call for a new bot operator as the current operator has asking that someone else take over the task, due to their being overloaded with work.

There are requests to restore bots: 'Featured Portal closings' [W548]. This proposal resulted in a link to a suggestion on taking over old bots and the suggestion of user 'Theopolisme' [W376] for clearly named (i.e. identifiable) bots undertaking multiple discrete/sequential tasks: 'Bot takeover' [W585].

Sanctions against a bot operator can have significant effect, as can be seen in the scenario arsing in 'BRFA denied [W433]' (bot-op not task). Specifically:

> "**Rich Farmbrough's administrator status revoked**
>
> "1.2) Rich Farmbrough's administrator status is revoked. At any time after the closing of this case, Rich Farmbrough may request that his administrator status be restored by filing a request for adminship. Passed 9 to 0, 19:17, 15 May 2012 (UTC)
>
> "**Rich Farmbrough prohibited from using automation**
>
> "Passed 9 to 0, 19:17, 15 May 2012 (UTC)
>
> "Amended by motion at 08:45, 27 December 2015 (UTC)

"Vacated by motion [W435] at 18:36, 9 July 2016 (UTC) - latter is below"
(Rich Farmbrough prohibited from using automation [W434])

The above case resulted in a failed BRFA for another request, 'Asteroid Thingy' [W537], whereby 'Helpful Pixie Bot' [W610] task #50 was denied.

If a bot is banned, it can be a problem to get a task renewed. For example, 'People to pick up after RF bot's task' [W542]. This was referred elsewhere, leading to comments at 'Bots/Noticeboard/Archive 7: Rich Farmbrough's bots' [W578] in turn leading to 'successful BRFA' [W606]

'Redirects to sections' [W551] shows a case of a request, happily taken up and BRFA submitted for 'RileyBot' [W621] only for it to be withdrawn by the operator for the reason "Bot operator has retired".

The bot operator may not have retired/withdrawn from bot work but may simply be uncommunicative or busy. The latter is the case in 'Trouble communicating with bot operator, request for new bot maintainer' [W563], which had been using the services of 'Theo's Little Bot' [W375] but the operator was too busy to make new code choices. Despite the bot code not being proprietary, other operators seem unwilling to step in. That could reflect coders enjoying writing new code rather than re-working existing code. As they volunteer their time such a choice seems only fair, even if frustrating to the user who is now bereft of their pre-existing bot service.

In the case of 'WP 1.0 bot' [W576], the operator's requirement is explicitly known. Again there is no appetite for someone to take over the task, although two solutions are proposed: migrating code to the Toolserver and a method requiring no code, by simply improving category usage.

Bot Conflict

Lack of formal co-ordination can lead to bots potentially repeating work, as with 'Template:Infobox dam' [W555]. Here, this let to a BRFA being withdrawn: "Looks like annother [sic] bot came in and did the replacements. I'm withdrawing this request since the base request is fufilled." The identity of the other bot is not known.

Similarly, a lack of co-ordination leads to double tasking with an AutoWikiBrowser solution: '? a bot to convert {{Somerset}} to {{WikiProject Somerset}}' [W470]. The result is edit conflicts by 2 bot operators.

Bots for Tracking Bots

A proposal from 2007, 'Requests for bot approval request tracker' [W460] posits a bot to track approval (i.e. BRFA) status and most recent activity, and was well received. Though described as resolved, it is not clear as to the actual result. Since the time of the request, the BRFA and BAG processes have been much improved. Neither of requested features was in place during the sampling period. A similar request appears in 2011: 'Bot to keep track of bots' [W529].

Toolserver

Wikipedia's 'Toolserver' [W683] was offered from 2005 until 2014 when it was closed down. Its servers hosted software-based tools written by volunteers for the Wikipedia projects. Not well known to the wider user audience, it appears in a number of proposals. In 'List of doping cases in sport' [W487] the request is fulfilled but the user doing so noted the solution was effected using the 'CheckLinks' tool[29].

The Toolserver did suffer from reliability and resource issues. This situation is reflected in thread 'User:ArticleAlertbot has been inoperable for a month' [W519]. The original maintainer of the bot had retired and although a user had proposed a fix to the bot in question, this was stopped by inability to get access to the Toolserver. The loss of such a service is reflected by another request for the same bot: 'User:ArticleAlertbot' [W518]. The latter also shows requesters are not checking past proposals as this post-dates the former by only a day—and the latter proposer is directed to the former thread.

Discussion of the request 'Most Redlinked Bot' [W493] includes an extended mention of the sort of numbers and timing that occur in use of the Toolserver's 'Report tool'[29] by bots, even with the higher API rate limited allowed to bots.

Code running on the Toolserver was initially ported to 'Wikimedia Cloud Services' [W711] from June 2013 until the Toolserver was decommissioned on 1 June 2014 (see 'Migration from Toolserver to Wikimedia Cloud Services' [W712]). Within the Wikimedia Cloud Services the tool services were provided via '(Wikimedia) Labs' and 'Tool Labs' before both were 'rebranded' [W136] in July 2017 as 'Cloud VPS' [W726] and 'Toolforge' [W727] respectively. As at 10 January 2019, the Toolforge listed [W725] c.2,200 tools hosted on their systems though few tools offer information about their use. In many cases, these are essentially the 'private' code repositories of the bot operators for their bots. However, it shows yet again the nature of bots and other back-end tools is highly opaque and essentially a closed group on this otherwise open system. There is reference to a plan for 'Toolhub' [W126], which would give a richer interface to what is available on the Toolforge but the project has yet to come to fruition.

The Job Queue

Unseen by most users of Wikipedia there is a background 'Job queue' [W50] that performs long-running tasks asynchronously. Bots tasks being requested could have an effect on the performance of the job queue and thus the performance of the site as a whole. An area of relevance here is that changes to templates that are used by many pages. The effects on the job queue are also discussed in 'Replacing calender templates' [W495]. The request was effectively withdrawn, but the discussion included the observation by editor 'Zzyzx11' [W388]: "last time I heard, the job queue takes about 40 days to fully complete a round of all the pages on Wikipedia". The job queue is also referenced in 'Template:Canada CP 2006' [W497].

[29] This tool is no longer available since the closure of the Toolserver.

BRFA Problems

The BRFA approvals process for bots is important, but things can go awry. In the thread 'Undated Articles' [W522] there is confusion over why a bot was blocked, and a denied BRFA is mentioned: 'BRFA for KarlsenBot 6 [W614].

In this thread there are multiple rejections of BRFAs 'Redirects to lists, from the things they are lists of' [W573]. 3 BRFAs are involved. The first, for MoohanBOT, expired [W617] whilst the remaining two BRFAs, both for Redirectbot, were denied. MoohanBOT's operator (Jamesmcmahon0) appeared to be busy and not on Wikipedia.

> "**Denied**. Look, I was giving you the benefit of the doubt earlier, but since you have (1) posted the bot's password online publicly again after I told you not to, (2) disregarded my clear instruction in the email to Please change [your bot's password now that I have access to it], and (3) somehow don't see a problem with any of this, I have lost my confidence that you will operate this bot responsibly. No judgement regarding the task itself or you as an editor. Please reevaluate [sic] your account security practices."
> (Rirectbot (op: Philroc) denied: 'Bots/Requests for approval/Redirectbot' [W618])

> "**Denied**. On further investigation, apparently the issues related to the account's security still haven't been rectified. Based on that, alone, there's a basic security issue and editing-experience issue that's self-evident and incompatible with the safe functioning of an automated account on this project. Ignoring that and dealing solely with the concerns raised above, I can't confidently say that this bot has consensus to run within the bounds of the bot policy and therefore the expectations of the community."
> (Redirectbot #2 denied: 'Bots/Requests for approval/Redirectbot2' [W619])

An impediment to servicing bot requests is that unless an existing bot already has approval that covers the proposed task, progress has to await the separate BRFA approvals process. 'MTA (New York City) dead links' [W558], shows this can be slower than optimal.

A delay may also occur if involved parties do not keep a watch on the process moving forward, as here: 'add ref section to language stubs' [W536]. A BRFA was submitted for 'Helpful Pixie Bot' [W609] by its operator 'Rich Farmbrough' [W338]. The operator was then banned from using automation (i.e. a bot) at the 'Administrators' noticeboard' [W429] several days later. The BRFA was debated, approved in principle but formally denied due to the bot operators current ban. However, this reverse was not reported back at the bot request forum and the outcome of the original task is unknown. Other issues arising from sanctions against this operator have already been explored in the topic 'Loss of Bot Operator', above.

Automation does not always have to occur within the current, here 'en', wiki. Such a case is 'Looking for a Commons bot operator to do a simple category merger' [W544] where the bot 'AvicBot' [W393] was (apparently) approved for the task over at the'Wikimedia Commons' [W392], which is the Wikipedia organisation's repository for media freely usable within its wikis.

Employment of Bots

Not all users understand nor, indeed, like bots. Some would like more intervention into bot activities. In thread 'Quis custodiet ipsos custodes?' [W535], from 2012, an editor comments:

> "I'm concerned that after approval, there is no oversight of bot operations; it is assumed that an approved bot will only perform the tasks to which it is approved. Bot edits are flagged and generally ignored, so there's a reasonable scope for a bot account to go off reservation without anyone noticing, at least for a while. There have also been incidents lately of this occurring and causing a ruckus.
>
> "I'd like a dev to propose a minimally intrusive bot to monitor all bot accounts on Wikipedia. I imagine that over time it will grow more sophisticated, but for a first iteration I'd like to see at least monitoring of the namespaces edited and use of edit summaries. There ought to be some thought given to security/monitoring/auditing/-generation of the data that configures the monitor's operation.
>
> "A retrospective report of all bot accounts that have operated outside approved tasks would be of interest. It would not be able to be generated until appropriate configuration data had been assembled." (Comment by user 'Josh Parris' [W281])

Note how this is indirectly repeating a call for a bot to track other bots, as described above in the topic 'Bots for Tracking Bots'. Here, the editor makes plain a sense of doubt about bot edits overstepping the bounds of their approval. Regardless of whether that is justified, my general analysis shows a ad hoc bot environment that is opaque in function to all except those routinely and intimately involved in it.

Occasionally there is a glimpse into the background tech support of some bots, as in 'We need another User:WebCiteBOT' [W521]. Unusual, but seen, is a bot being created by one editor for operator by a another editor. 'JerryBot' [W492], was coded by user FoxyLoxy (User:Fl) for user Jerry (see approved BRFA for 'BRFA: JerryBot' [W612].

Re-engagement with bots (and their operators) is an interested aspect of bot requests. In 'wikiproject articles by size' [W540], the proposer describes past help by 3 different bots and is unsure how to re-engage their help: there are no replies.

At the time (April 2012), two of those bots were no longer available. 'BetacommandBot' [W235] was banned indefinitely in April 2009 and '718 Bot' [W214] appears to have last edited in December 2010[30]. The third bot, 'ClueBot II' [W239] was still active but it is unknown if it subsequently assisted with this task.

5.3.2 Analysis of Task Types And Outcomes

Table 5.2 below shows the top 25 categories of request as found at Wikipedia's 'Bot requests' [W455] during the sample period. These 3,769 threads represent 88.143% of the total requests. The entire listing of 76 categories is reproduced at Appendix J in Table J.1.

TABLE 5.2: Categorisation of 4,276 Requests for Bot Assistance 'en' Wikipedia: 'Bot Requests' Page (August 2005–July 2017). Largest 25 Categories only.

Category	Count	%
Tagging	565	13.213
Link management	389	9.097
Bot how-to	362	8.466
Clerking	258	6.034
Image management	233	5.449
List creation	223	5.215
Redirects	184	4.303
Infoboxes	181	4.233
Category changes	168	3.929
Templates	142	3.321
Bot repair	137	3.204
Subediting	125	2.923
Tag parameters	110	2.572
Notifications	108	2.526
References	84	1.964
Vandalism	60	1.403
Article move	58	1.356
Archiving	56	1.310
Assessment	55	1.286
Interwiki	53	1.239
Citations	52	1.216
Code cleanup	46	1.076
List order	41	0.959
Formatting article text	40	0.935
Data input	39	0.912

(continued on next page)

[30] '718 Bot' was not formally retired until June 2018, as part of a clean-up of moribund and abandoned bot accounts.

TABLE 5.2: (continued from previous page)

Category	Count	%
Total	**3,769**	**88.143**

In Table 5.3 below, the same 4,276 discrete requests at Wikipedia's Wikipedia's 'Bot requests' [W455] were evaluated as to the outcome of each request, i.e. what resulted as opposed to what was being done. Once again the incomplete and inconsistent nature of the records is reflected by the allocation of types of outcome. The size of the sample precluded trying to track and verify edits for poorly documented requests. At the same time, in keeping with the non-interventionist methodology what could not be found was essentially out of scope. Thus, the 'Unknown' category represents requests which matched no other type of outcome. In general that meant there was some discussion but no *reported* resolution, often due to a lack of consensus or an engaged party dropping out of the discussion.

In order to clarify the main types of outcome, the original 17 discrete outcome codes from the analysis—as listed in the three left-hand columns—are further aggregated into 8 discrete outcomes, as listing in the three right-hand columns.

TABLE 5.3: Categorisation of 4,276 Outcomes of User Requests for Bot Assistance Via Wikipedia's 'Bot Requests' Page (August 2005 to July 2017)

Outcome	Count	%	Aggregated Outcome	Count	%
Thanks	1	0.02	Misc	1	0.02
Withdrawn	77	1.80	Withdrawn	77	1.80
Moved	11	0.26	Moved	11	0.26
Not appropriate task	110	2.57			
Rejected	237	5.54	Rejected	347	8.12
Referred elsewhere	249	5.82			
Not needed	9	0.21			
Alternate solution	121	2.83			
Solution exists	138	3.23			
Already fixed	5	0.12	Resolved	522	12.21
Partial fix	42	0.98			
Requester fixed	52	1.22			
Fixed manually	74	1.73			
Fix in progress	367	8.51			

(continued on next page)

TABLE 5.3: (continued from previous page)

Outcome	Count	%	Aggregated Outcome	Count	%
Reported fixed	795	18.66	Fixed	1,330	31.10
Unanswered	573	13.40	Unanswered	573	13.40
Unknown	1,415	33.09	Unknown	1,415	33.09
Total Requests:	4,276			4,276	

Of immediate note is that over 1 in 10 (13.40%) requests get no answer to, or acknowledgement of, the original question. This is potentially off-putting to a new editor looking for assistance. A further 8.12% of requests were explicitly rejected. The latter occurs for a variety of reasons, such as asking in the wrong forum, or submitting a proposal known to be mentioned at 'Frequently denied bots' [W577].

Further aggregating the three minor outcomes 'Misc', 'Withdrawn' and 'Moved' into a single outcome 'Misc.', the data from Table 5.3 is plotted in Figure 5.3. The requests are aggregated by the year in which the each request was initiated. After the boom years of the late 200s, the request numbers begin to tail off. The percentage of fixed items per year has generally been in the 30–40%, thus at about a 1 in 3 recorded completion rate. The busiest years were during the period 2007–2009 during which the wiki was undergoing rapid growth and many new editors were joining.

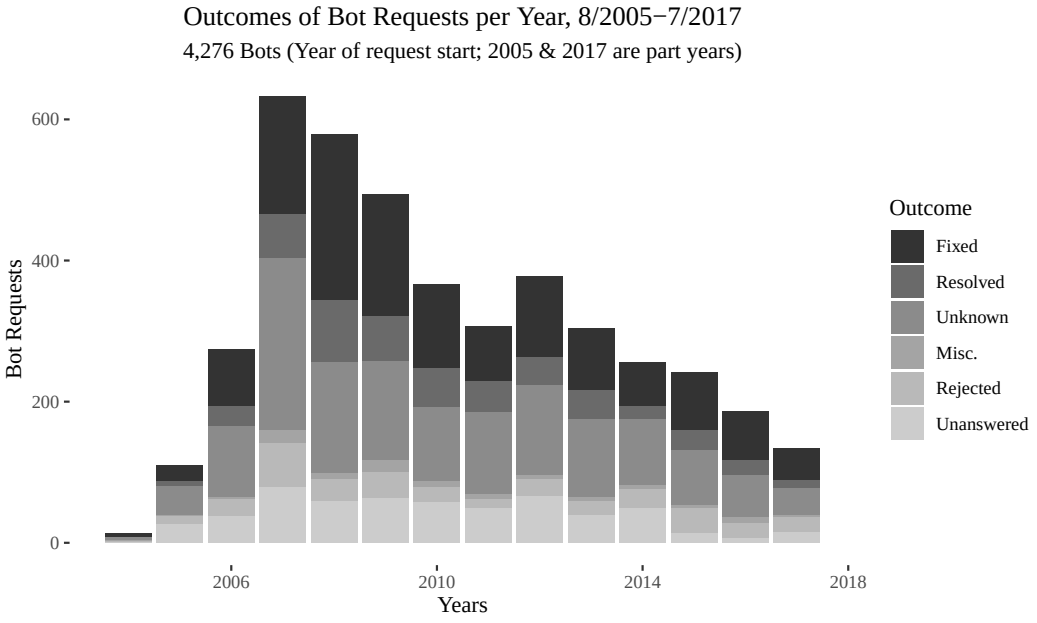

FIGURE 5.3: Outcomes of 4,276 User Requests for Bot Assistance Via 'en' Wiki's 'Bot Requests' Page (August 2005 to July 2017)

5.3.3 Interactions of Task Requesters And Bot Operators

As part of the analysis of the requests, all participating users were noted and, having resolved the results for issues such as name changes, it was assessed that 2,735 discrete (human) editors took part in the 4,276 request threads. At first sight, this averages to editor participation in c.1.5 threads but in reality, participation is heavily skewed to a small number of editors. The most prolific contributor (regardless of request outcome) is user 'Anomie' [W226]. This editor participated (one or more times) in 278 of the discrete request threads (6.50%) from 26 August 2008 up until 28 July 2017, this likely still actively engaged at the sample cut-off. User 'Anomie' is also the operator of three bots[31].

By comparison, 1,655 (60.88%) contributed only once and therefore engaged in—and likely started—only 38.70% of the overall requests. This would imply over 60% of requests were from repeat use of 'Bot requests'. Interestingly, 57 requests were raised by bot operator (27.94% of bot operators participating) also raised their own bot requests.

318 discrete users were identified as providing or describing an outcome to a 'Fixed' request as per Table 5.3. Of these, 195 could be identified as bot operators. Solutions attributed to individual bot user ranged from 62 discrete solutions by bot 'GoingBatty' [W256]' down to 72 bots with a single fix each. The top 30 bots, by fix count, account for 54.63% of the recorded fixes; the top 10 account for 33.66%.

The overall count of users involved in any thread was 2,735 of which 204 (7.46%) were identified as bot operators. 2,881 (67.38%) request threads featured 5 or fewer posts (including the first) and 3,801 threads (88.89%) involved 10 or fewer edits. As revealing is that 4,020 threads (94.01%) involved 5 or fewer discrete editors. Many eyes might be looking at these requests yet few choose to comment, which is pertinent in terms of consensus seeking. In fact, 3,311 threads (77.43%) involved 3 or fewer editors, meaning that less than a quarter of all requests receive comments (of any kind) from more than two other users.

By comparison, in the same sample period, the 'Bot requests' talk page [W410] accumulated only 57 discrete threads. The smaller number of threads doubtless reflects the fact that the main page takes the same conversational form as many talk pages. However, it reveals some comments worth drawing out. Of initial note was a proposal from 9 September 2007, 'Extending Archiving to 7 days?' [W413], which was agreed and actioned 3 days later. This means that requests if the *last-dated* edit to a request was more than a week old it would be removed to the current archive page (often by a bot). Another archival suggestion, from April 2007, was for themed archival:

> "This pages [sic] archiving is less than ideal in my opinion. I think requests with no responses should be left on the page longer; conversely I think requests that have been fulfilled should be archived faster. As it is, unfulfilled requests are archived in

[31] 'Anomie' operates: 'AnomieBOT' [W225], 'AnomieBOT II' [W227], and 'AnomieBOT III' [W228]

> a week, when it often takes at least two weeks for anyone to do anything about them, and fulfilled requests sit on the page for a week after they have been fulfilled clogging up the page. I recently set up archiving for the graphics lab in which fulfilled requests (requests with either done or resolved somewhere in the section) are archived in 3 days and unfulfilled requests are archived in three weeks. I think here we could go with 1 day and 2 weeks respectively and it would be a great improvement."
>
> ('archival suggestion' [W412] by user 'Dycedarg' [W251])

The idea was not adopted, but reappears again [W411] in December 2008, via a different editor 'Pigsonthewing' [W327]. That same user bumps the idea [W415] in April 2010 pointing to further discussion of the topic at the Wikipedia 'Village pump' [W703] page. Despite some agreement, this eminently sensible suggestion was not taken forward. Compared to the way BRFAs are filed and classified, the archival of Bot requests is somewhat haphazard. Data is retained but no review of the archive appears to be undertaken leading to duplicate (and possible missing) threads as well as broken formatting rendering the pages hard to read.

Confusion about archiving continues, as seen in two talk page threads from 2017. In 'Erm, what happens here?' [W416] a confused user asks "The instructions don't make clear to visitors what happens after a request is placed. Do all requests get denied or taken forward or do some linger here for months, or do requests that don't find favour with a bot programmer get automatically denied?". Citing an answer to that thread "They generally linger until someone takes them up.", in 'Archived without action' [W417] another user exhibits surprise that far from lingering, their request was archived, unanswered. Response to the thread does note the at time of posting the archival window had moved from 7 days to 60 (c.3 months). Thus, even after over 10 years of operation, the process was still confusing potential requesters of bot assistance.

In the same 'Village pump' [W703] thread from 2010 is this revealing comment:

> "Bot operators check that page very often, if its something they are interested in, or already have code to do a task they perform the request. But if its something no operator is interested in then it gets archived with everything else as stale. This works very well and has served its purpose very well for a long time. The main issue is the lack of good bot programmers who are willing to get their hands dirty with requests that no one else wants." (Comment by user ':Δ' [W213])

This reinforces the fact that demand, in the form of bot requests, outstrips the supply of *volunteer* bot operators willing or able to make a contribution. The general tenor of many requests would suggest the requesters see themselves drawing on a service rather than indirectly engaging with the efforts of fellow volunteer editors, the bot operators.

5.4 Discussion

The result of the analysis of bot use within a (wiki) hypertext is as confusing as was found in analysis of human editors' use of content transclusion (see Section 4.6.2). In this research the task was made more complex by the confusing and incomplete nature of information regarding bots and their management. In addition, the work looking at 'Bot requests' required significant use of archived talk pages which in turn revealed these too are in often sorry state with broken formatting, dead links, missing/duplicate content. I will discuss these aspects of my analysis in the a wider synthesis of both Research Questions 1 & 2 in the Chapter 6 that follows this.

5.4.1 Work Undertaken by Bots (RQ2.1)

My analysis of the Bots in the 'en' wiki reveals a functional but somewhat disorganised ecosystem. Whilst the emergent BRFA approvals process and controlling BAG team have worked reasonably well since 2006, *understanding* the ecosystem has been challenging as so many records are incomplete or missing. In the period 2002–2006, the number of bots was vary small albeit growing fast, and their role was yet to be fully understood and formalised. Then from 2006 until 2011 there was continued growth in the number of bots as seen in Figure 5.2

In tracking down the bot accounts, the 'Pareto principle' [W133] was very evident with noticeable proportion of (possible) bots accounts taking significant effort to assess owing to incomplete data, or as often, links broken by poorly considered subsequent changes to the record. The eventual record is described in detail in Section 5.2. 1,111 accounts were shown or, where records are incomplete, assessed to be bots, which resolve to a slightly lesser figure 1,086 when renaming of a counts is taken into consideration.

Classification of bot tasks proved equally troublesome as some bots declined to define their work whilst others undertook a number of task with no clear primary role Some of the busiest bots were essentially odd-job bots picking up bother user requests in 'Bot requests' [W455] but more often through their operators interactions in other loci of discussion within the wiki. Despite the emergence of the BFRA bot approvals process (see Section 5.1.3) and the BAG control group (see Section 5.1.4) to oversee bot activities, the whole is still somewhat disorganised. Having matured, the BRFA process is now much more clearly administered and recorded. More confusing are the varied exceptions that mean there is presently no accurate report of approved bots, especially active approved bots. I have shown that often-cited Special namespace report of 'users with the bot flag' [W147] cannot be regarded as reliable. It is to be hoped current BAG members know the true tally, but if they do, they chose not to share that fact.

An unexpectedly useful by-product of the survey of bot accounts was the running tally of active bots as shown at Figure 5.2. Stripping out erstwhile–though popular— 'Interwiki' [W61], it can be seen that since 2006 the 'en' wiki has likely supported a steady but changing total of c.100–200 bots. As many bots remain flagged for two years or more after last active use, that count is

likely lower but would additional new research to refine the data. Although the lifespans of bots were also assessed (see Figure 5.1), this proved less insightful that the tally of active accounts, not least as it draws more directly on the start and end of a bots life, which are now shown to not be accurately recorded across the data set.

In terms of participation, the 'Interwiki' [W61] linking task—cross-connecting articles on the same matter (though usually *not* simple translations) is the activity, that attracted the most bots. At least it did so up until 2013, when the interwiki task was transferred to 'Wikidata'[32] [W390]. Thereafter many bots, especially those only doing interwiki work were de-flagged, causing the drop in active accounts that years as seen in Figure 5.2. Given the large number wikis for of discrete languages[33], even if interwiki bots only served one pairing of 'en' and another wiki it would easily account for most of the bots, though in reality many interwiki bots operated across several languages. In its time, although doing work *between* hypertexts, it did represent a deliberate work to maintain coherence between discrete wikis within the overall Wikipedia domain.

The two other significant categories of work were '(anti-)Vandalism' and 'General Tasks', the latter reflecting the odd-job bots who will take, usually sequentially, a range of tasks. The anti-vandalism task has become a high profile task undertaken by a diminishing number of increasingly more complex bots reflecting the complexity of the task. The role is now undertaken primarily a single bot 'ClueBot NG' [W240] with some minor administrative assistance (maintaining account blocks) from other bots. Though poorly configured bots may apparent vandalism, this is tiny compared to human malfeasance in this regard. Thus the task, though important, is one I consider secondarily related to the RQ as were human editors to better behaved such fixing of the hypertext would not be required.

Harder to gauge were the 'General Tasks'. As noted in Section 5.2.3, although 14 bots were adjudged to operate in the role the vast majority of the edits were done by only 3 of the 14 bots. Furthermore, the operators of two of those bots were in the top ten editors by edit volume for their own work. This points again to the fact that any one metric does not really tell the full story.

Viewed from the nature of the task undertaken, it is the case that some of the work being done can be held, in its effect, to be assistive of the hypertext's overall structure and stability. Fixing broken links by amending the target at link source is definitely useful however, fixing links via re-directs can be seen as kicking the task forward into the future and essentially adding technical debt whilst giving a present-day fix. Given that 59% of 'en' wiki articles, i.e. `<page>` are re-directs, and doubtless increasing, this does not seem a sensible trajectory if one cares about the health of the hypertext.

Thus, as with RQ1, the picture is not unambiguously clear. It cannot be stated that there is no evidence of activity that may be considered beneficial to the hypertext. However, it is less clear

[32] As explained in the Interwiki article talk page in section 'Wikidata and Interwiki links' [W242].
[33] Over 280 as per 'List of Wikipedias', section 'Detailed list' [W105]

that such activity as was detected is requested or planned with more in mind than maintaining the state of the viewed article pages. There are no distinct activities identifiable that start from a hypertext perspective opposed to one assessing the wiki by its viewed content. That is not to say those two perspectives are in opposition, but the lack of the former may help explain some of the accrued technical debt observed during the analysis.

5.4.2 Bots As Supporters of Human Hypertext Editors (RQ2.2)

Analysis of Wikipedia's 'Bot requests' [W455] resolved the participants in all 4,267 discrete threads but divining the nature of requests and resolving the outcomes proved as challenging as for RQ2.1 given the comparative inattention to data logging in non-article pages compared to the main article space.

As detailed in sections 5.3.2 & 5.3.3 above, a picture emerges of a highly skewed set of results. Fewer than a quarter of requests result in a discussion with more than two other editors. Surprising as this is, these less busy rates show slightly higher rates of positive outcomes ('Fixed', 'Resolved') compared to the percentages in the Aggregated Outcomes averaged for the whole data set as seen in Table 5.3 below. Clearly in some cases, the request is resolved quickly and without wider discussion.

The categories of task undertaken for the requests, as shown in Table 5.2 showed the same mixing of functional and outcome-related descriptions as was encountered in RQ2.1 (as discussed above). A synthesis of the tasks from the RQs 2.1 and 2.2 is given in Appendix K and it shows a degree of alignment. Indeed, had both tasks been undertaken at the same time—not possible with one primary investigator—it might have been possible to get better alignment of task categories. This comparison has some limitations. The bot activities list the *primary* task and a bot may be approved to undertake several discrete tasks. Conversely, a bot request might describe the outcome in terms of its affect on an article's rendered text as opposed to *how* that change is effected.

Concentrating more closely on RQ2.2, the evidence is that whilst bots are being used to undertake task that help the health and ongoing use of the (wiki) hypertext, it is far less clear that such actions are happening due to the deliberate intent of those making the requests. Put another way, the tasks being requested and resolved may be having hypertextually beneficial effects but it is not unambiguously clear that this reflects a deliberate rather than an accidental result of user's request.

Regardless of the tenor of interactions, it is clear that the servicing of bot requests is entirely at the gift of the bot operators and further restricted via the BRFA/BAG process. Thus there might be a task where there is consensus amongst editors but which might not find favour from, or similar consensus between, BAG members who are the back-stop to bot regulation. To understand this process would take deeper examination of the request threads, relevant discussions in other parts of the wiki and the resulting edits arising from resolving a request.

My perspective from this analysis is that if general editors have a viewpoint when requesting bot help it is that they are essentially engaging with a service, whereas in reality they are engaging with the (goodwill) of other human editors who happen to be the bot's operator. In that sense, a bot is somewhat like the eighteenth century automaton 'The Turk'[34] [306, pp.154–163], not in the sense that a human is doing the work but in the unseen human guiding hand on a seemingly automated task.

This blurred interface between the roles and actions of human and bot automations is something I will explore further in Chapter 6.

5.5 Summary

An unexpected revelation from the RQ2 is the degree to which unresolved technical debt is quietly rotting the hypertextual coherence of the wiki, especially in non-article pages. These may be less visible to the 'consumer' of the wiki's articles but it nonetheless erodes the hypertext from within. This problem is described in detail in the issue of 'Links to Moved Threads Not Resolving'.

In this chapter I have examined the role of bots within the English-language wiki of Wikipedia at some length. In order to analyse bots in any meaningful way, I first had to map out what is meant by a 'bot' within Wikipedia, and then how such an actor is defined, authorised and monitored. In doing so, the difficulties encountered also illustrated some of the ways in which available data and systems are problematic for the wider audience's understanding of bot activities. This work is contained in Section 5.1.

Building on that reconnaissance I was able to trace—as far as incomplete documentation allows—all bot accounts from their first use in late 2002 up to a sampling cut-off in mid-2017. This revealed over 2,000 possible accounts. Of those, 1,086 accounts (54.30%) proved to be accounts approved at least once as a bot. It should be noted that in some cases approval had to be inferred from limited available data (see Section 5.2.1); this was especially true for records prior to 2007. That listing enabled me to calculate, or in some cases estimate, the likely lifespans of bots and the number of bots concurrently approved over the sample period (see Section 5.2.2). Approval does not relate directly to activity. Some bots are active constantly, whilst others only run occasionally by dint of their given task. This makes the latter more difficult to tell apart from bots that have ceased to be used. Further work with this dataset allowed investigation of RQ2.1. This involved tabulation of the types of tasks being undertaken by individual bots as shown in Section 5.2.3, Appendix H, and Appendix J.

Looking at human tasking of bots (RQ2), the last part of the research in this chapter looked at the 'en' wiki's 'Bots Requests' page where any editor may seek the assistance of a bot (operator). After reviewing data quality issues encountered whilst analysing the requests (see Section 5.3.1)

[34] also see 'The Turk' [W201]

as with the work for RQ2.1, I tabulated the types of tasks requested and, separately, the outcomes of tasks. These are shown at Section 5.3.2, Appendix I, and Appendix J). The interactions of bots and human editors users were reported in Section 5.3.3.

As intimated at outset, proving the *absence* of something is not simple as its presence. It is clearly the case that despite the presence of large numbers of bots and, most recently, a defined control structure there is little clarity as to their use. Though in keeping with Wikipedia's open and false structure that allows contributors a high degree of freedom, the loose degree of control and lack of standardisation means there cannot be said to be organised, widespread, of hypertextual techniques overall. Work undertaken by bots, such as dealing with re-directs as a result of page renaming/moving, definitely does help maintain the overall health of the hypertext. At the same time this activity is not seen to be driven by a vision of the coherence of the whole hypertext (and by which I do not mean aspects like page styling or rendered layout). On the evidence collected, this would suggest the structurally positive outcomes are essentially arising indirectly rather than intentionally. That is a positive result for Wikipedia's current functioning yet does suggest a regime of good-enough fixes, which in turn may not be the collective desire but pragmatic acceptance of the (volunteer) resources at hand.

It is the case that one significant type of bot, thought doing 'interwiki' alignment work were rendered surplus by one of the few deliberate structural changes. That was migrating into Wikidata the management of the cross-linking of similar pages in different-language wikis. It should be noted the latter are not just translations but language-native articles on the same topic in different language wikis.

Whilst considering Wikidata, RQ1 work found no evidence of transclusional use of Wikidata, so that resource has much yet to offer. RQ2's work showed no direct evidence, nor inference, of bots being used for content transclusion.

The case is less clear with the bots requests (RQ2.2). At first glance, many tasks that have been run by bots can be legitimately run in support of hypertext. The richer picture comes from the multiple read-through of the original request threads, which points to the same conclusion as above, that positive contributions are as much incidental as by direct design.. The metadata collected for Appendices G and I helped identity regular contributors and bot-account holders. That is useful to understanding in this area as actual bot account names are not mentioned regularly in the request threads.

Despite the lack of an absolute answer to RQ2, I contend that the research has been successful in establishing a deeper picture of bot accounts than has hitherto been documented. That allowed me to examine tasks undertaken both as submitted by bot operators and a wider set of requests from the wider group of ordinary Wikipedia editors. Thus while bots appears to do no harm, it cannot be said that they are being tasked efficiently in relation to maintaining the quality of the hypertext.

In the next chapter, I will consider the findings of my investigations, into content transclusion and the employment bots, within the context of a social machine.

Chapter 6

Social Machines, Knowledge—and Hypertext?

'*Never send a human to do a machine's job.*'

Agent Smith in *The Matrix* (film, 1999)

'*Power corrupts and obsolete power corrupts obsoletely*'

Ted Nelson's 'TEDtalk', TED2 Conference 1990 [341] (1990, phrase coined c.1987)

In this Chapter, I will review the my research findings from the previous two chapters within the context of a social machine and to explore how and if this broad socio-technical concept is making deliberate use of the hypertextual, in the context of my research within Wikipedia (q.v. Section 2.11 for description of the broad concepts of the Social Machine and the Semantic Web). Why so? By observing a subject in too small a frame it is possible to miss the greater picture, yet it is in that wider context that our knowledge generally resides—especially in today's networked world.

A central notion of Berners-Lee's original description of the social machine is that "Computers can help humans if we work with them to create abstract social machines on or via the Web: processes in which the people do the creative work and the machine does the administration" [35, p.172]. This view of the participants' contribution is not necessarily the current view (see Section 2.11) of the field of Social Machines where the role of the machine is now in the ascendant, in the form of Artificial Intelligence (AI), or more realistically rather Machine Learning (ML) mechanisms. Faith, or hope, in the superiority of machine processes tends to position the human in a more subordinate role within the combine.

Surprisingly, in 217 pages of Shadbolt *et al.*'s recent (2019) book *The Theory and Practice of Social Machines* [311] 'hypertext' there are only two mentions of hypertext and even those occur only tangentially in the opening theory chapter. More surprisingly the Semantic Web is not referred to at all[1]. Semantic drift has occurred in the conception of a social machine and it would seem the roles are reversed from Berners-Lee's vision, with technology ascendant and AI to the fore. This begs a question: at what point does point does the model of a platform like 'Zooniverse'[2] [398] tip from being citizen science into being a mere people-farm, with an army of 'mechanical turks'[3] feeding the machines' algorithmic training needs?

This same absence of any reference to the a hypertextual medium was notable in most of the papers that I examined in relation to social machines and the Web. Few papers even mentioned the word 'hypertext' (or hypermedia), or alluded to it. Whilst I had not expected to find copious mention of the term, the limited and incidental nature of its occurrence does elicit surprise. This may reflect the fact that—for many authors and researchers—whilst the Web whilst hypertextual in form it may be seen as just the technical substrate upon which the social machines exist and, as such, unworthy of direct acknowledgement. I believe this to be an oversight, and a misunderstanding of the role of hypertext. Our society is currently transitioning from a print era, where our knowledge lives predominantly on paper, to a world of digitally-based knowledge storage. Our cultural perception of how and where knowledge is curated may lag the realities of its storage, but to fail to understand the affordance of hypermedia is concerning, not least as doing so forces us to confront what does not easily or simply map from old media to new.

6.1 Wikipedia And The Social Machine

Turning now to my research into Wikipedia reported in Chapters 4 (transclusion) and Chapter 5 (bots) I consider several aspects of this work before discussing topic as a whole. I will comment mainly on the 'en' wiki. Whereas in my transclusion study, the mark-up used to denote transclusion was accessible to machine based identification and tabulation, RQ2's work on bots and requests for their use required much closer reading with first-language facility with the studied text. In the first instance I will look at how humans or bots do, or do not share, and the implications for collaborative knowledge work.

6.1.1 Ownership, Territoriality And Re-use

Wikipedia is very clear in its essay 'WP:OWN' [W664] on ownership of content that (original emphasis): "**No one**, no matter how skilled, or how high-standing in the community, has the right to act as though they are the *owner* of a particular page.". Yet that aspiration is not necessarily

[1] More precisely, the term does occur but only in the titles [sic] of cited references.
[2] This is a 'people powered research' platform, that grew out of the Galaxy Zoo project.
[3] See 'The Turk' [W201] and the 'Amazon Mechanical Turk' [W1].

reflected in the activity of editors. Differing viewpoints, especially on divisive social views, can lead to combative editing but that can not account for all the editing behaviour seen during the research.

As reported for RQ1 in Section 4.6.2.6, some editors are uneasy about 're-using' content from elsewhere in the wiki, i.e. transcluding content. Their concerns appear to be that the output of other editors will degrade their own efforts. The latter may indicate a degree of laziness or technical ignorance as much as a principled objection to transclusion as editors may edit any (non-protected) article, such as those from which they might transclude.

Those editors actively engaged with the wiki appear set store by their edit record. Although 'Edit counts' [W641] in Wikipedia are not formally described as a ranking system, when reading through talk pages, and pages in the wikis' 'Wikipedia' namespace it is hard to escape a sense that edit counts are nonetheless a proxy form of editor ranking. Wikipedia also makes clear that 'subject matter experts' [W645], whilst desirable contributors, have no special standing amongst editors as a whole. Though this is a well-intentioned policy to avoid such editors assuming authority through claimed expertise, it also weakens the use of fact in the face of edits based alternative opinions that lack substance. Disagreements become proxy fights via the assertion of missing or poor referencing, or other procedural infringements. In an era of 'fake news' this is not necessarily a helpful policy within a social machine.

It is hard not to become sceptical about Wikipedia's assertions in relation to non-ownership of content (q.v. WP:OWN, above) in the light of observed behaviours, such as I have described in Section 4.6.2.6. These appear to undermine non-ownership principles, even if the cause of such non-observance of Wikipedia rules is not fully clear.

Kittur & Kraut (2008), recognised the disparity between the open-access environment of Wikipedia and other peer production settings:

> "*Management implications*. Both the public perception and the ideology of Wikipedia are of a free and open environment, where everyone is equally eligible to contribute. However, even a free and open environment needs coordination, and peer-to-peer communication is ineffective if too many people are involved. Our results show that it is highly beneficial to have people who can set direction and provide a structure to which others can contribute. In Wikipedia this leadership group emerges naturally and informally based on editors' pre-existing interest in the topic, expertise, availability, and commitment. However, in other peer production environments it may be valuable to appoint leaders to positions of authority and control and to recognize their roles formally. This may be especially true for tasks where it is critical to get things right."
>
> ('Harnessing the Wisdom of Crowds in Wikipedia' [199, p.45])

Their latter point is an important one. The gentle anarchy of Wikipedia's very flat authority structure is not necessarily well fitted to social machine contexts where the outcome may affect the

health and well-being of those participating. Some defined structure, or at least the allowance for it, is also a necessary restraint against in-groups excluding those new to, or at the periphery of, such systems. A problem in a Wikipedia-type structure is that the "people who can set direction and provide a structure" are essentially a self-selecting group and who can easily become protective of their acquired privilege in the system.

Although not directly present in the areas of the wiki I investigated, another challenge faced by Wikipedia arises from the fact that it allows articles about very recent, or even current events. Despite its 'WP:NOTNEWS' [W707] policy, articles touching on areas of disputed social politics can become toxic as seen in articles such as 'Gamergate controversy' [W32] and 'Climate change' [W15] and their associated talk pages. Contested national views on history, or disputed borders can have a similar disrupting effect on otherwise normal editing behaviour. These situations are less likely to encountered within organisational hypertexts but the possibility of such discordance needs to be recognised and planned for.

A complication arising when such adverse behaviour occurs is that the small number of administrators[4] in any wiki are often assumed as being the leaders, in the manner Kittur & Kraut describe, even though that is not their role. The confusion likely arises because the very flat hierarchy of roles in mediawiki-based wikis means the general editor has no other formal authority figure to whom to turn. Pertinent is the Wikipedia's 'arbitration' [W432] processes for dispute resolution are complex and forbidding at first encounter. It is thus unsurprising if editors seek to try to resolve issues via administrators, despite that not being the latter's explicit responsibility (Expectations of adminship [W428]).

These possible issues of 'ownership' of articles caused me to look into the psychology of ownership[5], but it has proved hard to pin down in the general literature. This is echoed in the conclusions of Wang *et al.* whose 'The Role of Psychological Ownership and Ownership Markers in Collaborative Working Environment' notes "The possession of objects has been one of mankind's most puzzling behaviors. Various previous works suggest that the psychology of ownership is well-rooted in humans." [367, p.231].

Exploration of the concept of ownership within boundaries of individual disciplines, such as psychology or sociology, proved to be uninformative for this broader societal, collaborative context. A contributing factor observed was over-focus on method and process, leading to narrow scope and rather than viewing systems both as a whole *and* as collection of individual actors, as opposed to one or other. The mismatch in the outcomes I was seeing contrasted with the Wikipedia's guidance began to suggest the issue was not a notion of *ownership* but rather one of *territoriality*.

Thom-Santelli, and co-authors, appear to be some of the few researchers addressing the issue of

[4] the full structure of Wikipedia's user roles are described in Appendix D.

[5] In a general human behavioural context rather than in a narrow scientific sense.

territoriality in collaborative systems. In 'Social Tagging Roles: Publishers, Evangelists, Leaders' [346], an exploration by Thom-Santelli, Muller & Millen of the narrow role of 'tagging'[6], an interesting observation is the egocentricity of the editors' described motivations [346, pp.1042–1043]. This echoes Wang (above) in concerns about ownership.

Thom-Santelli, Cosley & Gay develop this further in [345, 342, 343] where they note, with reference to the social roles previously mentioned in [346]:

> "Our data revealed that territoriality does emerge within online space—and that when a hierarchical style of collaboration is crucial to success, territorial behavior may be valuable. However, it can also have a negative effect by deterring new member participation. Instead of negotiating with a Maintainer who reverts their contributions, new editors of an article might just give up on the article, or the community as a whole. To maintain the health of a collaborative social system, encouraging a diverse pool of participants to help maintain documents may help slow the decay of artifact quality." ([342, p.1484])

From my research, this has a familiarity with the behaviours I observed in the Wikipedia data as reported in Chapters 4 and 5. Territoriality appears to work at a deeper, more visceral level than higher level social behaviours akin to assertion of ownership, leading to misinterpretation of the cause. Whilst the two behaviours are essentially the same in terms of generating potential conflict between editors, the underlying drivers may actually be different. I believe gaining an understanding of that difference and applying that insight is important in terms of trying to improve socio-technical interactions within a social machine. Not all actors are open-handed or communitarian in their dealings with others and our social machines should account for this, though it is not clear that they do so (by intent).

In Wikipedia's case, the reaction to disagreements has been incremental growth of the now-complex set of arbitration processes, as mentioned above. Few, except a small in-group, show facility with understanding these complex rules and navigating the attendant processes. This leads, not unsurprisingly, to accusations of bias and unfairness. Doubtless the procedures were written with good intent, but the outcome appears not quite as envisaged whereby the arbitration process is applied to problems not suited to its consensus-based approach.

In the wider social machine context there seems to be a naïvety that the right blend of human and machine will somehow magically render a useful and concordant outcome. Complex community rules that are hard to understand, such as generated by Wikipedia, appear a poor bulwark against the effect of deep, atavistic, behaviours—especially when those behaviours are not necessarily consciously manifested. Put differently, policing the outcome rather than addressing the cause is perhaps the wrong choice. This does, of course, assume that territoriality is understood in the

[6] This term describes a folksonomic form of article categorisation but insertion of templates that setting membership of a particular category for that page

context of collaborative systems like Wikipedia, whereas the subject appears under-studied at present. In 'What Do You Know?: Experts, Novices and Territoriality in Collaborative Systems' [344] Thom-Santelli, Cosley & Gay sensibly conclude:

> "However, it is important to ensure that novice contributors do not feel marginalized or excluded because of the territorial expressions from expert collaborators. The challenge for designers lies in managing the inevitable points of tension that will arise when a novice asserts his or her particular perspective and the expert feels threatened by novice attitudes or activities. Successful collaborative groups will need to strike this balance in order to remain healthy and vital."
>
> ([344, p.1693])

The above call for a light touch ('balance') managing user interactions is very familiar from my years of experience of online forum moderation. Territoriality is hard to control in others, and just as hard in oneself where its actualisation can simply feel like application of expertise when correcting—or confounding—the efforts of others. From my own research, I sense that territoriality is not necessarily the preserve of a 'experts', in terms of subject-matter expertise. Indeed, and though this needs more validation, territorial behaviour is as likely triggered by the degree to which an editor feels invested in an article. This might be because it is close to their own expertise, but also may just reflect the editor's time invested in getting an article to a good state of completion. The latter is likely linked to the (again, informal) kudos attached to Featured or Good Articles[7] and reflects the nature of user edit counts as a proxy for expertise or authority as discussed above.

There are many quantitative studies involving Wikipedia, even if the actual research subject is tangential to Wikipedia itself. By comparison, but there are far fewer qualitative or ethnographic studies of Wikipedia that look at the nature of the system, i.e. the actual labour of editors within the wikis, rather than studies looking to map sociological constructs onto the system. In this regard, Sundin's paper of 2011 'Janitors of Knowledge' [336] is a departure from the norm. The paper's title reflects the unseen, less glamorous work that goes on unseen by most of the general public using the website. Amongst the activities Sundin records are: cleaning articles (formatting, typos, vandalism), acquiring references, applying citations to article text in support of retaining content. Compared to the early years of Wikipedia, there is now much more focus on ensuring facts are footnoted with credible reference sources. By contrast, some of these other 'janitorial' roles are likely under-resourced which helps contribute to observed technical debt[8] accrued by applying quick fixes to structural problems that would benefit from better maintenance, yet the task falls to no one and passes un-actioned. Such tasks are also familiar experience, from maintaining an ongoing hypertext in other setting: the 'content' is just the visible part of the iceberg.

[7] Respectively, Wikipedia's 'Featured articles' [W647] and 'Good articles' [W654]. These are considered the articles displaying the best of Wikipedia's content.

[8] Interestingly, this term was coined by Ward Cunningham, inventor of the wiki.

6.2 Collaboration And Dysfunctional Actors in Wikipedia

I turn next to the nature of some of the interactions between editors, both human and bot. In a social machine, disconnects or friction between the actors therein are unproductive. This does not preclude differences of view or, in the case of bots, differences over edit behaviours. Within the Wikipedia commons the main process to manage this is a principle of 'Consensus' [W633]. Editors are not supposed to act without consensus, but that laudable notion relies on there being others willing to discuss—and reach broad agreement upon—the issue at hand. The difficulty of achieving this in a voluntary setting was seen regularly in my review of bot requests. Editors cannot force others into discussion and even if discussion does take place, consensus is held to occur—or not—on the contribution of editors often numbered in only single digits. Thus an *assumption* of sufficient debate underwriting acts which are then deemed as consented to, does not hold up to scrutiny. The issue is not people trying to avoid consensus, but false assumptions about how the process works in practice as opposed to theory. Given the size of the endeavour as a whole, consensus-building participation is surprisingly small. But, these are all volunteers; even the bots only function at the behest of their volunteer operators.

It might be assumed that low number of editors engaged in consensus-forming does not matter due to experienced users being involved. But it is actually difficult to assess experience with clarity. As has already been related, user edit counts [W641] are not indicative of what the user knows. Nor is the longevity of the user's account. Nonetheless, regular participants are likely to recognise other regulars. This informal ranking works against the occasional user who lacks such system kudos though that is in fact totally unrelated to the potential contribution at hand.

Within the system (or social machine) this privileges system/process knowledge significantly over domain knowledge. Whilst the former may deliver better structure to the input, the result may be a loss of good content which has been shut out by the system's in-group. It might be expected that experienced users would impart their system knowledge to the less-experienced, but I saw little evidence of this beyond policing of actions or referencing Wikipedia's many procedural essays. The latter are often complex, weighted Wikipedia's due process in favour of 'experienced' users, or more precisely those with deep knowledge of procedure which can then easily be deployed to frustrate newer users of the system. Indeed, knowledge of the system—or rather its rules—is not a guarantee of better structure overall, but simply favours the biases of the in-group with expertise.

When human and technical (bot) activities first showed friction it was mainly over edit statuses. An early complaint against bots was that their editing style and speed made edit logs hard to read and cluttered human editors' Watchlists [W58]. This issue was partially resolved in 2008 by the addition of a Template 'Bots' [W174] whose purpose is to "tell bots and other automated editing tools and scripts that they either should or should not edit a page that has the template". In the same year, user watchlist options were added to allow users to suppress records of bot edits on watched pages.

Some of the early resistance by human editors to content transclusion appeared based on technical concerns that may have been valid at the time, but were which never clearly elucidated (Section 4.6.1.6). Thus, even if such concerns were contemporaneously correct, there is no suitably annotated reference to the problem such as might ensure later review and correction to guidance about transclusion.

In mid-2009, a now-retired editor 'GoneAwayNowAndRetired' [W257] is notable for their description of Wikipedia's dysfunctional nature in the essay 'Wikipedia is broken and failing' [W258]. Polemical in tone, nonetheless it offers a useful adumbration of the ongoing dilemma in the Wikipedia's regulation. The difference between the intent and exercise of consensus can be stark. Another earlier general essay, started in 2007, 'Wikipedia is failing' [W713] covers some of the same issue in a more granular manner. Whilst some of the issues mentioned have since improved, it was not clear during my research that the consensus and adjudication system is held in any better general regard by editors.

Friction between human and non-human actors also occurred over the perceived overreach by bots—or their operators. This is seen in RQ2 (Section 5). The nature of such disagreements are not easily or neatly categorised, but this may reflect the above observation about territoriality being misread as ownership.

Meanwhile Wikipedia does try to hold the line using both stick, e.g. 'What Wikipedia is not' [W706] and carrot, e.g. 'Ignore all rules' [W656] (or 'WP:IAR'). The last enshrines a concept well known to workers in structured organisations: knowing which rules to break, and when to do so. Here, however, such rules seem like weak containment for general process failure and whilst offering useful leverage to the process-savvy actor.

Thomas-Hunt, Ogden & Neale's 2003 paper, 'Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange within Groups' [347] concludes:

> "Based on this study, it is possible to conclude that some of the decrements in performance experienced by heterogeneous groups may be caused by self-censoring the potential contributions of socially connected members rather than self-censoring by socially isolated members as is typically believed. Socially-connected members may become so focused on maintaining their social connections that they focus more on contributing the knowledge possessed in common with those to whom they are socially connected rather than focusing on or sharing the knowledge they uniquely possess." ([347, p.475])

This point about affinity-seeking trumping actual useful purpose is important as it flags the fickle nature of human actors in the social machine. If the human actors do not behave equitably, what hope is there for the machines to do so—especially as they grow in ability and learn mimic human foibles? A mirror is being held up to the supposed harmony of consensus.

Niederer & van Dijck's 'Wisdom of the Crowd or Technicity of Content? Wikipedia as a Socio-otechnical System' [278] (2010) is atypical in taking a centre-ground approach between human and machine actors: "a third category: that of the many active users assisted by administrative and monitoring tools, also referred to as software-assisted human editors. Bots are Wikipedians' co-authors of many entries." [278, p.1378]

They make the point that looking only at one or other group of those actors gives a false picture if trying to comment on the whole. This brings readily to mind the metaphor of the blind men and an elephant [W7], whereby assessing (feeling) only parts of the system (animal) leads to fallacious assumptions about the whole. Depending which group of edits you take, your observational experience may differ when arguably it should not, especially if edits are essentially a unit currency of article construction.

Though well cited, the paper delves but shallowly into Wikipedia's structure limiting its insights. The weakness therein lies in the presumption that high-level quantitative abstractions reflect *accurately* the nuances with the observed data; it is not that their measures lack merit but unless the authors' qualify the rigour of their findings—i.e. in using only high-level counts—appropriate caution should be applied to use their findings. Despite that caveat, Niederer & van Dijck's point about over-concentration on the actors versus the overall system[9] does echo the picture emerging from my research here. High-level abstractions break down when contextualised by the working level structure the social machine.

From a high-level perspective, O'Hara's 'Trust in social machines: the challenges' discusses trust "it is likely that public trust in such machines will be highest when the public has had a say in their design and operation" [281, p.6] whilst also commenting that "Importing human interaction into the programming environment . . . presents a major challenge." [281, p.4]. These two observations suggest some contradiction, unremarkable at high-level abstraction but not ignorable at an operational level. The latter quote shows another instance of a shift from Berners-Lee's social machine as a balance of human and machine to one more in favour of the machines, in this case due to the unpredictability of humans.

6.3 Breaks In The Record

As noted above, it is clear that users set store by edit counts. This is apparent in the case of user 'HighInBC' who changed their username five times. Whilst effort was applied in the administration of the name changes to ensure the user's edits under the old name were still accredited to new account, the same could not be said of links from edit signature blocks in talk pages and in general edit logs. The disparity in the two outcomes is telling, with respect to editors' priorities.

[9] Again a subject for further work: this issue most likely occurs through narrow focus via over-concentration on methodology.

Archival activity shows similar functional breaks. Talk pages, as well as some articles pages in the Wikipedia namespace, like 'Bot requests' [W455], are archived regularly to remove moribund conversation threads from active pages. Unfortunately, such archival actions appears to be undertaken with little forethought as to later use of the archived information. Finding items that have been archived is time consuming, due to unnecessarily broken links pointing to the current page and not its relevant archive. Some users provide a dated archival index section for their talk pages such as is seen with user 'Anomie' [W203]. The most active talk pages can require a finer-grained archive index, such as using a discrete user sub-page to index the archive of the user talk page contents of user 'Rich Farmbrough' [W339]. Such affordances are the exception though, and while they neatly curate the user's talk page content, they do so with no regard to inbound links from else where in the hypertext. Individually initiatives like these help, to some degree, as if the date of the sought-for information is known, it is easier to narrow the search for where it may lie. However, even whilst allowing for the volunteer effort aspect of Wikipedia, it is not a good exemplar of archival for the long-term record.

An example of the hoops through which the reader needs to jump to trace editor activity is seem in the analysis of the thread involving RileyBot (q.v. Section 5.3.1.5 'Loss of Bot Operator'). In fact, for brevity, the analysis omits some of the additional work undertaken to resolve the case at hand. Such an experience was very common during my bot research and would suggest either few people trace activity using the hypertext (e.g. they use un-documented search methods) or they simply do not attempt these tasks except in extremis. That this problem is not seen written about would also suggest it's rectification is seen as either too hard or too big to resource, thus technical debt is left to increase.

It may well be the case that Wikipedia's archival practices to date have been a blend of expediency, system constraints and ignorance of good practice. Nonetheless the hypertextual breakage created is not self-repairing and is occurring in a part of the hypertext away from the 'many eyeballs' working on the wikis public articles. As (article) edits are an apparent if unacknowledged system for kudos and standing amongst editors, it is unsurprising that fixing elements of the hypertext away from the public get less attention. Talk pages are essentially the minutes of the Wikipedia docuverse. They contain information about, or intimations of, the many ongoing choices that inform the growth and structure of the whole. Quite apart from their interest to future researchers, the pages are far from yesterday's (digital) fish-wrappers and deserve better attention if only to avoid the significant degree of link breakage and opacity to the site's search features.

Summers & Punzalan's 'Bots, Seeds and People' [335] (2018) paper on archival is set in a wider Web context but their observations are pertinent to large social machines. They note the need for some of those engaged in the activity to have an understanding of what is implied by archival of content. In a Wikipedia context some editors may be archivists by trade but this does not mean that the skill is necessarily recognised, or called upon for betterment of the hypertext. The authors recognise that optimisations of human and machine contributions are not naturally

aligned so thought needs to be applied if trying to support deliberate archival within a large social machine.

As regards logging, only in recent years has the logging of important events such as the granting or removal of a bot flag been improved. In a very loosely controlled social machine like Wikipedia, it might be argued that those who need regular operational knowledge of this information already do so, albeit favouring elite human actors. Realistically, it ought to be an expectation that such information is efficiently recorded, with semantic structure, and with thought to enabling the non-human actors within the system to assist the social machine without recourse to human triage of (often partial) logs written for human interpretation. The 'reading' needs of the human and machine differ in the same way their 'intelligence' does.

6.4 Bots

In Chapter 5 I investigated bots from two aspects, the first as a stepping stone to the latter. The first task (RQ2.1) was to ascertain what bots are doing. The approach was not to judge their editorial contribution or 'social' behaviour, but instead tried to ascertain what they are actually doing within the overall Wikipedia social machine. Unexpectedly, but as has been shown in my research results, Wikipedia is unable to give a comprehensive report on its bot population. This occurs due to incomplete logging of key events (which is improving), poor archival practice (not improving), and a lack of clear instruction as the (minimum) information made available to other editors by the bot, or rather its human operator. Doubtless, the Bot Approvals Group (BAG) has a working overview, but that ignores the fact that such a report is only possible (even if accurate) by personal contact with the BAG. At the same time the BAG is a voluntary, self administering group. Malfeasance is not the concern, rather the failure to understand that at scale some facts are needed without having to wait for a person to wake up and answer the phone/email/message. A detailed report of bot activities (numbers, tasks, etc.) *should* be an automated report would the BA would monitor for accuracy.

Given the prominent role of bots in my research, it can be easy to accord them greater status than that given by their fellow editors. In reviewing the small number of books about Wikipedia, as described in Appendix F, Darlby's *The World And Wikipedia: How We Are Editing Reality* [95] is interesting as written from the perspective of an active Wikipedia editor. Although bot numbers started to climb rapidly in 2006, writing in 2009, the author makes no mention of bots. Were it antipathy, it would likely have created comment. Rather, they were clearly seen as unimportant, or at least to the editor's focus of article content. I find this a timely reminder that it can be too easy to focus on the work of bots simply because they are an easily identifiable sub-group of editors.

As Geiger's [134] 'Are Computers Merely "Supporting" Cooperative Work' points out, we have a poor understanding of the ethnography of bots in a Computer Supported Cooperative Work

(CSCW) context and thus must contribute, if only in part to the confused nature of out understanding of the bot ecosystem within Wikipedia. This echoed more recently by Gorwa & Guilbeault [146] discussing a taxonomy of bots for social media: "much of the recent theoretical and ethnographic work on bots is not in conversation with current quantitative efforts to measure bots and their impact." [146, p.22][10].

Müller-Birn, Dobusch & Herbsleb (2013) [253] investigated 'Algorithmic Governance'. An insight they give is of the degree of drift that can occur in a bot's editing focus. The difficulty in understanding what a bot was doing as compared to what it was intended to do is something I encountered in RQ2 (see Chapter 5). In my investigation of bot accounts, I was concentrating on the visible information on the bot's user page—and sub-pages and talk pages where necessary. But, in a social machine, bot-to-bot interactions are possible. It seems odd that groups like the BAG are not considering some form of machine readable data in the account pages to help bots understand what other such editors are doing. Bots are editors too, and in many respects are treated exactly as such. Thus, defining a bot in such machine-readable terms might also be useful in causing the controllers (BAG) and operators to think about how they classify bots' work. At present there is a mélange of descriptive approaches. Some describe the task in terms of the action "add a template to pages…", others in terms of (visible) outcome "tag pages belonging to…". Whether only one form or both are used, the fact the two may describe the same task is not at all clear. Once again, that is except for those who already know. They clearly have no such problem, yet have no incentive to think about other actors within the machine, and the way the lack of clear information obfuscates understanding of bots work; simply looking at the bot's edit record is not sufficient explanation. Not least, it is important to understand the intent and authorised scope of a bots work so that the bot's activities can be properly monitored. Expecting a general user to wade through the (often jargon-heavy) discussion on BRFA pages just to have a some sense of bot tasks shows a lack of consideration for other parts of the social machine.

Krafft's [206] 'Bots As Virtual Confederates' (2017) delves into the aspect of bots as assistants within the broader collaborative scope of CSCW and the biases we may add to them. In most cases, Wikipedia bots are essentially defined mechanical tasks where the degree of subjective judgment is little to nil[11]. But even seemingly rote quantitative work can be biased if there is a degree of choice as to what is enumerated and what choice is made, if those choices cannot be subjected to revue. However, it is not clear that the current BAG has the remit, expertise or interest to judge on wider issues like bias; in contrast, the wider editor base may lack the expertise to understand issues of bias as might occur in a bot content. Importantly, the preceding observation does not devalue BAG members' expertise or individual judgment but simply flags the wider context of consideration needed.

Long *et al.*'s 2017 paper ' "Could You Define That in Bot Terms?" [219] is most revealing. As discussed in Section 2.10 it draws attention to a fallacious assumption that most people know

[10] See also, Section 2.10.

[11] A notable exception is a bot like 'ClueBot NG' [W240] that patrols vandalism using AI/ML technology; most bots are far less sophisticated, down to the level of being operator triggered simple scripts.

"what bots are at a functional level, how they act and react to data, or how sophisticated they can and cannot be." [219, p.3496]. Their locus of study was within the 'reddit' [300] forum website. Their observation of a gap between the needs of the bot's customers, i.e. general users (human and bot), and the bot-makers was echoed in my analysis of bot requests in RQ2.2. The issue is not one of animus or laziness. Would-be users of a bot simply are unaware how much pertinent knowledge they lack to task a bot effectively. At the same time, creating and running bots clearly attracts people interested in that activity: that interest does not automatically map through to consideration of how others might use their bot or how easily they might engage with it. Thus machine resource exists within the social machine but its application is not effective due to this communication barrier.

A different aspect of this is considered in Cheon & Su's 'Configuring the User: "Robots have Needs Too" ' [83] which investigates how operators 'configure' their (ro)bot's users. In the simplest terms this can be seen as the range and manner of what they 'allow' their users to do. Cheon & Su state that "We have painted a picture in which roboticists vested in the human side of robots conceptualize users as in a transition from a naïve to sensible state. These same roboticists do not view their ideal robot as a machine that must become socially intelligent for users." [83, p.203]. Though the paper is written from the wider contexts of robots as a whole (i.e. not just software-based) the broad observation holds. In the particular context of Wikipedia this may seem harsh, as the human bot-operating editors are all volunteers, making it hard to place too heavy a remit upon the operator. By comparison, current bot oversight is more about making sure bots do not damage the hypertext by poor practice or over-stepping the remit of their approved task(s). Nonetheless, it does seem consideration about a bots users (customers?) may become more pertinent if the range of bot activities broaden. As is stated by Young, Wigdor & Kane "In fact, as computers become more anthropomorphized, human weaknesses such as bias may begin to apply to computers. Notably, bot weaknesses are applied consistently and efficiently. A human may become aware of bias or apply a biased view only some of the time. A biased bot will act with bias in an unwavering manner, and will do so on a scale not possible by human actors. Thus, bots pose both a great promise and a great threat to information quality in open collaboration communities." [392, p. 7] (2018).

That statement chimes with an observation in Reagle's *Good Faith Collaboration: The Culture of Wikipedia* (2012) that "when a person comes to be responsible for more than he or she can do by dint of will alone new responsibilities and authority pull taught a tightrope that must be carefully walked in front of an audience of one's peers." [298, p.135]. Although talking primarily about humans within the machine, this neatly encapsulates some of the early credibility problems of Wikipedia bots. The general audience treated them with some doubt whilst the bot operators saw themselves as simply being helpful and misunderstood any hostility. Longs's later paper (above, [219]) confirms this broken communication in a different large Web collaborative space. In Wikipedia's context, this is something the formation of the BAG tries to counterbalance.

As a social machine,Wikipedia is bound to an extent by what constraints may be enforced without volunteers withdrawing their labour. Ironically this can, indirectly, favour the users (at scale) as

opposed to the Wikimedia.org that runs the website. Thus the site can police a known action like copyright infringement without noticeable push-back from users, but it has little control over the subjects upon which editors choose to contribute. An example outcome of the latter is under-representation of articles about female scientists of note (though the root cause of that is far less clear, even if much conjectured about); there is a general social acceptance that the lack is bad, yet this has not resulted in significant new effort on the subject area. As science is generally a well-referenced subject area, it ought not to pose a challenge for the generalist editor to populate. Ergo, system/customer demand does not necessarily command individuals' contributions.

Perhaps reflecting the volunteer-sourcing of bots, I was surprised to find no use of a bot to track and report on bots: the accounts active, the tasks undertaken, and the like. Something of this sort was suggested at 'Bot requests' in 2007 and again in 2011, though never fully actioned in either case. Unlike many potential tasks this seems well-bounded, and would also help drive better documentation/description of bot activities. Such accounting of function is not a negative as some might assume; it can also offer an opportunity to show the volume and range of work undertaken. Work that would simply not be done to such an extent if reliant on only human editors.

6.4.1 Bots in A Collaborative setting

Kennedy [196] (2009) in 'Textual machinery' looks at the authorial legitimacy and agency of bots. The assertion "Bots are the unspoken textual curators of Wikipedia" [196, p.305] seems over-reaching as it ignores the agency of their human operators. Livingstone's 2012 thesis 'Network of knowledge: Wikipedia as a sociotechnical system of intelligence' gives a useful summary of bot organisation and activity [218, pp.182–191] though sadly the sourcing is almost entirely unreferenced[12].

Halfaker & Reidl [159] 'Bots and Cyborgs: Wikipedia's Immune System' returns to the issue of anti-vandalism as a bot role but in doing so gives an informative insight into the normalisation of the place of bots within the community of Wikipedia editors. Indeed, anti-vandalism is a prime role of bots (or mainly a single, complex bot, as explained below) within the social machine. Hall, Terveen & Halfaker note "If even a relatively small portion of these are vandalism, the effect on data quality in the community can be large. Further, such damaging edits could propagate downstream to applications that use Wikidata such as Wikipedia and Google Knowledge Graph." [160, p.64:11] (2018). Anti-vandalism bots started to be used as early as April 2006[13].

[12] The failure here is the lack of linking to Wikipedia sources, as opposed to 'traditional' reference sources such as papers, books, etc. However, the utility to the fellow researcher of the information given was diminished by the fact the reader had to hunt around Wikipedia to follow-up on the observations made. This is something I have attempted to avoid in the is thesis by linking where possible, in-text (for digital readers) and in W-prefixed references (for paper readers).

[13] See 'Tawkerbot 2' [W373].

At present the anti-vandalism task is carried out almost entirely by a single bot, 'ClueBot NG' [W240]. Though operating as a single bot account, its efforts actually represent sophisticated software including AI/MLroutines. The bot actually runs on a number of computers, all centred on the single task of detecting and reverting vandalism. Geiger [137] (2013) looked at the effect of a service outage of 'ClueBot NG'. Interestingly, although response time for fixes grew it was shown that the wider (human) editor group still gave the hypertext resilience. Some will read that as an implication that the humans can do a bot's job, but this ignores the aspect of slower response time, nor the other task from which human editors were diverted to back-fill for ClueBot NG. In fact it shows that, notwithstanding some contested reversions by bots, a machine action within Wikipedia is doing useful work. Bots may not be improving the hypertext directly but it is definitely acting as a bulwark against its degradation by deliberate vandalism or mis-directed edits.

Despite such useful work on behalf of the wiki, even this anti-vandalism tasking is not without initial problems. In the early days of such work, human editors would report unhappily that their well-intended edits were being reverted by bots, essentially being false positives to the bot. That experience and the inclusion of some AI routines has improved the service and anti-vandalism defence remains the most prestigious role undertaken by Wikipedia's bot editors.

Clément and Guitton [84] and de Laat [207], both in 2015, address supposed opacity and secrecy of bot governance, and bot code. From my own research, I found that while the governance is arcane and hard for the ordinary user to find, it is not *hidden* from them: there is opacity, yes, but secrecy seems an over-strong charge. As bots can represent significant personal effort by their creator, it seems not unreasonable in a context like Wikipedia for operator's to choose to not expose their code—at least as long as the wider Wikipedia community does not *require* them to do so. Geiger (2014) [136] reported on bots, their 'bespoke code' and its meaning within the wider Wikipedia. But, as I have already noted, objectively it is perhaps not fair to judge bot operators for not sharing their code when there is no requirement (or incentive) to do so. Lack of visibility may be just that rather than automatically implying some darker reason. In contrast, in cases of Retirement or Deactivation of bots there is a valid concern about lack of access to code. For bots with important roles within the overall system, it is sensible to consider some sort of code escrow, to give the community access to a bot's code in case of sudden loss of the bot or its operator.

Tsvetkova *et al.* (2017) [358] investigated edit reversions and the interesting notion that bots might 'fight' amidst themselves (rather than with human editors). A mainly quantitative study it falls short of useful conclusions as the numbers alone do not tell a full story. A generalised perception of quantitative Wikipedia researchers is that reverts are a just reaction to mis-behaviour. These may be the larger cause for such events, but without triaging out all the reverts for other reasons, the quantitative analysis is devalued. A simple example is that two bots working on the same articles but with different tasks and/or methods, might misinterpret each other's edits and revert them. In such a case there is no fight as such but simply poor configuration, texting and oversight. Geiger and Halfaker [138] (2017), building on their previous paper, from the same

sociological perspective, usefully recognise the weaknesses in the data when simply taken as a quantitative whole without bothering to tease out some of the underlying nuance [138, p.49:28].

Steiner's 2014 papers [326, 327] created a method for reporting bots vs. human editor activity, looking to report which group was making which edits and how many. The short papers are essentially an announcement of the 'tool' [58], which is still operational. Missing from this work is a sense of what these counts tell us, beyond tabulation on-screen. Again there seems over-promotion of the importance of edits as the primary measure of activity. It may reveal the presence of unexpected actors but little more. Edits are counted easily; assessing the cause and impact of the edits is much less easy.

6.5 Unclear Structures

I have already noted, especially in Chapter 5, the research difficulties posed by deficient logging and archival processes. This situation is not totally unknown as Jullien's 2012 review of Wikipedia literature [192] 'What We Know About Wikipedia', cites Doyle [107] back in 2008 commenting on the lack of standardisation in different types of wikis, (over-)reliance on search [sic] as means of finding content, and on a failure to consider re-use and data structure. My own research certainly reflected the tyranny of a system where access other than by direct API query, relies on keyword-based search, that resting informally as it does on a weak assumption that a simple keyword search will find any needed item. Not only is that assumption over-optimistic in practice but in turn this devalues effort with hypertext structure: why bother to attend to the latter when a user can 'just search for it'? But structure is more than a form of index or outline. It is important to a coherent hypertext, otherwise the Wikipedia docuverse verges on being a just a large bundle of co-located small texts. Considered and understood structure enables other features too. For instance, semantic data—even if unseen on screen—enables machine actors to operate more efficiently than simple word-level indexing and string matching can offer. In this sense, the Wikipedia favours the needs of the humans within the machine, albeit more by omission than commission.

A lack of description and understanding of tasks and structures, as described in Chapter 5, can lead to some potentially flawed analysis in both quantitative and qualitative approaches. One such example is the use of Wikipedia Categories[14], where editors and external researchers overlook two important aspects. Firstly, the Categories cover (at least) two discrete activities—proxy-indexing of article content, and the management of back-of-house content. There is seemingly no method or intent to de-interleave these tasks. Secondly, the Categories are folksonomic in creation, organisation and management with the inconsistencies inherent in such a method. I have also described some of the idiosyncrasies of category use in Section 3.1.

[14] See Wikipedia namespace article 'Categorization' [W631] and Help article'Help:Category' [W44] give an overview of the feature.

Outside small areas of common practice, such as some WikiProjects, application of categorisation templates is incomplete and not encyclopædic. Thus the act of simply fetching articles within a given category cannot safely be assumed to be all or even most of articles implied by that category. Yet categories are often treated unquestioningly as being complete listings of relevant articles. An example is Tramullas, Garrido-Picazo & Sánchez-Casabón's 2018 paper 'Use of Wikipedia Categories on Information Retrieval Research' [352]; I cite this not to take issue with the research subject but simply the lack of evidence that the authors comprehend they are pulling incomplete datasets when relying on categories to scope their search. The paper gives no indication of inspection of the completeness of the categories being researched. As use of categorisation is part of the research this seems an odd but familiar oversight. If the authors did indeed investigate the completeness of category coverage, it is something worthy of mention in the paper.

6.6 Whose 'Social Machine'—Democracy or Technocracy?

Turning, more directly to my own research, what does it reveal about Wikipedia? Before addressing this, it is necessary to point out that this is not a critique on Wikipedia's purpose or intent. Indeed, the totemic nature of the site often leads to zero-sum reactions to any comment about it.

Currently, Wikipedia is a prime source of information, with free public access, and available in more languages than there are officially recognised nations on the planet. Some may have concerns about validity and bias in articles, others about civility, yet others about diversity amongst both articles an editors. Those concerns lie outside the scope of my observations here. Instead my comments relate to Wikipedia's function as both as social machine (but *not* its role as a quasi-social network) and as a collaborative hypertextual repository of knowledge.

Having addressed the scope of my reply to the opening question, I will first use the perspective of a social machine. In this context, Wikipedia's mediawiki software storing the various wikis and their data forms a technical substrate to the social machine. From a socio-technical perspective that software's evolution reflect some co-creation by human and machine but nonetheless that occurs in the underlying environment system wherein Wikipedia exists. As such, the pertinent part of my research here is RQ2 and the role of software actors within Wikipedia.

Wikipedia employs a very flat structure. It has no real management structures but rather remarkably few layers of enhanced roles for some users (described in more detail in Appendix D). Roles are based around the additional powers accorded an account, which are implemented as 'flags' on the account[15]. In fact there is only one managerial role at scale with a few more assigned only to a handful of users across all the wikis. The broad managerial role is that of 'administrator' which, as some editors have pointed out, is something of a misnomer: indeed, some users employ the term 'sysop' interchangeably with the former. The article 'Expectations of adminship' [W428]

[15] The 'bot' flag is one such though an outlier in this context.

states that administrators "are Wikipedia editors who have been granted the technical ability to perform certain special actions on the English Wikipedia". Thus by described intent, they have no explicit social managerial role such as with a message board moderator or administrator. These editors are not like a teacher patrolling the school playground to keep order. Though administrators are involved in the bureaucracy of disputes, both for due process and judgment, that is mainly because there is no other designated group able to perform such a role. The 'sysop' label better defines their role and also reminds us that Wikipedia roles are essentially technocratic.

Though Wikipedia's software-role based role structure might suggest an ascendancy of the technocrat, the system's only machine actors—bots—are firmly under human control. This is true insofar as a bot account *must* have a human operator who is also held accountable for the action of their bot(s). My investigation into RQ2 did leave a sense though that at times the operators act as proxy for their bots in defending the 'rights' of their bots. This comes to the fore where there is friction between bots and human editors as touched upon above (in Sections 6.2 and 6.4).

It appears there were possible technical implications for use of (content) transclusion in the early days of Wikipedia (see Chapter 4 and Section 6.2), in terms of loading on the underlying database/website, but the facts behind such claims are not well evidenced. This has contributed to likely now-obsolete concerns still being present in the record without subsequent clarification or correction. That is unhelpful for the general user trying to make proper and effective use of structurally useful techniques like transclusion. Worse, it has made it harder for those trying to create a more maintainable hypertext structure. Wikipedia, and other wikis using mediawiki have been existence long enough that it ought to be possible to have created appropriate documentation on these matters. Once again, the volunteer nature of Wikipedia means such tasks, not involving mainspace articles, languish unless an editor is motivated to address the problem.

The structure of mediawiki's roles does show a weakness in that it has no mechanism to empower roles that cannot be expressed in narrow technical terms. This makes it hard to engage people with the skill and intent to work across the wiki's hypertext to ensure it remains effective and not accruing technical debt as appears to be the case at present. It further means that thoughtful nuanced use of hypertext such as by transclusion of article content or by calling smaller segments of data from Wikidata, are not incentivised (q.v. Section 4.6.1.6). At some point in growth a large resource starts to need more than ad hoc organisation. As the article count grows, there is increasing benefit in taking an overview so that content is planned in such a way that changes do not require multiple edits of the same data. Wikipedia's role as an encyclopædia means a requirement to cover a very diverse range of topics, one which perhaps makes the benefits of such structural work less immediately obvious to its editors.

So, is Wikipedia a social machine? Yes, if going by Berners-Lee's original definition "processes in which the people do the creative work and the machine does the administration" [35, p.172]. It is also evidently a socio-technical combine. These elements come together in two ways. Firstly, in that the entire enterprise runs entirely in a computer software environment, the Wikipedia website being its public face. The mediawiki software used to run Wikipedia, provides the both

the substrate for all the generalised activity and at a closer scope provides all the affordances needed by users. Mediawiki's code is forked from Cunningham's original wiki code and has evolved alongside its community of users, evincing a clear case of co-creation as users' feedback has been represented in the improvements made to the software.

The second element is the presence of automated accounts, the bots. Though this means there are both human and machine actors within the system, it must be recalled that the latter are only allowed with a responsible human operator. Thus whilst, bots are just one special form of editor, few (no?) bots really have the independence afforded the general human editor. Whilst bots may not (yet) be particularly contributory to the creative aspect of Wikipedia, the range of tasks shown in my research (q.v. Chapter 5) makes clear they are helping in an administrative capacity. By comparison, whether bots are, or are being directed to, undertake work deliberately in support of maintaining the knowledge in the hypertext is less clear.

Taking the more recent perspective on social machines, as related in the literature in Section 2.11 and at the opening of this Chapter, we see a one that subverts Berners-Lee's concept, with the technological (machine) side of the combine moving to the primary role. The original definition might thus be re-stated as "processes in which the people assist the machines to produce outcomes useful to society in general". In that framing, Wikipedia seems less of a fit as social machine, as the humans remain the controlling partner. Contrast that with human 'turkers'[16] making validated datasets in order to train AI, in the likes of 'Galaxy Zoo' [131]. Indeed, exactly to whom the benefit then applies is less clear. In the case of Wikipedia, it is one of the most frequently accessed websites in the world, visited by both humans and software agents. Machine visitors may use web-scraping or API accesses but they are effectively still users of the Wikipedia website.

Considering the question in this Section's heading, it thus appears Wikipedia social machine manages to be both. Its control hierarchy structure is remarkably flat, compared to most organisations. The lack of a large hierarchy makes it, notwithstanding the complaints of some editors about the complex arbitration processes and difficulties achieving meaningful consensus, a very democratic system (in intent, if not in practice). This is all the more impressive when acknowledging how coverage of very contemporary subjects inevitably draws in current social politics. In contrast, as described here and in Appendix D the user roles, the means by which the control of the site is effected, are decidedly technocratic. As a member of the community, you are what you may create, edit or delete. As stated above, 'sysop' is a more descriptive term for the 'administrators' who form the control backbone of the Wikipedia community.

As to discussion of how these structures work—or do not—in the context of maintaining knowledge in hypertextual systems, I will discuss that in the next Chapter.

[16] Such as those working for the 'Amazon Mechanical Turk' [W1].

6.7 Summary

Does the synthesis above tell us anything other than that the social machine concept provides a portmanteau high-level term for human and computer actors operating in the same context? I believe not. Rather, the discussion above indicates such high-level constructs are less useful at giving insight into the internals of such systems, not least because there is methodological pressure to make the reality fit the model rather than make the model reflect reality. The generalities of high-level constructs mean they lack sufficient nuance to reflect the finer processes therein.

Clearer from the research is that, to me, the social machine is not a helpful model for exploration of preserving and maintaining digital knowledge. In the main it is due to the emphasis on the actors, but disconnected from to the outcomes of their efforts. I would also observe that the primary gain of using Wikipedia as the subject of the research is to draw, non-judgmentally, from its mistakes and weaknesses in order to build better structures. This is the subject to which the next chapter turns.

Chapter 7

Sustaining Knowledge in Hypertext

> '*Paper simulation is retrograde. Imitating paper is to me like tearing the wings off a 747 and driving it on the highway like a bus. WHERE ARE THE CONNECTIONS?*'
>
> Nelson's memoir *Possiplex* [272, p.332]: 'How do we improve on paper?'

> '*Using a computer should be like flying a jet plane through information space.*'
>
> Allan Kay, quoted: 50 years of Hypertext at Brown, Session 3 @41:26-41:33 [9]

Having reviewed my research as a whole in the previous chapter, I now wish to explore its implications in the wider context of hypertexts. I will also consider where some now-abandoned pre-Web hypertextual ideas might usefully be applied to longitudinal support of ongoing hypertextual resources. How then does this relate to hypertext both in the context of an organisation's ongoing knowledge of its business processes and to my thesis? Using large, open, collaborative sources such as Wikipedia for research makes for easier access to a suitably large dataset, and review/replication of findings. Nonetheless, Wikipedia is a quite unusual community, being part encyclopædia and (to a lesser degree) part social network. As such, it is not necessarily a close reflection of large information systems within public or private sector organisations. Thus my use of Wikipedia as my research source is as an accessible exemplar of a large collaborative hypertext rather than being about Wikipedia per se.

7.1 What Has Been Lost?

During my research, I was reminded of the absence of several earlier hypertext techniques that are pertinent to the sustaining a large hypertext. The first of these is the concept of link types (q.v. Section 2.4) which were applied in early hypertext systems that were exploring ideas of capturing

and analysing argumentation in a document. This task of adding such augmentative link metadata can to be seen as an encumbrance to authoring in many settings and the idea fell from favour. In today's larger systems, links may be thought of optional link metadata to mark key paths (trails [sic]) through the an organisation's docuverse. Applying link-type metadata was not easily possible in early HTML, whereas today such metadata could now easily be applied via CSS (it then being also accessible to JavaScript). Such a pragmatic method of adding metadata does not invalidate use more formal argumentation e.g. via Resource Description Framework (RDF), but for the general user the latter is complex and arcane for a general user base. By contrast, CSS is familiar to a wider base from its use in visual styling webpages and doesn't negate parallel or later use of more powerful methods. Regardless of technique, methods making it easier for authors and maintainers to capture link intent or, relationships otherwise unclear to semantic parsing, would aid the hypertext. Not only would human actors benefit, but such metadata could help formal analytic processes establish a better connection between human and machine understandings of the docuverse's linkage structure and purpose.

A contingent aspect of link-typing might be a light form of trail marking (in the absence of a formal linkbase). This is also a pragmatic approach. Non-Web hypertext's use of link-bases is rich, as seen in Microcosm's pre-Web use of multiple, swappable, link-bases allowing contextual changes in links offered to the user. Nonetheless, even in contemporary Web specifications, a light method of using CSS to mark chains links between of notes of interest offers a simple bridge to implementation of more heavyweight formal methods and gives a means to indicate deliberate paths through the hypertext. This makes a step towards implementing trails in the hypertext. Personal experience making hypertext-based documentation for public access indicates different groups of readers often need different paths through the same nodes in order to understand a process: indeed, mapping reader perspective and vernacular at the boundary is also useful in allowing users to boot-strap themselves so a comfort with using the corpus. This acknowledges differences in user expertise, task vocabulary, and perspective.

Another interesting area of hypertext languishing in the shadows is spatial hypertext (SH) (q.v. Section B.3). In the SH concept, information is plotted on a (normally) 2-D plane–or map—meaning is inherent in, and derived from, relative positioning on the map. Formal links between notes may be drawn between items or, alternatively, existing links may be hidden so as not to imply unintended non-SH relationships. This sort of open-ended mapping space can be useful for investigating complex hypertext resources allowing experimentation with different structures (or linkage).

Currently, SH systems tend to map smaller, closed hypertext systems but the concept could also be useful for exploring *parts* of larger hypertexts to review distribution of content across nodes and possible (improved link paths). Such localised views also aid link triage when revising the hypertext, especially the link triage involved when a node is split or partially duplicated. In the latter case both inbound and outbound links might need to replicated and that can become complex if only working from a textual representation of a node , e.g. the source of webpage or wiki-page, or the rendered form of a document.

A different visualisation is a link map, whereby *only* notes with links to the current (sub-)network of the hypertext are drawn. This is helpful when looking at the structure and completeness of a hypertext in terms of how nodes are linked—or note. The view cap map all links types or just some and an gives an abstraction of the hypertext as a set of links.

None of the features described above are readily accessible as tools to today's maintainer of hypertextual data stored and viewed in web (compatible) format, an issue I will return to further below.

7.1.1 *Où Sont Les Bibliologistes d'Otlet?*

Paul Otlet's work (q.v.Section 2.2.1) mostly took place in the early twentieth century before both mass communication at scale[1]) and the advent of computing. His work is thus rather overlooked, but his late-career (1934) oeuvre *Traité de Documentation: Le Livre sur le Livre* [284] has interesting utopian ideas for the recording and managing of knowledge. The work is only available in the original French, but via Wikisource and Google translate, I have included short sections from the book's Chapter 1 (introducing 'bibliology') and Chapter 2 (the definition of 'biblion') which are included at Appendix N and recommended to those unfamiliar with the text. Looking beyond the linguistic style and the book being the erstwhile main 'technology', Otlet's overall vision has worn well.

His notion of the 'biblion' [284, p.43, Section 211.3][2] was a unit of knowledge akin to the atomic particle of matter. Using biblions derived from existing knowledge, new and more useful knowledge resources could be created. It is not much of a stretch to see the biblion in our current notions of structured data and resources like Wikidata. The work of deriving and re-assembling these biblions was the work of what he termed 'bibliology' [284, pp.9–11, Section 11][3]

Thus Otlet also saw a need for new specialists other than authors, domain experts and librarians. He realised that the overall knowledge ecosystem, the docuverse itself, needs care and attention. Order in the system is not simply an emergent property of adding more content.

In the light of my findings here, Otlet's ideas and those of Nelson seem to mesh well, with the shortfall in the care and attention being given to our emerging organisational docuverses.

---

[1] The technology for this was invented but not in use at the scale of today.

[2] Further detail is at Appendix N.3.

[3] Further detail is at Appendix N.2.

7.2 What Does Wikipedia Teach Us About Knowledge Preservation?

The strongest insights come from some of the less successful aspects the observed in researching Wikipedia. First, however, it is important to repeat that whilst those insights are useful to the overarching question of the effective preservation of knowledge in digital hypertext, it would be quite wrong to read that as a critique of Wikipedia. Why so? Wikipedia operates in an unusual space. It is open to all, an almost entirely voluntary effort, and attempts to cover all aspects of human knowledge in encyclopædic form. This places burdens on the organisation (or community) which are less applicable in the docuverses of our society's large public and private organisations. The context of my discussion here is about the preservation of professional, process and organisational knowledge in large organisations, public or private. For instance: national or local government bodies, hospitals, emergency services, large charities, commercial companies or professional bodies. In all these other settings, there are other but differing constraints of privacy, sensitivity, and legislation. So whilst a knowledge system may be fairly open within an organisation, publicly it cannot be as open a resource as Wikipedia.

A positive aspect of such limitations is that tighter constraints make it easier to set considered rules which might seem burdensome in a fully open system. The latter, with care, may thus address some of the short-comings I will describe without the challenges faced by doing so in a Wikipedia-like environment. It is also important not to conflate systems running on Free and Open-Source Software (FOSS) with the voluntary nature of Wikipedia participation. Each of these does not require the use of the other to function normally.

I will turn to address my findings via a series of topics, noting that none are truly discrete and inevitably there is some overlap between the ideas in different sections. Being parts of a whole there is also no obvious order of presentation.

7.2.1 Docuverse or Digital Heap?

As stated in the outset, society is in transition from paper-based storage of its knowledge into increasingly digitally-native, and indeed digital-only, systems. This movement has occurred without a guiding plan and this has inevitably led to some gaps in our understanding of the new medium. A result has been that organisational knowledge which used to be in a printed form (and likely out of date) is now stored and used, in digital form. In the past, letters, reports and documentation would be stored discretely with minutes stores alongside their relevant documents. New digital systems and tools now blur these distinctions.

This means that there is a blurring of these old print-era distinctions of purpose as documents now reside in a common docuverse and minutes and similar notation—the documentary metadata—is often dispersed amidst sprawling email archives, apart from the source material. At the same time, human organisational knowledge stores such as registry staff have been lost along with the

paper they curated. In the absence of such a trusted resource, we are now overly dependent on keyword search. As yet, yet our efforts with Artificial Intelligence (AI) and Machine Learning (ML) that are supposed help our searching are still only at the early stage of understanding the nuances of human written and spoken language.

This move from documentary bureaucratic practice of old to a form of large hypertext, also means skeins of implied, if not literal, links between documents and their minutes. These marginal notes have moved mostly into email, though as evidenced by a experience of secondment to a UK Government Ministry in 2019, it is hard to link minutes back to documents effectively without now-missing metadata, i.e. the minutes. In physical storage the 'link' would have been co-located in a common folder[4].

In modern office applications, documents have aped the wiki-like system of being a massive database where a rendered document is a cumulative series of edits (though the latest render may be cached for efficiency purposes). In principle, the human users annotate the important events in the system, but that task is made difficult by weak affordances for capturing metadata and limited effective training on the needs of this new medium. Instead I observe people *thinking* in terms of paper but storing their knowledge and processes in digital systems in a manner that lacked such clear delineation as seen in older paper-era secretariats. Further obfuscation of document relationships occur as emails (minutes) may link to a document at creation but not expressly to the then-current version at time of linkage; the true linkage is thus not necessarily captured.

Within this digital 'heap' an organisations staff are forced back onto brute force keyword search and trying to guess the keyword needed to find content known to exist but not easily brought to view. As Web-based search engines appear better at search[5], unsurprisingly staff will search for public-facing documents from *outside* the organisation, whilst email inboxes serve as proxy filing cabinets with emails holding links to otherwise hard-to-find information.

It may give comfort to some to insist that different digital data within a common store is discrete in the way that it was with paper, but that is to cling unhelpfully to concepts of an era now past. Technology may have made IT infrastructure faster, more pervasive and more stable, but is less clear it has made information any more easily usable than in paper-based systems. This experience exposes a lacuna with respect to how information is organised and supported *within* the overall IT-supported storage medium. Our organisations' digital docuverses lack a clear champion, and the attendant problems are only increasing whist this issue is unaddressed.

Increasingly stable, pervasive and persistent IT systems make digital systems available—and restorable in case of failure—but the IT function does not necessarily collect or curate the knowledge within the organisation's docuverse, that being a job for 'somebody else' in the organisation.

[4] Traditionally, in UK, most often with minutes separated and placed on the left of the folder and with the main papers on the right.

[5] This reflects the invisible amount of hidden hardware and software applied to the task, but at cost beyond the reach of most individual organisations.

7.2.2 A Voice for The Docuverse?

Yet if our organisations' knowledge now resides in a docuverse, whom within the organisation speaks for the docuverse and the *knowledge* therein? IT services within an organisation provide the means of storage, availability and access control to an organisation's docuverse, but it is far less clear as to the their role and responsibilities towards the *content* within it. Realistically, access to information within the docuverse is more of a 'business' task though there is no commonly known and recognised owner of that role. Subject Matter Experts (SME) may contribute to the corpus of documents but are less likely to have requirement to nurture its general health. 'Knowledge Management' business functions are generally closer to management *accounting* so are more concerned with higher-level considerations of risk, value and cost than with the details underpinning that overview accounting, i.e. the docuverse's content. Merely accounting from *outside* the docuverse does not impart much insight as to what lies *within* it.

Security staff may control access and set policies for privacy and release of information but again may have no responsibility for the docuverse content beyond that; policies look the scope of access granted to a user, not the detail of the content within that scope. The organisation may have archival responsibilities but these often relate to removing and 'freezing' content so it may be stored without being subject to further change (reflecting issues of the paper era). The main business function, be it public or private, is mainly ensuring the systems work efficiently to deliver their needed role. Thus they too have little need to understand the docuverse, as long as it works.

As systems grow and information storage makes content less tangible and visible only in small parts at a time, there is thus a need for someone within an organisation to have an understanding of the structure and content of the docuverse. This begs the question of how to fill that role, and a question to which I now turn.

7.2.3 Roles for the content maintainers

The research undertaken here has shown that within Wikipedia there is little organized and directed control of the hypertextual structure. A wiki is an essentially flat structure, without an index; new articles are added as needed, to create another item in the flat bundle of articles. Wikipedia has folksonomic categories that mix content classification (a incomplete quasi-index) with flagging articles for bureaucratic attention. The emphasis, unsurprising given Wikipedia's genesis, appears to be focussed on the 'shop-front' by keeping up a steady supply of Featured and Good articles. Behind the scenes, the quality of finish is not the same. A utilitarian argument would be if it works, simply leave it be. But that way technical debt grows. Currently, over *half* (58.52%[6]) of article pages in the English wiki are actually re-directs, normally arising from articles being

[6] Of 12,188,486 articles in the 'en' wiki main namespace, 7,132,675—58.52% are redirects. or put another way only 41.48% of the wikis articles are actually articles. In fairness, Wikipedia does not account re-directs as articles. However, this hidden rump of articles is a good example of accumulating unaddressed technical debt.

renamed or merged, or improvements in the syntax allowed in URLs. As yet they do not cause problems, yet it seems counter-productive to have so many instant fixes behind the scenes. This brings to mind Alan Kay's comment in 'The Early History of Smalltalk':

> 'A twentieth century problem is that technology has become too "easy". When it was hard to do *anything* whether good or bad, enough time was taken so that the result was usually good. Now we can make things almost trivially, especially in software, but most of the designs are trivial as well. This is inverse vandalism: the making of things because you can. Couple this to even less sophisticated buyers and you have generated an exploitation marketplace similar to that set up for teenagers. A counter to this is to generate enormous dissatisfaction with one's designs using the entire history of human art as a standard and goal. Then the trick is to decouple the dissatisfaction from self worth—otherwise it is either too depressing or one stops too soon with trivial results.' ([195, p.563 para 2])

Though Kay wrote of a different context, there is analogy in the way articles are made easily, because people may do so. Meanwhile, more complex hypertext work towards structure and stability requires greater thought and effort but actually goes under-resourced. Thus we see sticking-plaster fixes of chained redirects, instead of pruning the excess material or providing better link resolution. The latter is a more complex task than just writing or sub-editing articles, but the work is no less important.

Yates, Wagner & Majchrzak 'Factors affecting shapers of organizational wikis' [391] offers the notion of 'shapers' within the system, summarising them thus:

> "Shapers are community-asset builders today and into the future. They do not simply organize knowledge; they ensure that the "signal" does not get lost from the knowledge noise. They do not organize people, meetings, or tasks. Instead, they use Web-based technologies to organize language, words, and links. They do not move the community forward through technical leadership by asking the "right" questions or offering answers or solutions. They instead move the community forward by looking for differences and similarities in perspectives, integrating where they can, and highlighting the differences when they cannot. They do not see themselves as thought-leaders or experts, as idea-generators or innovators, but are looking for new business opportunities, and recognize that the place to find them is no longer by walking around the laboratories. They recognize that new business opportunities are found on the corporate intranet among individuals who are sharing knowledge not because they have altruistic interests but because it is the way work gets done."
>
> ([391, p.553])

This is another notable difference from most interpretation of collaborative space and reflects some of the missing skills already identified. In a knowledge docuverse the issue often is not a

lack of domain expertise but the inability (or unwillingness) of engaged members of the group to integrate individual inputs. Thus, I would suggest it is worthwhile that shapers are investigated to see whether they demonstrate a lessened sense of territoriality.

Faraj, Jarvenpaa & Majchrzak 'Knowledge Collaboration in Online Communities' [120]

> "In most OCs [online communities], knowledge collaboration takes place despite the absence of existing social relationships. The lack of traditional structural mechanisms appears to partly free the collaboration from concerns of social conventions, ownership, and hierarchy." ([120, p.1225])

This is interesting counterpoint to contemporary presumptions about online communities needing these 'missing' social structures in. Such activity may simply reflect that for those who like social networks, every online community is imported improved by becoming a social network, an evidenced notion.

Close inspection of the Wikipedia's roles, or 'user groups' [W699], shows that even if some of the role names sound managerial they are essentially a set of technical switches. Co-ordination and curation are emergent by-products rather than an a role-based outcome, and this strikes me a lost opportunity. The use of edit counts as a proxy for experience and standing may have had some pragmatic sense for Wikipedia, especially in its early growth phase. Yet it is less compelling in a mature system, or those where there is not a need to attract and retain volunteers. Such incentives to participation also have potential negative effects as they privilege those with longest involvement and a better understanding of procedure: his can be exacerbated if process documentation is not clear and constantly maintained.

iRegardless of the motivations of general users, a harder task is to find appropriately skilled personnel to sustain the docuverse. In systems not bound by the need of (presumed) open consensus or dependency on volunteer effort there is scope to add roles that support those with and interest in or motivation to sustain the stability and structure of the hypertextual content and to resolve technical debt. For those working across the hypertext as a whole, this represents a mis-alignment with boundaries presumed by those only working within a particular area of expertise. That leads to issues of trust between users.

7.2.3.1 Better Trust Structures

Although I believe further research is needed to establish the full detail, the indication from my research here (q.v. Section 6.1.1) is that territoriality is an unrecognised and unresolved issue in large collaborative systems. Given that hypertextual maintainers ideally need a purview encompassing the whole docuverse, there is a need to create trust structures so that the needs of the SME and the generalist are not a source of friction. The domain specialist will not necessarily understand what may appear to be interference in 'their' work even if it occurs for the betterment of the whole system.

7.2.4 Support for Transclusion and Semantic Interfaces

In Appendix C, I note how wiki-style transclusion lacks Nelson's original notion of a fixed source, thus degrading the provenance-preserving aspect of the method. However, there appears no practical reason why a wiki system like mediawiki could not adopt a variant of its existing transclusion method, one that uses a 'fixed' source. For instance, the transclusional mark-up code could add—as a parameter—the edit ID of a specific edit to act as the source of transcluded data, as opposed to the current behaviour of using the source page's most recent edit.

Such a enhancement might need some adjustment in caching behaviour to ensure transcluded past edits are as accessible as the current page versions. Another accommodation, with the reader in mind, is making it plain in the rendered content that the type of transclusion used is from a fixed (edit) source, i.e. differing from hitherto 'normal' mediawiki transclusional behaviour. With suitable descriptive markers both types of transclusion could feasibly be employed side-by-side as both mechanisms have valid uses.

With all the effort being poured into AI techniques like ML and Natural Language Processing (NLP), more use could be made of adding semantic markers so that even where the topic does not lend itself to use of semantic data (such as Wikidata as a source for Wikipedia) we should be looking for ways to aid software agents to better understand human structuring of data.

7.2.5 Visibility & Control of Bots and Machine Processes

Despite the work they do, outside several high-profile tasks like anti-vandalism, bots have a somewhat limited role in Wikipedia. Though hard to evidence by direct reference, it appears bots had something of a false start in the hypertext, being seen as editors with a unfair advantages (e.g. speed) over human editors. All that in the context where the editor's number of edits may play into their standing within the community. In the wider context of an organisation, hypertextual edit-kudos is unhelpful as not all contributors will have the same need to contribute in the same volume. Thus, it means alternate and more appropriate methods of encouraging involvement and rewarding useful input need to be found.

A notable non-use of bots is in accounting the action and purpose of bot processes. For this, the Wikipedia model of bots having an operator (i.e. owner) merits re-visiting, at least in systems other than Wikipedia. Bots should not be in a position of working against human users rather than working for, and alongside them. Having clear and easy insight as to bots' work would help in this regard. Were the documentation and set-up of bots to be better formalised, and were bot (task) approval to require clear human and machine readable descriptions, this would result in a less opaque bot ecosystem within the docuverse. Likewise, better control and reporting of automated process like bots. If, as suggested, the approval process demanded a full and understandable description of the task it would be much easier to generate reports and dashboards showing approved bots, bots actually currently at work and tasks being undertaken.

The issue of recording the bots purpose more clearly is non-trivial in its potential gain, when set against the confused picture seen in Chapter 5. The act of describing something generally gives insight—to the attentive author, at least—about the task. Clearly documented, it is more likely to accessible to a wider readership transferring to them a clear picture of the bot's work.

A volunteer-based system like Wikipedia suffers from standardisation problems because any user can start a new would-be standard. Users do this, as evidenced by the mess of transclusion mark-up as summarised in Figure 4.2 revealing duplications of process and abandoned/incomplete methods. A useful initiative here would be to ensure standardisations are—where possible—encapsulated in a template or other standardised format so ensuring both human bot contributors can produce consistent outcomes.

Noting Long's description of bot creator/customer mismatch (q.v. section 6.4), the docuverse's maintainer need to ensure focus remains as much on bot process and function (for the customer) as it does the writing of new code (the creator's pay-off).

7.2.6 Boundaries and Trails

At less than whole Web scale, any hypertext will have boundaries. Indeed within a large organisation, there may be sub-divisions although such delimitations are worth regular revisit to see whether they have a useful function or are simply inherited past practice and no longer needed. With new digital systems, it makes sense to not place constraints where they have no function reason to be.

Any organisation will still have outer boundaries and this presents yet another task for the hypertext maintainer. A primary boundary issue is link breakage, 'link rot': links that used to work no longer do so. the breakage may occur outbound or inbound. The former are slightly easier to address as the arise inside the system and can be regularly tested for breakage, though this leave a process (human or machine) to revolve and correct the error. Has the target object disappeared or just been replaced? The latter leads to a slightly harder outbound case where a link works but is superseded by a new document with a different URL. The old URL may work but no longer meet the rationale for the link, e.g. linking to version of the reference *current* at link creation, rather to any superseding version of the reference. This in turn reflects a need for better curation of such key external references so that changes in the latter can cascade as changes or updates within the local docuverse.

The problem of maintaining hypertext linkage was explored by Meneses et al [244] who looked at errors in outbound links from the ACM Digital Library[7]. They showed the difficulty in knowing when links require attention and the degree to which human intervention may be required. Whilst the paper concentrated on the performance of the techniques applied to detect and correct

[7] 2,001 discrete links from the ACM Digital Library to ACM conferences and workshops were analysed. Only 75% of those returned '200/OK'. On inspection of the latter, only 61% were actually correct. Overall, less than half (46%) of the links sampled were correct.

incorrect links, it begged a wider question as to the roles involved—whose task is it to manage this sort of hypertextual maintenance? Clearly, as they show, machines are as yet not fully competent to manage the whole process without human input and supervision.

Inbound links may be harder to diagnose; if such a link is not used the recipient system may be unaware of the possible inbound link. But, within a system, there is scope to provide triage especially for the most accessed public URLs. The ability to do this connects with the point above about building in as much semantic structure as possible, short of forcing looser human-generated content structure into the comparative inflexibility of machine logical strictures.

The hypertext's boundary and portal page(s) also an offer a place to handle issues of vocabulary and vernacular. Visitors from outside the organisation may use different terminology and this is another consideration that the hypertexualist must consider as individual editors within the system are unlikely to do so. If the system offers trails of any kind this is a further task for the hypertext maintainer to consider being another task benefiting from having an overview of the hypertext as a whole rather than imagined via the abstraction reports or rendered pages.

7.2.7 Tools: Understanding The Task of Maintaining A Hypertext

As the last section reveals, unlike a general contributor, the hypertext maintainer needs to have an broad overview of the system. A complicating factor is the lack of knowledge tools available that allow mapping and annotation of the docuverse and processes therein. Network visualisation tools exist but are generally optimised for uses like mapping social networks, or subjects with rigid data structures - neither are an ideal method for mapping a hypertext which generally lacks a fixed structure or hierarchy and has interconnections that can not be unravelled into a simple graph.

As well as tools, those operating in this space are likely to require a different and more pervasive access to data than the ordinary staff member. In turn, this implies new consideration at the Information Technology (IT) level as to how to mix this richer service provision with the standard offering to office workers. It is a nuance too easily missed. To work across the organisation's docuverse requires access to all its content. Also overlooked in that is the attendant consideration of the privacy ans security clearance of the human actors. Even if not interested in content beyond its role in the docuverse, they may still need to read it to understand its relevance.

The ability of hypertextual workers to protect and nurture stored knowledge would be greatly helped by better means to visualise *parts* of its structure as the whole may form too large or complex a diagram for the task at hand. The need is not to map fixed (semantic) structures, but rather tools that help investigate the less orderly manner of human thinking and writing. Similarly if splitting/merging pieces of knowledge in the corpus, careful link triage is needed. Our lack of recording our link intent and purpose, removes access to visualising it making it harder to hand this task of off to an algorithmic function. Algorithms are powerful but also dumb, as in having no wider intelligence in the human sense; here better semantic hinting or explicit metadata could

help close the gap. I believe SH is also overlooked in this regard as it offers an exploration space not replicated well in other visualisation methods.

7.3 Summon The Hypercorps

Ted Nelson had a notion of a 'Hypercorps', which he described in *Literary Machines*:

> "The Xanadu Hypercorps is expected to be an unusual and elite group. They will circulate amongst Xanadu stations transmitting skills and outlook. They will not be people who can program or repair a computer; rather, ...Like good librarians the Hypercorps will have an understanding of what materials are available, but they will know how to deal with an avalanche, rather than a trickle, of ideas and information. Like a good teachers they will have a sense of how to convey ideas. Like good woodsman they will have a sense of the trails and byways of the territory to be explored. And like academics they will have a personal love for one or more topics that they will watch and study in their free time on the system." ([266, pp. 3/17])

This concept has resonated with me since first reading Nelson's book. The sustenance of knowledge needs something like this but a problem we face is it is not yet clear what those skills are or where we should start to look for them, though look we should. Today's hypertexts may not be Xanadu [266], but the need for people who know, care about, and are happy to explain a hypertext's structure is both pertinent and pressing as we move into the post-print age. This notion of a Hypercorps seems to blend with the bibliological work envisaged by Otlet and perhaps not (yet) well serviced in our current hypertextual docuverses. Whilst the discrete skills needed are not new, the combination of them and the context of their application is.

7.3.1 Skills for a 'Hypercorps'

In advance of research in this area, there is no definitive list of such skills, but drawing on my experience during researching this thesis and from 20 years of previous hypertext work, I would suggest the following skills are pertinent for those who might be called upon to help us make sense of large organisational hypertexts:

- A lessened sense of territoriality
- Being as comfortable looking at the big picture as much as the minutiae
- A sense for the balance of node size and trail opportunities
- An understanding of link networks

- A sense for granularity within knowledge
- A good grasp of the role of metadata
- An understanding of co-creation of content within social machines, by humans and computers

7.4 Summary

In terms of the observations here, a simple summation is that our print era house is now resting upon post-print sand. We need to work on new (digital) foundations. The experiences described above show me that we have decoupled our technology from our understanding of it, albeit without deliberate intent, but such an outcome is still the by-product of progress. The loss of paper records and their attendant secretariat, with the institutional and process knowledge enshrined therein, is not necessarily being replaced with document metadata and process semantics that offer a similarly detailed knowledge of an organisation's documents.

Not having clarity as to the skills typifying a person who understands the digital docuverse is problematic. Without a clear description of that role and the tasks arising from it invites the question of how people with the requisite abilities are to be found and recruited. Understanding the organisation's knowledge is too valuable to simply be sub-contracted out, all the more so if some information is sensitive or requires limited distribution. A tendency in Computer Supported Cooperative Work (CSCW) settings to leaving tasks to 'someone else' is not credible in this setting.

An unseen change with the increase in digital use is how the support role of IT and the users of their systems have drifted apart. At organisational scale IT is generally much more stable and persistent than in the past. If systems fail, IT can rapidly return it to last used state. However, IT staff are not the librarians of the docuverse, though many users may assume that they are. Without easily accessible information, users are reminded that internal search tools are weak, forcing users to ask other *humans* for help or relying on using their email client as a proxy filing system.

There is a lacuna of expertise and support within the docuverse itself. The IT function acts as an underlying substrate to the docuverse, whilst the users work atop it, interacting to the degree that their application skills allow (thus generally in a shallow fashion). This is the gap we need to resolve and appropriately resource.

Chapter 8

Conclusions

'*The past of paper is a foreign country; they do things differently there.*'

. . . after L. P. Hartley's *The Go-Between* [168, p.1]

'*No note is an island,*
Entire of itself,
Every note is a piece of the hypertext,
A part of the main.
If a note be unlinked by a slip,
The docuverse is the less.'

. . . after John Donne (1572-1631)

In this final chapter I will summarise my thesis' contribution and then note my recommendations for further work in this area that arise directly and indirectly from my research. I will then give some draw out some general recommendations, not already covered under future work) that may be pertinent to those running knowledge containing collaborative hypertexts. To close, I will offer a finally summary.

8.1 Contribution

Although Wikipedia has formed the focus of my research, it has been in the guise of a large, open, collaborative hypertextual docuverse. Therefore the thesis is not a judgment on Wikipedia but rather the functional and organisation problems posed for sustaining our stored our knowledge as it moves from paper into digital hypertextual spaces.

In the course of my research and its subsequent analysis, I believe I have made a number of contributions, as described below. Firstly, I will cover the contributions arising from the two primary pieces of research, before raising some emergent points that add to the overall contribution.

8.1.1 Planned re-use of article content

Though mediawiki makes significant use of a (non-Nelsonian) form of transclusion, I have shown that transclusion of either type (tied to original or current source version) is not used in any pervasive or consistently planned manner within Wikipedia (q.v. Chapter 4). This finding has also been replicated in the wikis for ten discrete languages within the overall Wikipedia website. The sampled wikis represented the most of the largest largest wikis by article count and included languages that do not use the Roman alphabet and some representing societies outside Western Europe and North America. I believe diversity of sampling is important else all too easily, findings may be skewed by biases arising from data from just one wiki. Of significance is that the single wiki showing absolutely no content transclusion (Cebuano) is also notable for being created almost entirely by bots. Otherwise, the occurrence of articles with transclusional mark-up (i.e. either calling or supplying content) by percentage of articles was less than 1%. Different wikis showed differences in the manner of mark-up use but not such as to suggest significantly different practice. Indeed, the differences might be attributed to common practice by small groups of editors rather than a difference common to a wider base of editors.

Within wikis there was generally one topic area that predominated for transclusion use and these primary areas varied between wikis. Again the general scale of use was sufficiently small that it would be incorrect to make strong assertions about different styles of practice in different wikis. The fact one topic tended to use much more transclusion is most likely the halo effect of small groups of common practice - or even individuals.

Patterns of transclusion were investigated, as a means of assisting easier detection of transclusional structures and to suggest useful contexts for their application. As yet, transclusional use is generally too limited and haphazard to define such patterns: one exception was a simple hierarchical pattern for linking articles on about episodic media (e.g. TV shows). Surprisingly, there has been no effort to apply this pattern to a wider number of articles. Potentially this is a task where a bot might assist, even if only spotting locales wherein transclusion could then be applied by a human author or by offering pattern templates to assist editors in making such transclusional structures.

Analysis of the documentation and editor attitudes was only pursued in the English wiki. Here, the documentation was shown to be disorganised and not well curated, but this could be easily fixed (were resource applied). Editor attitude was hard to evidence but where found it shaded to the negative on transclusion, most likely reflecting that the technique is, at first encounter, a confusing factor for those who want close control of articles of interest. Changing attitudes may be difficult as long as edit count status is a pre-eminent measure of editor value to the system.

Separately from the documentation, no clear evidence was found of general encouragement of content transclusion, by way of pointers to explanatory documentation, examples and community leadership (this finding is within the English language wiki). I have mapped hitherto unreconciled localisations of media wiki namespace-related terminology (q.v. Appendices L & M) which will further aid investigations of structural aspects of other mediawiki-based wikis. Despite the structured content in Wikidata that is available to Wikipedia there was scant evidence of such data re-use or attempts at semantic enhancement of content to make it more accessible to non-human users.

Results notwithstanding, the forms of transclusion and attendant mark-up detected in Wikipedia have been recorded in detail allowing this investigation to be replicated in Wikipedia beyond the ten wikis investigated, or in wikis running on mediawiki software within other domains.

8.1.2 Bot work in support of hypertext structure

My investigation of bots has traced and identified all affirmed and inferred bots within the English Wikipedia up to 30 June 2017 (q.v. Chapter 5). This was necessary precursor work to enable the main investigation as to bots' work. That this initial work was needed is significant in revealing how poorly curated bot information is, the lack of descriptive consistency about bots, and an absence of resolved reports on bots' activities and actions. Though Wikipedia's bot control documentation is revealed as a poor, despite this there is evident good work done by bots, their operators and the Bot Approvals Group (BAG).

Using this identification of bots accounts I was able to document the range of roles they undertake. This reveals further systemic weakness in the documentation and management of bots with inconsistent descriptions of tasks making difficult the verification of bot activities beyond the blunt tool of studying edit records[1]. I have shown the absence of evidence to suggest bots are being deliberately tasked, either directly by operators or on behalf of other users, with an intent to improve the clarity and resilience of the overall hypertext. That finding does not imply bots are not making useful contributions to Wikipedia, but speaks to the manner in which they are employed. It is not clear that bots are working at the service of the general user community so much as working at the service of their (community-minded) operators, a small but significant difference.

A surprising absence from the range of tasks given to bots was using a bot to track and report on bots, their number, status, and activities. Whilst it is fair to assume (there being no independent measure) that the volunteer BAG members have current bot operations under oversight, ongoing operational detail is neither faithfully recorded for later study nor clearly described. This has revealed precisely the fragility of undocumented process *knowledge* within a large hypertext, and instead reliance on a presumption of sufficient 'eyeballs' on the task at any time. Whilst it

[1] Edits are not always given informative edit comments, further adding to the task.

may work in a volunteer-based commons of Wikipedia's size, is not a model commended for other hypertexts to follow.

8.1.3 Additional Findings

As part of the main findings described above the research itself produced some further, unexpected, but pertinent observations. The first of these is the degree of broken hypertext 'behind the curtain', i.e. away from the public content articles. This has been significant and thought-provoking finding given our increasing use of digital/hypermedia repositories to house our knowledge that previously lived on paper, and maintained by librarians.

Wikipedia's service to the Web, involving fast growth in an evolving technical environment and with an almost entirely volunteer workforce, has been and remains impressive. Yet it is hard to ignore the mounting and unserviced technical debt arising as an unintended by-product of over-focus on the delivery boundary and the edit-kudos economy that offers little reward for maintaining the underlying parts of the hypertext. That is, the hypertext as discrete from the underlying technical substrate of servers, storage and networks, etc.

Attempting to trace links through areas such as talk pages—in many ways the 'minutes' of the hypertext's function—there is found to be a significant degree of link breakage. It is likely archival choices were for reasons of expediency rather than with regard to the long-term record; this has been hard to evidence due to lack of a clear record of intent. Such hypertextual breakage is not self-healing and harder to fix as time passes and the contemporaneous human memories that understand the cause of such breakage begin to fade.

Though Wikipedia's user roles (q.v. Appendix D) appear to function adequately, it can be seen that they are both technocratic in conception and do little to enable and incentivise those who need to work on the structure of the organisation's docuverse. Thus, even if using a Wikipedia model, (non-voluntary) organisations should give due regard to reviewing and altering the existing role structure the better to sustain their hypertext.

Human territoriality is under-researched but does appear to have a role in how well a docuverse with multiple editors function. Given the preceding point about making space for informational maintenance roles (as discrete from IT maintenance), making trust structures that acknowledge and reduce territorial behaviour will become increasingly important. Whereas most users are acting as contributors of domain expertise or editors thereof, much of the work is localised to specific parts of the overall hypertext. It is the hypertextual maintainers who have a different organisational task, their perspective being orthogonal to silos of per-domain contribution or content responsibility. Such an un-aligned view is bound to create tensions as unseen territorial boundaries are crossed.

I found little evidence of semantic enhancement being used in Wikipedia. Though there is data available to allow transclusion from Wikidata, although it does occur (q.v. Sections 4.3.4 and

4.6.1.6). As has also been seen, some editors seem averse to re-use of any text not written by themselves, or editors they trust (that trust also being an undefined measure); this is often cloaked in terms of provenance but appears more closely linked to territoriality and the edit-count-based kudos ecosystem.

In the round, this thesis' contribution has been to paint an unvarnished picture of a large functioning hypertextual system containing knowledge. The value to general society, and researchers in particular, is to create an source upon and against which to base future work on hypertextually-stored knowledge. This is the *tabula rasa* alluded to in my introduction.

Each doctoral thesis is but a stepping stone for others to build upon: without this step, I contend that study of hypertextual structures in organisational contexts would be significantly harder due to a lack general reference. Indeed, I have undertaken this study for that lack of just such a reference. I believe I have shown here that whilst our technology may keep our knowledge stored safely, it is far from clear that what is stored therein is safe from loss of coherence as links break and different parts of the contained knowledge evolve at different paces and in different directions. That weakness reflects a failure to recognise the task of maintaining a hypertext's structure as distinct from its underlying technical substrate or the factual content within the individual nodes.

8.2 Recommendations for Further Study

I believe my choice of a publicly accessible source as the subject of my thesis' investigation is important so that later research, especially if into large but less public systems, has a common point of reference. Past research into large hypertexts has generally been into their function (technical/quantitative) or behaviours (humanities/qualitative), rather that the content of the system as a whole. It is perhaps unsurprising the results are not as clear-cut as we would all hope. But, to make progress in this area it is necessary to survey the status quo, as in the work undertaken here.

With evidence at hand about a large hypertext's knowledge from the perspective of the content, rather than its creation, it should now be possible to make more meaningful research into the editors' perceptions of their work and its coverage of the problem space via methods such as interviews and surveys. That would allow better assessment of the alignment of users' intentions and motivations as compared to the needs of the hypertext. Such interviews should also allow issues of territoriality to be explored and whether this affects how editors perceive best use of transclusion mark-up. Indeed, direct questioning may give greater clarity to whether editors can conceptualise the hypertext as a whole or simply address individual issues as they become clear, such as by running watchlists only for certain pages; it may be that in large collaborative systems training needs to be directed to addressing this.

An ambiguity evident during my research is that of the degree to which an editor's edit count forms a significant part of their standing within the community. This is important as it plays

into the problem of generating sufficient consensus discussion and who decides if consensus has been reached. It is also important to obtain greater clarity on the degree to which edit counts are necessary or useful in sustaining user engagement in a collaborative hypertext, and how well such mechanisms work as incentives the hypertext's internal maintenance tasks. Edit-count kudos may be a necessary in Wikipedia's all-volunteer context but less divisive systems need to be explored and developed for collaborative hypertexts within organisations. One's standing within a system would be more useful if it reflected appropriate application of expertise rather than sheer volume of activity, not least because not all tasks occur to the same degree or pervasiveness.

There is a need for research to understand better the link triage necessary during archival and events like user name changes. Whereas in a public encyclopædia the current version of the main namespace articles are the version of record, organisations will have more nuanced needs and therefore we need to study how our hypertextual systems better embrace these via non-onerous methods and make such metadata accessible in semantic Web forms.

Despite the technical advances made with the Semantic Web, its features are still not widely seen integrated into large hypertexts. The degree to which this is due to the seeming complexity for the non-programmer of the likes of Resource Description Framework (RDF) is an aspect of use that warrants study> More and better Semantic Web integration will aid our hypertexts as they grow and also enable us to map better our intent for input to Natural Language Processing (NLP) and Machine Learning (ML) processes being run on these knowledge sources. For organisations with unavoidable responsibilities, e.g. adherence to laws and regulations, better embedded semantic structuring ought to enable better tracking of the implication for an organisations docuverse of changes in these externally-applied constraints.

Research is recommended into the lack of use of readily available semantic data resource, e.g. from Wikidata or through content transclusion (q.v. Section 4.6.1.6, especially the latter part). This would be research clarify if he lack of (re-)use is due to any or all of, a lack of understanding of semantically curated data, the means of its use, or an antipathy towards its use (e.g. through territoriality).

Lastly, there is also a research need to understand the visualisation of hypertext maintenance tasks. For instance, when working on the corpus as a whole and splitting/merging articles the link triage that arises can be complex, easily resulting in lost or broken links. Visualisation of link paths between the nodes may also help identify missing link pathways, or improvements to pathways between areas of the hypertext. In this regard an overlooked areas of research is spatial hypertext (SH) (q.v. Section B.3).

8.3 General Recommendations

Despite the limited use of content transclusion observed, the results presented show it is possible to find mark-up indicators of transclusional activity. This provides a useful marker against

which to undertake more detailed search for localised communities of (good) practice, allowing identification and encouragement of editors contributing to the health of the hypertext.

The hypertextual damage caused by poorly considered archival processes has also been shown. Increasing amounts of broken links within the hypertext that such slipshod activity creates increasing unpaid technical debt. Archival, along with reports of operational of activities, are important to those who need to maintain the hypertext and should be appropriately resourced. Reports that are not stored or cannot be run against anything but current data further degrade the longitudinal coherence of the hypertext, a non-trivial degradation to any such resource subject to legislative or regulatory oversight and archival requirements.

To be able to use structural techniques like transclusion, clear documentation is needed and proactive training may be required to seed uptake of new practices. Documentation is not nugatory work as besides informing those new to the subject, if well-written the process should help provide greater clarity of understanding to existing users and expose overlooked weaknesses in the system. An example of the latter might be a lack of simple tools to effect structural processes or methods to apply visual tell-backs in articles where these methods are used.

Where bot activities are supported, there needs to be clear reporting accessible to both humans and other bots. This reporting needs to explain what a bot is authorised to do, the scope of the authority (all or only part of the corpus, etc.) and the broad logic driving the task. As 'Artificial Intelligence (AI)' is used more widely, the latter becomes important when trying to understand bots' mistakes and biases.

Allied to the recommendation above re archival, better consider should be made to ensure process changes and archival do not routinely break links. When mended within that process or via a separate process, the aim should be to minimise the number of internal broken links. A task for hypertext maintenance staff should be to advise technical coding staff (who may have little understanding of the hypertext medium) on process changes with predictable destructive outcomes. It may be that such staff may net to vet process changes against such unforced errors.

The disciplinary boundaries between Computer Science and Library/Archival Studies warrant study in the light of the new hypertextual medium. Whether the emergent new skills required are absorbed into noe or both of these disciplinary areas or forms a subject of its own, resourcing for the task is long overdue.

8.4 Final Thoughts

I write this surrounded by books, a life-long source of both education and entertainment. Books and printed material will be around for some while yet, but print technology casts a long shadow: despite the speed at which much of our work and knowledge is moving into digital-only form, paper and print influences still feature heavily in digital-native designs. But, such design is not our future.

We must avoid embracing the familiar too tightly: the comfort of McLuhan's 'Rear-view mirror thinking' will serve us ill. As hypertext pioneer Ted Nelson has noted [341, 00:30-00:35] "Power corrupts and obsolete power corrupts obsoletely"[2]. Though his comment dates from 1980s and is about PCs rather than the digital docuverse, its tenor is as pertinent today in relation to our transition from the print to the digital era.

An abiding memory from the period of study for this thesis is that those most insistent that "there is no problem" were those who proved to have little–or more often no—real experience of creating a hypertext. Reading Web pages *cannot*, alone, reasonably be accepted as giving an understanding of the structure of hypertext. An irony of the post Web-age is that likely fewer of us *write* hypertext as opposed to 'create' it by inputting data into a database. Linking has, and *should* have, intentional meaning otherwise we—with current AI—have little better than linking on probabilistic best guess; those who lived through the late 1990s period when webpages were spammed with keyword-based created by auto-link-generating Web-bots will well understand how poorly computers understand the complexity of human-written text; even now NLP is little further forward: human language is unexpectedly complex. Linking, in the *hypertextual* concept has never been that ill-considered. We are some way yet from letting an algorithm usefully decide what links (relates) to '*this*'. Yet we, for reasons of habit or zeitgeist(?), choose to privilege primitive algorithms over well-developed human thought. Being able to stand up a web server is *not*, of itself, an indication of the ability to understand the complexity of the information that then may be stored upon it. Understanding of hypertextual structure needs some understanding of the underlying technology but less so than a good grasp of human reasoning and thought processes. The failure of the Semantic Web to make the advances assumed at its advent speak to a degree of technical solipsism.

The lack of hypertextually-skilled people is a pressing problem for society, especially as we live though a soon-to-be-realised informational dark age; at least until effective, pervasive, information systems are available. Ill-applied presumed expertise serves us all ill. Knowledge is not a zero-sum game; if we choose think less of others who lack our expertise whist discounting theirs, we are collectively the poorer for it. Thus a challenge for our collective society is to embrace the new medium in ways that free us from the hold of print-era concepts that are no longer constructive and to think openly about making best use of our newer media's affordances. The docuverse's ongoing health will benefit from some new skills. It may not be Nelson's exact imagining of a Hypercorps (Section 7.3), but there is a skills gap within the docuverse that we would do well to resource soon, by defining the skills of the hypertext worker and then selecting, nurturing and training those skills. The problems are not new. Otlet's idea of 'bibliology' (q.v. Section 7.1.1) is almost a century old and we have yet to understand the roles needed in the era of mass communication and replication.

[2] Nelson is understood to have originally coined the phrase in an article for *New Media* Magazine in the late 1980s in response to the increasing dominance of the Microsoft PC, and he has used it since as it the referenced video in 1990. The exact original source is not known.

We need the likes of a Hypercorps, or bibliologists. We can and must do better, therein lies a better collective future.

FIGURE 8.1: A Call To Arms
(© Mariana De Oliveira [226], reproduced with permission)

Appendix A

Xanadu Manifesto - 1966

Below is part the text of Ted Nelson's 'Xanadu Manifesto' [261, pp.7–8] from December 1966 in which Ted Nelson laid out his thinking about the capabilities of a hypertext system. The overall document was originally written as a proposed system for use by the New York publishing Harcourt Brace[1].

The document was referenced, and displayed [8], by Terry Gross during the Brown University's celebration of 50 years of hypertext research where he showed how Ted Nelson's early prediction of much of modern hypertext design.

Note: the document's text is reproduced here with the author's permission[2]. Emphasis is as per the original and few literal errors have been corrected. The original was type-written and the underlining was done manually[3]. Source underlining is not vertically displaced by descenders as seen when rendered in LaTeX.

A.1 The Manifesto

From the point of view of the system's actual design, however, these particular subjects and facts are unimportant. The parts that have to be designed into the system are more like the following.

1. Text editing. Contained texts may be modified with such instructions as "insert", "delete", "move", "copy".

2. User-defined file structure. The user may define his own information types and links, and the way the machine will move and handle them.

[1] Acquired in 2007 by Houghton Mifflin, and now subsumed in the (2010) renamed Houghton Mifflin Harcourt [177] company.

[2] Personal correspondence with Nelson, 24 July 2019

[3] Manual typewriters had no italic or bold method. Underlining was the normal manner of showing emphasis.

3. User-defined workspaces and file disciplines. The user may define such abstract working arrangements as push-down stacks, temporary holding spaces, "folders", indexes, and windows into his file. He may even specify overall file disciplines that the system will thereafter hold to him.

4. User-defined screen formats. The user may "carve" the screen into separate areas where his files and workspaces will appear.

5. User-defined branching actions. The user may arrange interchange between various presentations by drawing branch graphs, and specifying the conditions under which the machine is to branch; these may include throttle, pushbuttons or screen pointing. This branch-defining technique may be used either for creating work-systems or for creating hypertexts.

6. User-specified presentation modes. The user may have text segments or streams presented as frames, drum rolls, and streamers; and specifically which are to blink, how brightly, and when.

7. Continual Modifiability. The system a person is using my be modified ad lib.

8. Reversible processes. Both text and modification commands may be stored in such a way that changes can be undone as far back as records are kept.

9. Evolutionary text generation. Alternative versions may be spun off, and linked to each other so that their corresponding parts may be compared.

10. Complex Indexing. The system will have a complex indexing ability for either straight text complexes, permitting the creation of any number of indexes for a given body of information, from which the user may immediately jump to the information being indexed. This indexing facility permits the use of lists or texts to index one another, so that the user may proceed to a corresponding part in the other text or listing. This may be regarded as "backward" and "forwards" indexing, but that implies a priority which may not necessarily exist in the material.

11. Alternative drafts. The system will also have the ability to store alternative drafts of either straight text or text complexes.

12. Search. There will also be a word search facility, with which the suer may look for the occurrence of specific words and phrases, tally them, index them, process them, etc.

13. Complex Connective Structures. Such connective facilities as super-footnoting, the ability to attach text chunks to an ordinary text and have them available under many different categories; parallel texts, where ordinary texts are linked side by side so that they may be compared or their relations noted; array formats, where other types of information may be stored in matrices and graph structures; and a

cross-cutting facility, enabling the user to jump from a given part of the text to any linked part of the text.

14. Arrange and Printout. The system will be able to sort and arrange the material it contains, and produce printed reports on demand, even interactively with a user. This will make possible catalogues concordances, quick indexes and annotated manuscripts with some swiftness.

Appendix B

Early Hypertext Systems And Concepts

This Appendix forms extended notes to section 2.3 on early hypertext systems within the main literature review (Chapter 2). There are three discrete sections in this Appendix. The initial section looks at research into hypertext pre-Web, looking at the research and the systems arising from it. This includes techniques and systems building upon that work and which has been allowed to lie fallow since the advent of the Web.

The second section looks at narrative in hypertext and the link from the technical perspective of hypertext to Electronic Literature (ELit) and post-modern literary forms. Lastly, Ii will cover the often overlooked subject of spatial hypertext. Here, a note's position within the hypertext in 2D (or 3D) space is used to infer structure and interrelation of otherwise discrete notes.

B.1 Early Hypertext Systems

Recall that in the late 1960s, personal calculators were yet to arrive. Computing occurred on large and very costly central computers. Displays were rare and expensive. Programs would be written on punched cards and then would have to wait for a time slot to run on the computer. This form of 'time-sharing' was then improved so that multiple users could share concurrently, with the computer rapidly shifting its focus between users. Personal computing was only beginning to arrive in volume in the 1980s. The screens that are the most visible part of most people's computers today were rare and costly items in that period and networks were in their infancy.

If early hypertext seems limited in scope, that is to overlook the significant challenges of the day. Indeed, it makes the achievements all the more impressive. Although some organisations created a series of systems over time, most noticeably at Brown University in Rhode Island, I will describe the systems by their initial data (insofar as they can be fixed). This allows for seeing the addition of new features within the hypertext arena as time passed.

Some of the dates that follow for systems are inferred. Very few canonical papers about early hypertext systems give an accurate timeline for the system in question and a number had a gestation period of some years between project inception and a workable system being shown or published. Likewise some of the monographs about key systems use a discursive narrative style and often omit key dates. As a generality, the exact dates are unimportant but the sequence of systems does have some import in understanding the genesis of some hypertextual features we see today.

B.1.1 1967: HES, PROMIS

'**Hypertext Editing System (HES)**'. Brown University might seem an odd place as a cradle of hypertext, but here it was that HES was created. Andries ('Andy') van Dam, from Brown, had met Ted Nelson at the 1967 Spring Joint Computer Conference (SJCC)[1] and had decided to run a demo project to allow them both to explore Nelson's ideas about hypertext and also to investigate text editing for print documentation which at the time was a task done with typewriters. The result was the first hypertext system running on a commercial computer[2] and which could be used by computer novices. Although Nelson had seen Engelbart's oNLine System (NLS) in 1966, seemingly no explicit mention of it was made during the HES project so the degree of influence, if any, is unknown[3].

HES might not have happened at all as some senior staff at Brown viewed computers as too expensive and scare a resource to be used by mere undergraduates, as van Dam planned to do. As an unofficial project with no real funding the program only ran for a period during 1968. Some two dozen pages of Nelson' handwritten 'Hypertext Implementation Notes' survive[4] [262]. Nelson also claims HES had the first implementation of a back button [271, pp.94–95] [272, illustration, p.173], a feature now so commonplace as to pass without notice. It was also unusual for the period in that undergraduates were involved in the research and construction of the program.

Further detail of HES is given in Carmody *et al.*'s 'A Hypertext Editing System for the /360' [77]. The divergent goals of the two project leads eventually resulted in a frustrated Nelson leaving the project. Ever forthright, he reflected in a 2006 letter to the *New Scientist* magazine:

> "That project dumbed down hypertext to one-way, embedded, non-overlapping links. Its broken and deficient model of hypertext became by turns the structure of NoteCards and HyperCard programs, the World Wide Web, and XML."
>
> ([271, p.94])

[1] They also both went to Swarthmore, not quite overlapping, and so knew of one another.

[2] An IBM System/360 Model 50.

[3] Nelson states [30, p.98] he never mentioned NLS "…as its hierarchical system was unacceptable, and thus irrelevant.…".

[4] Nelson's comments to Barnet [30, p.100] imply these notes relate to HES.

Nelson's view was that a hypertext system should not be built around the limitations of paper (size) and printing. That parting of ways should not be read as a zero-sum contest. There were good reasons for van Dam's push for better documentation and printing. Nelson's dissenting side of the story is told in [30, Ch.5], [272, pp.172–174], [271, pp.93–95]. It is unfortunate the computers of the time lacked the capability to have supported more of both the planned outcomes. HES may have been a short unofficial 'demo' but the seeds of interest in hypertext had been sown[5] and in 2018 Brown was celebrating 50 years of involvement in hypertext.

'**PRoblem-Oriented Medical Information System (PROMIS)**'. At the University of Vermont in Burlington, a system called PROMIS was coming into service. Although the work of a research lab, this was not a research tool but a patient records system [308]. At a time when display screens were rare and something of a novelty unusually it used touch screens for display and data entry [308, pp.446–448][6]. PROMIS was based on information 'frames' (a large medical database) and patient record 'tables'. The frame or table filled the display and navigation was from screen to screen [7] [308, pp.450–453]. This node-per-screen approach was to influence the builders of the 'ZOG'. Though PROMIS was not conceived as a hypertext system, its inherent design has a hypertextual aspect. While HES introduced computerised hypertext to the academic student community, PROMIS was perhaps the first hypertextual system deployed in a public setting.

B.1.2 1968: FRESS

'**File Retrieval and Editing System (FRESS)**'[8] was Brown's second hypertext system and a follow-on from HES. Unlike HES, FRESS was funded (q.v.) allowing for more thorough design. By the end of 1968, van Dam had been to the Fall Joint Computer Conference (FJCC) and met Doug Englebart .

Rather as with Nelson's reference to other proto-hypertext systems, DeRose and van Dam's 'Document Structure and Markup in the FRESS Hypertext System' acknowledges that in this early stage of hypertext there was cross-over with other emergent software designs:

> "Many systems that have contributed to wide acceptance of hypertext were not conceived of as such. For example, Apple's HyperCard originally made no claim to being hypertext yet may have brought an idea of hypertext to more people than any other. Halasz notes that "Neither [GNU Emacs nor HyperCard] is a true hypermedia

[5] Not known at the time, IBM added HES code to its Type Three Library, with the result that HES was installed at National Aeronautical and Space Administration (NASA) who used to produce documentation for the Apollo space programme [386, pp.217–218].

[6] This choice was a research-based response to lessen the need for users to (mis)type inputs with potentially severe medical effects.

[7] The latter might seen odd viewed from the twenty-first century but like display monitors, the method of how to display things on a monitor was also work-in-progress.

[8] FRESS is also a play on words. The Yiddish '*fresn*' [W30] means to devour or gobble and the FRESS system was resource hungry [7].

> system (although both include some hypermedia features)"[9]. Similarly, "Guide did not start as a 'hypertext' project"[10]. Thus, they lack some of the conceptual foundations we consider necessary for fully coherent hypertext." ([104, p.8])

An interesting aspect of van Dam's approach to hypertext was that FRESS was used by arts students on a poetry course, a cross-pollination of disciplines less seen since; there is surviving contemporary footage [183] of this course[11]. The system is the first known use of an 'undo' feature [30, p.108], another feature now common to today's user of a web browser. Whereas the 'back' button reverses across a link, the 'undo' rolls back to a previous saved state of the system as a whole. FRESS was also the first system to use two-way links [104], i.e. a link between two lexia that may be traversed in either direction, whereas in earlier systems this would require two discrete links of opposite polarity. Ironically it was the a lack of these that was one of Nelson's disappointments about HES. FRESS continued to be used at Brown right up until the computers upon which it ran became obsolete and unmaintainable, a reminder that newer and faster is not always best.

FRESS also implemented stretchtext, "through dynamically updated conditional text" [104, p.14]

B.1.3 1972: ZOG

'**ZOG**'[12] originated at Carnegie-Mellon[302]. After early generally unsuccessful experiments [277], ZOG's developers came across the PROMIS system [302, p.461] and used it as a design paradigm for their interface, including use of a touch screen. ZOG used a series of full screen 'frames', with navigation moving the user from frame to frame, with a dataset being termed a 'ZOGNet'. The system was deliberately experimental, investigating how big a ZOGNet might practically become and the necessary response time for continued user engagement. Robertson's team reported several issues: "Users readily get lost in using ZOG", "Users fail to read information on frames", and "The limited short term memory of the user is everywhere evident in ZOG" [302, p.483]. The first of these, being 'lost', was common to many systems of the era, not least because this sort of interaction was very new. The latter issues are interesting as they flag up other erstwhile issues: screen resolutions were low and so information had to be placed across several frames (screens) thus also adding the user's confusion.

Nonetheless ZOG was deployed and validated in a number of work settings in information and process analysis roles [302, p.482] as well as being deployed in the aircraft carrier USS *Carl*

[9] [154, p.851].
[10] [66, p.33].
[11] From a 1976 documentary, no long before the system was decommissioned.
[12] ZOG is not an acronym: "which stands for nothing, but is short, easily pronounced and easily remembered" [303, p.1]

Vinson during 1983–84 after several years of shoreside training[13]. The initial deployment had some 20,000 data frames. As before this research group were atypical in clearly stating things that did not work alongside more positive findings. The findings report [13] from the deployment list has two development issues that stand out as still often overlooked:

> "R7: Insist that the developers use the system being developed.
>
> R13: Give documentors a central role in the project." ([13, p.2])

The first echoes Engelbart's approach (even if not intentionally), that the makers need to understand the practical differences between their intent and their outcome. The last is a point often overlooked, to the detriment of most documentation. The maker of something is often not the best to articulate it's function and means of use. In a parallel report, the *Vinson* experiment was further summarised [240, pp. 304–308]. It is interesting to note the cited limitations of existing hard-drives and network data rates (9.6kbps.). ZOG's experience was already contributing to new KMS system (see below) [240, p.309].

B.1.4 1978: Aspen Movie Map

The '**Aspen Movie Map**' was perhaps the first real piece of hypermedia [17]. Produced by Lippman and colleagues at the MIT Media Lab it allowed the user to take a simulated drive through the town of Aspen in Colorado, USA. This was achieved using videodisks that held all the town's streets in video form. In addition, some buildings had their interiors filmed allowing the user to enter the buildings and move around. Time-shifting was also possible as the town had been filmed in all seasons allowing a user to choose a different season to the current one.

B.1.5 1982: Guide, IGD

'**Guide**' began as a research project at Canterbury University, UK. It was initially aimed at finding a better way to read documentation and then developed further to add hypertextual features. Its design principle included a link traversal style that followed that of stretchtext (q.v. [65, 66]). After initial research and development, it caught the attention of Office Workstations Limited (OWL) who agreed to distribute it as a commercial offering. It was not until 1986 that Guide debuted on the Macintosh and it is generally held to be the first commercially available hypertext programme. Guide was soon also released for PCs.

'**Interactive Graphical Documents (IGD)**' (also called 'Electronic Document System (EDS)') was the next system created by van Dam's team at Brown. Part of its aim was the production

[13] The ship was launched in 1980 but did not join the United States Navy (USN) until 1982 and only made her maiden voyage in 1983. Thus the project had been embedded in the ship's activities from its very start and had years of experience before the first sea-going use.

of graphical documents (noting the heritage of HES and FRESS being used for documentation). The system is described and illustrated in van Dam's keynote for the Hypertext '87 [96], well as in Brown University's Hypertext@50 lectures in 2019 [6]. A step forward in the technology was the use of control of icons on screen rather than typing and the contextual elision of data to avoid visual clutter.

In 'Authoring Large Hypermedia Documents with IGD' [121] Feiner states: "The nodes in IGD's directed graph structure were called *pages*, modelled in part after the screen-sized pages of PROMIS, but emphasizing graphics, rather than text.", an interesting acknowledgement of prior work.

B.1.6 1983: Hyperties, SemNet, TextNet, KMS

'**Hyperties**' was a project by Ben Shneiderman and his team at the University of Maryland. It was first described in 'User Interface Design for the Hyperties Electronic Encyclopedia' [315]. Hyperties was originally text based with a screen either showing all or part of a lexia or an index screen but it was extended to enable touch-screen maps and mouse support. The program was also used to produce a (partial) set of proceedings for the ACM Hypertext '87 conference proceedings, featuring a subset of the papers presented. Hyperties also pioneered a move from referring to numbered links anchors, accessed by typing the number, to used 'anchor' text to indicate the presence of a link from that text: explained by Shneiderman in a 1999 video [324].

Notably, in 1989 Hyperties used to create the first software & book combination for *Hypertext on Hypertext* [316] which was produced in three forms, each including within the book the Hyperties software for a different operating system: Unix, PC(MS-DOS)[14] and Mac versions. Trans-media publishing of this sort was something of a novelty at the time but quite normal as viewed from a twenty-first century perspective. As the software contained the book in hypertext form, *Hypertext on Hypertext* is also held by some to be the first ebook, or 'electronic book'. In fact Hyperties was also used, a year earlier, by the ACM to publish part of the *Communications of the ACM* (or 'CACM') report on the first (1987) ACM Hypertext conference, which would seem to be prior art (and there are other contenders). Either way, Nielsen reports that:

> "For example, they have already sold 4,000 copies of their Hypertext Hands-On! book with its Hyperties diskettes. Also, ACM has sold 1,800 copies of the Hyperties version of Hypertext on Hypertext." (Hypertext'89 Trip Report [184])

'**SemNet**' [123]. This knowledge construction tool was published 1983–87. It offered a 3D-style presentation of hypertext linkages and is noted here as being regularly cited by contemporary papers.

[14] For the PC version, the software was provided on two 5.25-inch floppies.

‘**TextNet**’ [356]. A system produced by Randy Trigg as part of this PhD (the first in the field of hypertext). It was a precursor to the better-known Notecards (see below).

‘**KMS**’ by Akscyn [14] and his team. This was a follow on from the earlier ZOG system. It used the same presentational layout of screen-sized cards, the content essentially being tailored ot the display. That might seem egregious until you consider the lengths gone to by today’s website developers to represent the same source content for different screen form factors. Like ZOG, KMS used transclusion to bring linked information to screen.

B.1.7 1984: Intermedia, NoteCards

‘**Intermedia**’ was created by Brown University’s Institute for Research in Information and Scholarship (IRIS) group, which grew out from Brown’s ‘Scholar’s Workstation Project’ [247, p.186]. A development here was that linkages could occur *between* discrete documents, albeit within Intermedia system. This starts to extend the reach of a hypertext beyond earlier systems where the links were essentially internal within a single hypertext. Moreover, the system allowed multiple users to share the same docuverse. Odd from today’s perspective, but an important move forward towards collaborative/multi-user systems. Interestingly, in 1989 Meyerowitz is describing that only five years previously copy/cut/paste were a novelty and these are now available to any user [246, pp.112–113].

Rather than new content replacing the existing view, Intermedia allowed for multiple (document) windows within the application, with links between them. Link and anchor information was stored in a discrete database, outside the documents [151, p.43]. An innovative system, Intermedia suffered from running only on the Macintosh’s Unix implementation which limited scope for deployment.

‘**NoteCards**’. This program was produced at Xerox (Xerox) Palo Alto Research Center (PARC) and was a multi-user hypertext system that acknowledges similarities to Neptune and Intermedia. Although supporting multiple users it did not, in original form, support live collaboration. The system was built around ‘notecards’ and typed links and was described in [154, 157]. The latter had shown that users struggled with correct use of typed links[15]. At the time, hypertext search was still a nascent feature, meaning users were very dependent on the author’s links. This latter issue, of the problem of search and query, became Halasz’s first ‘issue’ (see section 2.4).

B.1.8 1985, DE

‘**Document Examiner**’ and ‘**Concordia**’. Document Examiner was intended as end-user interface for commercial hypertext documents, as part of Symbolics Genera a software development

[15] Bernstein recalls Trigg reporting NoteCards users found choosing the correct link type simply too hard and so tended to use no link typing at all, but this cannot be further sourced. Author’s email correspondence with Mark Bernstein, Eastgate Systems Inc. (Sep 2015).

environment for Symbolics computers. Concordia was the editing program and Document Examiner the program for reading documents [366, 338]. There is still video of the system being demonstrated [338]. The editor was built upon a derivative of the Unix 'emacs' text editor. The proposed benefits over paper were [366, p.313]:

1. Full indexing
2. Quick following for cross-references
3. Back referencing
4. History

A notable feature introduced by the system was bookmarks, again obvious after the fact to those brought up in the Web era. Consider that paper books may have indexes, though potentially expensive if done well. Meanwhile, the reader can also add ad hoc indexing mark by means of bookmark(s), most commonly used to show current progress through a book. Today, indexing can be automated given digital text, but Machine Learning (ML)'s ability to understand text semantics means human intervention is needed to monitor quality. However, supporting user-generated ad hoc bookmarks is far less complex as the user provides an unambiguous point of access to store and retrieve.

B.1.9 1986: Neptune/HAM

'**Neptune**' and '**Hypertext Abstract Machine (HAM)**' [103]. Delisle and Schwartz's system offered a hypertext system for use with Computer-Aided Design (CAD) applications, which is an interesting divergence from most hypertext programs of the time. HAM forming the hypertext underpinnings: "When we speak of Neptune, we are generally referring to the functionality provided by the HAM". They also cite Weyer's 1982 dynamic book [374] as inspiration.

B.1.10 1987: Hypercard, Storyspace, Thoth-II, WE, Hypergate, SuperBook

'**Hypercard**'. Goodman *et al.* [144]. Opinions are divided as to whether Hypercard was a real hypertext system but there is no doubting its presence at the time, not least as Apple choose to bundle it with their Macintosh computers. As the name implies the system was card based, by careful design users could make quite sophisticated applications. Working with fixed card sizes, the contents of one card would replace another, in a manner akin to Guide, where all content was in full screen pages.

The application was designed by Bill Atkinson. In suitably West Coast fashion the idea came to him whilst sitting out under the stars, having taken LSD, and wondering about interconnectedness: 'Triangulation 247: Bill Atkinson (part 2)' [25]

‘**Storyspace**’. At University of North Carolina (UNC) Chapel Hill, Bolter and Joyce had been collaborating since 1983 on a variety of non-linear narrative writing tools (GLOSSA, TALE-TELLER[16]), before settling on Storyspace as a name in c.1985. The original version of Storyspace [56, 42] was demonstrated at Association for Computing Machinery (ACM) Hypertext 1987 Conference, including an early version of ‘afternoon: a story’ [191] It offers another early example of the collaboration between the arts (Joyce) and computing (Bolter). Of the systems described in this section Storyspace is the only one still in general use and still supporting its original content[17]. Unusually for early hypertext systems, Storyspace was developed expressly with creative writing in mind, in contrast with other systems which were funded to look at documentation, Information Retrieval (IR), Knowledge Management (KM), and defence applications. However, as with HES and FRESS, the insights gained from writing non-linear narrative composition fed back into ‘non-fiction’ hypertext tools.

Eastgate took over the development of Storyspace in December 1990 [30, p.134] and it maintains the program to this day. In 2016 Eastgate created Storyspace v3[49] using a common code-base with the recently re-coded Tinderbox which should assist the longevity of the program. This is important as much of the canon of hypertext literature was written using Storyspace.

‘**Thoth-II**’ [85]. Billed as ‘Hypertext with Explicit Semantics’, it is interesting for its browsing mode. The system had two modes: reading and browsing. The latter uses a ‘Thoth Spiders’ display [85, p.278, fig.2]:

> “The design of Thoth Spiders is different. The space is a two dimensional plane rather than a three dimensional space. Location in the space is uninterpreted. Often in browsers such as SemNet, there is an attempt to map spatial location into some conceptual metric. For instance, the y axis might be used to represent abstractness. The most abstract objects would be plotted high up in the space while the more concrete sink to the bottom. In Spiders where a node is located within the display is purely a function of the users interaction with the system. Secondly, users interact with the structure that they view, creating the graphic objects as they browse. These properties flow from two basic decisions:
>
> “—The location of a graphic object has no particular meaning in terms of the abstract structure that is being displayed.
>
> “—There can be multiple graphic objects all representing the same abstract object.
>
> “This different use of the display space allows the drawing of the network to “flow out” over the 2D plane. In any system in which the nodes are multiply connected the placement of one node depends on the place of the nodes to which it is connected. By breaking these connections display is made inordinately easier. The display

[16] As described by Barnet [30, ch.6, pp.125–128].

[17] In part, this is because so many of the important early creative hypertext works were composed in Storyspace and need the programs’ player version to be ‘read’.

> algorithm breaks *cycles* by treating every reference to a node as unique. This potentially creates multiple *graphic* representations of the *same* underlying node in the network. The interface model is: the user views a two dimensional plane on which part of the larger graph structure is drawn." ([85, p.279])

This results in a display somewhat akin to a force-directed graph layout. Another interesting design aspect is that text (as viewed in a text display) is not simply the content of a 'card' as in some earlier systems but the text derived from a node (as of the type seen in browsing mode). This extra layer of abstraction is interesting as it begins to break away from the constraints of print/paper-informed design and was intended to facilitate reasoning/computation across the link graph.

'**Writing Experience (WE)**' [322] or 'A Writing Experience for Professionals' from UNC, looked at non-prose writing[18]. This was a tool intended for writing papers and documentation: "What is needed are tools designed specifically for the sophisticated professionals who use workstations within distributed environments." [322, p.727]. WE was not intended for screen reading, but instead assembled desired nodes in sequence to print out a linear document. The system used four modes "graph, hierarchy, content, and search—which are "held together" by a control panel" [322, p.731]. The modes had different roles:

> "The default layout for the screen shows five tiled windows (see Figure 2). The two largest are a graph window and a hierarchical window. The first supports operations that conform to the rules of a directed graph embedded in a Euclidean space. The second obeys the rules of hierarchies. A smaller window is available for either a text or a graphic editor used to write or draw the content of the document, associated as blocks of data with individual nodes. The fourth window is used to search the relational database for other structures or nodes that might be inserted into the current document. The last window is a control panel for managing the environment." ([322, p.729, fig.2])

The WE system shows that experimentation into the uses of a hypertextual environment were still ongoing and there was no single paradigm of use.

'**Hypergate**' [39] was another fledgling system demonstrated at the 1987 Hypertext conference. Eastgate were experimenting with the user experience of the reader and is notable for the first use of a 'breadcrumb'. Here, the breadcrumb gives an indication as to whether an available link takes the user back to a page previously visited [39, pp.42–43], as derived from a stored list (history) of visited pages. By exposing only the most recent entries, the session history thus becomes a list

[18] Barnet [30, p.127] notes that Bolter, who created 'Storyspace' was also an author on the WE paper and that the latter may had some design influence on the former, despite their explicitly different purposes.

of recent visits, to help resolve the thought "Where did I just visit? Have I been here before?". The breadcrumb concept later turned up in the Mosaic web browser[19] [30, p.133].

'**SuperBook**' [301] Not formally a hypertext, but rather a text browser was presented at ACM Hypertext '87 where it caught the eye of Halasz who referenced it in his 1991 conference keynote [155] for its use of search. SuperBook described itself, in part, thus:

> "*Different approaches to the same problem*. SuperBook attempts to accomplish the same ends as hypertext systems like ZOG and NoteCards by sophisticated key-word techniques applied to the full contents of a text, obtaining added intra-text relationships through annotation and adaptive aliasing, and by text navigation aids."
>
> (Discussion section (italics as per source) [301, p.187])

B.1.11 1988: Microcosm, Trellis, gIBIS, Hyper-G

'**Microcosm**' was created at Southampton University and grew out of investigations as to use of multimedia when the Mountbatten Archive was passed to the University. As well as the app using basic text documents, Microcosm could open a range of documents including Microsoft Office. Small plugins were developed for a number of applications, letting the user create and edit links in situ in the current app, including putting links onto video.

Microcosm [126] is interesting in that as well as having two-way links, which were stored in a discrete 'linkbase', it also supported running multiple link bases and the linkbase manager which lets the user set the priority of the various linkbases found with a dataset. The links being stored externally, changing the stack order determined which of the various linkbases would supply links to the user. Although the term linkbase has more recently defined as an XML

Microcosm overlapped the start of the Web and is an interesting vision of what might have been. From a research project, the program was spun out into a company, Microcosm Limited[20]. Additions were made to connect the app into the Web [79] but having been designed as a single user tool there proved to be no long term market.

'**Trellis**' [329]. Stotts and Furuta's system used Petri nets. Trellis separated content from structure and used the Petri net to control how the "association of content with the structure defines the context in which the content is presented to the reader" [129, p.2015].

[19] There, it took the form of using a different colour for visited link. Bernstein, writing in 2009, notes: "…breadcrumbs warn us when a new link is about to lead us back to a familiar page *not* because we don't want to waste time revisiting something we have seen before, but rather because the most important information in the link, for us, is that it leads to an issue we have already discussed." [50, p.11]. This bouleversement of original behaviour shows both how hypertextual navigation is now less of a novelty and how not all early hypertext notions translated naturally to the Web.

[20] Subsequently, this became Active Navigation Limited [11]. The company survives but not the Microcosm product.

'**Narrabases**'. Malloy created narrative databases, drawing on previous experience of exploratory stories using 3x5-inch paper card catalogues. *Uncle Roger* [223] (1988), was an online 'narrabase' using 3 discrete database files for different parts of the story.

'**gIBIS**' [89]. Conklin's system sought to investigate design deliberations, decisions and rationales. The name stood for 'graphical IBIS' and as basses on the Issue Based Information Systems (IBIS) system of issue mapping notation. This reflected a contemporary interest in Artificial Intelligence (AI).

'**Hyper-G**' (later 'HyperWave') [23, 237]. Developed at Graz University from 1989 onwards, Hyper-G's novelty stemmed from a shared/replicated link database. The original work started around the same time as the Web and offered a larger feature set. Hyper-G offered two-way links stored in a link base, page annotation (via its 'Harmony' browser), access control and multi-language support. The program was commercialised as 'Hyperwave' in the mid-1990s but though elegant in concept, the Web offered a cheaper, if less featured, alternative and like Microcosm it failed to compete at scale.

Pam's 'A comparison of WWW and Hyper-G' [285] from the early days of Hyper-G is interesting in putting Xanadu in the mix of competing hypertext systems.

Similarly, the company [185] formed to sell Hyperwave still exists survives without mention of its origin. Jaske's article [376], writing in 2000, provides a useful summary of some of the differences between Hyperwave and the Web.

B.1.12 1989: Scripted Documents, IDE

'**Scripted Documents**' [397]. This experimented with the notions of guided tours and paths as an attempt to reduce cognitive load and disorientation for the user. As well as browsing, sequential, branching, conditional paths could be offered.

'**Integrated Development Environment (IDE)**' [190], developed at Xerox PARC is interesting for its use of various features to help generation of structure in a hypertext. The idea was to aid the user by automatically building out sets of notes and auto-linking them, such as when rapidly building equipment manuals. 'Tailorability' was one of the program's aims—this being one of Halasz's '7 Issues': "IDE extends the tailorability of NoteCards by allowing the user to (1) create, modify or delete types defining cards and structures, and (2) modify the menu interface of the hypertext environment to suit his or her needs.".

B.1.13 1990: Dexter Hypertext Reference Model, Lace

'**Dexter Hypertext Reference Model**'. Presented in the 1990 Proceedings of the Second Hypertext Standardization Workshop held by National Institute of Standards and Technology (NIST),

Halasz and Schwartz's model [153] represented the outcome of two workshops[21] in late 1988. The meeting drew on the experience of the proceeding years and the problems of individual systems with no inter-compatibility, noting that this was some way in advance of modern networking and interoperable PCs and operating systems. The Dexter workshops drew people working on number of extant hypertext systems, including: Augment (NLS), Concordia/Document Examiner, IGD, FRESS, Intermedia, HyperCard, Hyperties, ZOG/KMS, Neptune/HAM,NoteCards, the Sun Link Service [286], and Textnet [356].

The Dexter model, in essence, proposed a generic 3 layer model intended to promote better interchange between hypertext system. It describes itself thus:

> "The goal of the model is to provide a principled basis for comparing systems as well as for developing interchange and interoperability standards. The model is divided into three layers. The storage layer describes the network of nodes and links that is the essence of hypertext. The runtime layer describes mechanisms supporting the user's interaction with the hypertext. The within-component layer covers the content and structures within hypertext nodes."
>
> (From paper's abstract, p.95 [153, pp.95–133])

The model gained further visibility when it was presented in the Communications of The ACM (CACM) magazine in February 1994 [158]. By this time however, the Web was already growing fast and research money and interest was beginning to pivot away from hypertext to the Web. It was not entirely ignored and several more publications followed up on the Dexter model: Grønbæk & Trigg 'Toward a Dexter-based model for open hypermedia' [149] (1994), De Bra *et al.* [98] (1994), Grønbæk & Trigg 'Toward a Dexter-Based Model for Open Hypermedia' [150] (1996), De Bra *et al.* [98] 'AHAM' (1999), Millard *et al.* 'FOHM' [250] (2000), and De Bra *et al.* 'AHA! Meets AHAM [99] (2002) De Bra *et al.* 'AHA' [100] (2003). De Bra's group were loking at the model in the context to Adaptive Hypertext, whilst Millard's group were taking an Open Hypertext Model (OHM) perspective. Most recently, Atzenbeck, Roßner & Tzagarakis' 'Mother' [26] (2018) is using a Dexter-type 3-layer model in a spatial hypertext system.

'**Lace**'. [293, 294] Another Southampton development, Lace aimed to combine the hypertext notion with structured documents. This choice reflected the simplicity of Web documents at the time. The system had three primary components. LaTeX formed the document creation environment, NeWS gave a windowed display system, and a document library server. Lace considered a document as a structured collection of nodes, showing some similarity with structured documents like Standard Generalised Markup Language (SGML).

[21] The name of the model comes from the fact the meetings were held at the Dexter Inn in New Hampshire.

B.2 Hypertext and Narrative

Writers had experimented with non-linear narrative before computers(see Section 2.2.1) but today's ELit needed the advent of the personal computer before the filed began to burgeon. In part that must reflect the fact that prior to the PC, using a computer meant access to a large, expensive machine on which a person was (time-)sharing a resource with other users. The PC brought computing to the more intimate surrounds of the author and offered them the complete service of that machine.

Barnet's *Memory Machines* [30, Ch.6] documents the early days of Bolter and Joyce's 'Storyspace' program [42], the software used for writing and publishing many of the early public hypertexts. It must be remembered that this was the mid–to-late 1980s and though an author could easily write sections of discrete text, until this point there was no tool to link those text together to provide narrative(s); again, this as half a decade before the arrival of the Web. Storyspace is the only surviving system from hypertext's pre-history, having recently debuted only its third version in 30+ years [49]. It remains important as besides still being an authoring tool, it is the only means to run some early ELit stories as these cannot simply be transliterated as Web hypertexts.

Bolter's *Writing Space* [55] seeks to give context to hypertext in contrast to existing print media and discussed literary theory for the (then) new medium. Landow's Victorian Web [211] is a long-lived hypertext resource that has journeyed to the Web, via Storyspace, having started life in Intermedia [390, Ch.4–5]. It is pertinent as one of the long-term projects[22] informing Landow's writing (above). Its survival is testament to hypertext's flexibility although it looks remarkably dated in design terms. Whilst design is no measure of the content's relevance, it can affect the impression of newer readers and those accessing the resource via newer devices such as a phone with a small screen and haptic interface. This is a reminder that the fettling of a hypertext resource needs to consider context of content as well as its structure.

Bernstein and Greco's *Reading Hypertext* [50] is also primarily concerned with literary hypertext but mixes published literary papers with more general content. Murray's *Hamlet on the Holodeck* [255] (revised/updated 2017 [256]), much referenced by other authors in the area, gives a look into the future of narrative, albeit now from the past; a future yet to be realised in part due to the divergence of hypertext study and the emerging presence of the Web. Whilst Hypermedia is often thought of in terms of audio and video, McCloud's *Understanding Comics* [238] gives a useful reflection on hypertext narrative, abstracted and viewed through the medium of the modern comic. McCloud uses the medium of the graphic novel (or 'comic') to show how the textual and graphical media can complement each other in a hypertextual fashion.

Waldrip-Fruin's compendium of essays and papers, *The New Media Reader* [368], helpfully juxtaposes the scientific and cultural aspects of the nascent new media that has essentially become

[22] Also see Landow's websites 'Contemporary, Postcolonial, & Postimperial Literature in English' [210] and 'Cyberspace, Hypertext, & Critical Theory'' [209].

hypertext and the Web. Landow's *Hypertext 3.0* [212], is the third edition[23] of his writing on the nature of hypertext, critical theory and narrative. For purposes of understanding hypertext much of the latter is abstruse, though it helps contextualise the thinking around the birth of some early hypertext systems and their literary goals. Much of the technical description is now dated (even for 2006). Worthy of note is his chapter discussing "The Politics of Hypertext: Who Controls the Text?" [212, Ch.8]. Focus dwells on political power and hierarchical authority. Oddly, from a socio-technical perspective, it overlooks the privileged role of the hypertext author instead focusing on editorial control. This shows the difficulty of simply mapping old publishing concepts onto new media.

Despite the title, Marshall's *Reading and Writing the Electronic Book* [229] actually has a wider scope, being the distilled experience of a researcher with the Aquanet/VIKI/Visual Knowledge Builder (VKB) spatial hypertext systems and subsequently as a digital ethnographer. Currently, ebooks are generally enclosed hypertexts thus they are not totally divorced from the wider Web experience. More recently ebooks have moved to allow web-based interactive content, further blurring the traditional notion of a 'book'[24]. Marshall explores how the experience of reading hypertext—including annotation—has an impact on how a (hyper)text is consumed. Marshall also contributes to *Reading Hypertext*, see 'Reading Spatial Hypertext' [50, pp.211–220] drawing on her combined experience of spatial hypertext and digital ethnography.

Although the focus of ACM Hypertext has drifted somewhat towards social networks[25], nonetheless there still continues to be discussion of hypertext fiction and narrative. Bernstein's 'On Hypertext Narrative' [45] from 2009 summarises a number of contemporary hypertext fiction considerations, including stretchtext. Hargood et al. 2009 [166] describe 'strange' [331] and 'fractal' hypertexts, the latter echoing in part the notion of stretchtext albeit in a larger context. Kolb [204] juxtaposes similar and inter-related content in book and Web form, to investigate the differences of the experience[26]. Extending from the 'Card Shark' narrative idea comes the notion of Sculptural Hypertexts [51] [251] the formalism of which has since been incorporated in Storyspace (v3), aiding exploration of that type of hypertext.

Hargood *et al.*'s 2016 paper [167] continues the theme of considering patterns in hypertext, this time in the context of their effect on using sculptural hypertext to create viable location-based hypertext narratives. The work follows on from a 2013 paper by Millard *et al.* [251] looking at use of the sculptural hypertext in location-based narrative. As yet under-explored is what if any of these narrative considerations can be applied to better structuring non-fiction knowledge in large hypertexts and their possible implications outside fictional narrative. Bush's idea of sharing trails [71, section 7] remains pertinent; the hypertext author, or maintainer, may consider looking

[23] Previous editions were published in 1992 (*Hypertext*), 1997 (*Hypertext 2.0*) prior to the current 2006 edition.

[24] q.v. Battle and Bernius, on ebook transquotation [33].

[25] It was suggested, in discussion around Hypertext 2016 that this may be because some groups of hypermedia authors/artists are antipathetic to the academic setting of the Hypermedia conference.

[26] This builds on his 'Sprawling Places' [202] and Story/Story [203] dual-context experiments.

for or explicitly creating link paths through a hypertext and then making those easily found at the edge of the hypertext (where readers may enter it).

An encouraging activity in the Hypertext Conference of 2016 was a creative track, which helped (re-)broaden perspectives as to the scope of hypertext. The return of a creative hypertext element in ACM 2019 is a welcome move. Such aspects of hypertext are a useful bulwark against an otherwise natural drift to a more tech-centric view given the digital environment of hypertext. Millard and Hargood's 2017 'Tiree Tales' [249] explored the issues surrounding the creation of locative hypertexts, an interesting blend of pure narrative with a real-world spatial element. Locative hypertext was explored further by Hargood and others in 2018 [164, 165].

Kitromili, Jordan and Millard [198] ask 'What is Hypertext Authoring?' Written with regard to (creative) hypertext narrative, my own experience of writing 'non-fictional' hypertext (e,g, documentation) would suggest the points raised speak to a wider degree of hypertext use.

Does this seeming niche matter to the hypertext medium in the broader context? Yes, it does. If for no other it reason, offers a bridge between the humanities and the technical disciplines. They are equally inward looking in the compass of their view but, a more holistic view is needed in the post print age; there is no merit in a binary judgment as to which disciplinary view better reflects the medium. Rather, drawing on both should give insights that can be used more widely in hypermedia, and outside contexts that are literary in intent.

B.3 Spatial Hypertext

In the early 1990s the world of hypertext (hypermedia) forked, the manner of which was not immediately significant. Tim Berners-Lee's new World Wide Web [35] offered a comparatively cheap and easy method to interlink hitherto discrete document collections and to do so without the quite complicated formalism then being developed within the hypertext community [158]. Hypertexts at the time were all essentially 'discrete systems' [78] whose linkage and search were intra-system and whose capabilities were at the behest of the system designer. New formalisms were envisaged to tie these disparate systems, but the simplicity and low implementation cost of the Web rapidly outflanked these initiatives in the initial instance.

In the current Web paradigm, links are one-way and embedded in the Web 'document', at least at point of render in a browser. This (technology-limited) approach precluded adopting the rich systems of typed, non-directional links stored in external link tables that were being developed in the mainstream hypertext research community. For instance, Trigg's PhD thesis [353, pp.2–3] had identified over 80 [sic] discrete types of link that hypertexts might use.

In 1988, Halasz, gave his 'Seven Issues' paper [154] that arose from his experience with NoteCards [157]. That had shown that users struggled with correct use of typed links[27]. At the time, hypertext search was still a nascent feature, meaning users were over-dependent on the author's links. This latter issue, of the problem of search and query, became Halasz's first 'issue'.

Ironically, World Wide Web was proposed not long after Halasz presented a follow-up 'Seven Issues: Reviewed' in 1991, included escaping 'the tyranny of the link' [155, p13]. The latter issue was a compression of three previous topics (search and query, composites and virtual structures) from his 1986 paper [157]. His point was to elevate use of the, then novel, search-derived links to be on a par with traditional deliberate (typed) links created by the hypertext's author. In the light of today's Web this seems unexceptional but was not at the time. Even if unintentionally, Berners-Lee's World Wide Web upended the status quo as regarded hypertext linking (mainly due to what it could not support due to technical limitations of the time).

During the preceding period, hypertext theory had invested as much value in the type and placement of links as in the texts (or other media) they connected. In part this stems from ideas having run ahead of the capabilities of existing systems: the usability of complex linkage methodologies was not easily self-evident in the hypertext systems of the time. Hypertext systems designers also perceived that hypertexts might encapsulate system design [89] or legal argumentation [354] and infer from such data, deriving meaning from the manner and type of linkage.

The new Web-based hypertext was consumed via a 'web browser' and with advances in computer power the browsers have grown in speed and capability. 'Search engines' not only broke the 'tyranny of the link' but started to tie together initially disparate island of content on the Web[28]. The original vendor specific feature differences have now generally given way to acceptance and adoption of open standards. Thus web browsers have a rich and fairly established Document Object Model (DOM) [383] and element selectors exposed via Cascading Style Sheets (CSS) and JavaScript offer the opportunity to reinsert some link typing, even if links are still one-way and not held in a discrete link table as in older hypertext systems. Whilst the early simplicity of Web features also initially removed the affordances for inference, numerous incremental improvements now allow for the Semantic Web. The latter is most often talked of in context of being a bridge to AI, but there is no reason to ignore use of deliberate authorial semantics being used to facilitate bridging to knowledge areas as yet unfathomable to AI.

Improvements in search engines mean the pages of curated links typical of the early Web are now rarely to be seen. Though servers may restrict access to some data, the documents of the Web—in terms of all the servers connecting via the HTTP protocol—are essentially one large hypertext; Marshall notes the 'growing heterogeneity of hypermedia genres, uses, and users' [228, pp.97]. The Web thus changes, if not subverts, Nelson's original vision for hypertext/media, creating a shift in meaning described by Wardrip-Fruin [369]:

[27] Bernstein recalls Trigg reporting NoteCards users found choosing the correct link type simply too hard and so tended to use no link typing at all, but this cannot be further sourced. Author's email correspondence with Mark Bernstein, Eastgate Systems Inc. (Sep 2015).

[28] Early on the Web, even finding content took some effort as there was no central directory.

> ‘those working in the literary community must reconsider hypertext definitions focused on the link, and those working in hypertext research must reconsider definitions that privilege knowledge work over media.’

If the world’s eye was turning to developments on the Web in the early 1990s, the hypertext community continued onwards with more hypertextual system focus and with aspects of narrative. Marshall et al.’s 1991 paper [231] from PARC presented Aquanet, which acknowledged a lineage from prior systems: Halasz’s collaborative NoteCards [157, 154], and Conklin’s gIBIS [89][29]. Aquanet hoped to help visualise the structuring of knowledge. Though not mentioned at the time this has a nod toward Polanyi’s concept of ‘Tacit Knowledge’ [288].

NoteCards (late ’80s) improved on predecessors like NLS/Augment. Besides the data notes (cards) it included a graph-style ‘browser’ view, giving an organisational overview of the interrelationships of the cards in the system. In reflecting on Aquanet, Marshall [231, p57–58] raised the notion of informal sorts and delayed linking during data creation. This might be seen as a different assault on the ‘tyranny of the link’ even if not thus described.

The ideas emerging in Aquanet were taken further by Marshall & Shipman at Texas A&M with the VIKI system [234, 233]. In the same period Shipman and Marshall had published ‘Formality considered harmful’ [313], a light re-working of their PARC tech note of the same title from 1993. This looked at tacit knowledge, emergent structure and the effects of premature formalisation. Indeed, a benison of spatial hypertext is that of avoiding the inherent structuring implied by the act of linking. This idea can be alien and hard to grasp for those trained from outset on preformalised building of systems and accreting structure on strictly linear or hierarchical form. Spatial hypertext allows for the sort of loose association that human minds do quite naturally.

Another program from this time looking at visualisation of the hypertext graph was Eastgate’s ‘Web Squirrel’ [40] which references both VIKI and Storyspace and describes itself as ‘spatial hypertext system designed specifically to help people manage the hundreds URLs, email addresses, and other internet resources they see every day’.

The same team that created VIKI followed it with VKB [314]. VKB built on the spatial hypertext notions of VIKI, retaining its spatial parser whilst adding the ability to make implicit type assignments based on note positioning and offering a navigable history. Recognising the need to interact with other systems, VKB also supported internal and external linking. Sadly, none of these systems are still available in running form.

Eastgate Systems is a publisher of hypertext literature and hypertext authoring tools that had taken over the maintenance of Storyspace [42][30] in 1990. Combining the spatial hypertext ideas

[29] gIBIS sought to investigate design deliberations, decisions and rationales.

[30] A necessary task for the hypertext fiction community as Storyspace was both the authoring and reading environment for most of the canons of early hypertext fiction. Even in v3, successfully loading ‘*afternoon, a story*’ remains a core unit test for new Storyspace builds. (Mark Bernstein, in conversation at ‘Hypertext 2106’ conference, Halifax, Canada. July 2016).

emerging from Shipman and Marshall's work with experience with maintaining Storyspace resulted in Tinderbox, hypertext tool [44, 46].

Eastgate's Chief Technologist and lead software author Mark Bernstein, in his book *The Tinderbox Way* [47] gives further insight as to his design intent. Tinderbox recast Storyspace's 'writing spaces' as 'notes' that may be displayed though a variety of views including (spatial hypertext) maps, treemaps, timelines as well as outlines. Document data is stored as an XML outline (the in-app outline view) with RTF text in RTFD encoding, which also makes it available for direct access/transform as XML. Links can be made both within a document or out to the wider Web, as in VKB. Although links are directional, they are stored as a separate link table rather than encoded into note data (i.e. unlike a webpage). There is a strong prototyping system allowing for rapid addition, or re-working of, emergent structure and avoiding pressure for premature formalisation. As an original design purpose was blogging (pre Web 2.0) there are strong export mechanisms including transclusions[31].

Ensslin, in 'Canonizing Hypertext: Explorations and Constructions' [117], pushes back against Storyspace dominance [32] but this seems churlish. Hypertext authors had to use the tools available and Storyspace was effectively the only tool at the time—recall that Storyspace pre-dates the Web and that in the early days of the Web, doing anything was partly trial and error, so hardly a stable platform for all but the most adventurous writers.

In 2010, a Web-based system called Emberlight[33] [235] implemented a PHP-based Web interface for hypertext derived from a Tinderbox data file. That file was stored in Github for versioning purposes and the Web interface also had a means for limited annotation input that could be merged back into the underlying data file. An Emberlight site's owner could also use the file in Tinderbox; as the latter is Mac-only, Emberlight allowed the sharing of spatial hypertext data with a larger non-Mac audience[34]. Intellectual Property ownership issues meant Emberlight closed in 2011 but the experiment proved a concept for moving spatial hypertexts out onto the wider Web. Separately, a 2009 paper by Matias [236] discussed issues of comparing and collating spatial hypertexts (again using Tinderbox).

Moving to the present, web browsers have seen the adoption of the HTML5 standard and improved support for browser-based spaces such as Scalable Vector Graphics (SVG) and the `<canvas>` tag. These technologies are not a necessity for Web-based spatial hypertexts, but their availability the opportunity to consider use of spatial hypertexts within Web documents without building completely new display and rendering mechanisms. Those looking to use spatial hypertext as a workspace for triage and fettling of existing hypertext, may be less interested (or skilled) in

[31] Personal experiments, in 2006 and 2009, to export Tinderbox spatial hypertexts as HTML document were failures. Tinderbox has flexible export methods [46] and whilst the necessary data could be exported correctly, the drawing capabilities of web browsers at the time were simply not developed enough to give a workable solution (and features were inconsistent across major browsers).

[32] via [30, p.132, footnote 27].

[33] Be aware the domain has been re-sold and is now an unrelated outdoor lighting business.

[34] As disclosure, I should note that I was an early beta tester, pre-release, and helped Matias and Cheung with writing the system's documentation.

coding so leveraging current technologies offers a less complex route to the actual task of getting a spatial hypertext displayed in a browser window.

Narrative has a long association with hypertext, not least because early pioneers were interested in hypertext fiction as a form of immersive non-linear story-telling [30, ch.4-5]. This also lead to the notion of Sculptural Hypertext [51], since implemented in Storyspace v3 [49]. As the latter is now being built on a shared code base with Tinderbox, two niches are united: the surviving hypertext literary tool with the surviving spatial hypertext tool[35].

During early use of hypertext fiction tools, the method of laying out the notes (writing spaces) on maps led to reflection on patterns[36] in the linkage and layout of the overall texts, culminating in Bernstein's influential 1998 paper 'Patterns of Hypertext' [41]. These patterns were offered as visual manifestations of various narrative structures. That concept was extended to non-fictional work by Bernstein's 2011 paper 'Can We Talk About Spatial Hypertext?' [48] but not taken further by others. In part this may be because fictional works drafted in environments like Storyspace's maps were, in effect, drawn on a canvas. Non-fictional work is not generally produced in such a way—often being written in, or output as, a hierarchical outline form—though there is no reason a more graphical approach might not be used. Patterns in (non-fictional) Web hypertext, if such there are, are as yet un-described.

From Mitchell and McGee worked on HypeDyn 'a procedural hypertext fiction authoring tool'; it also offered some spatial hypertext views to assist with story design.. The tool is best described in their 2012 paper 'The HypeDyn Hypertext Fiction Authoring Tool' [252]. The tool is still available [181].

As VIKI/VKB were research experiments, knowledge of spatial hypertext has remained largely within the hypertext community. The strength of spatial hypertext as an information exploration space is best experienced rather than described which likely explains in part why research funds dried up[37].

Hsieh and Shipman's 'VITE: A Visual Interface Supporting the Direct Manipulation of Structured Data Using Two-way Mappings'[38] [180], from the same lab as VIKI and VKB offers an 'interface for people to manipulate information in their own way and at their own pace' which will be familiar to the few who have tried and worked in such spaces. - Another of the few spatial hypertexts project is HyperSea by Styliaras and Christodoulou [333, 334] which aims to integrate Web 2.0 content within a spatial hypertext setting so the spatial element is a post-hoc visual element rather than the heart of the system.

[35] Although Storyspace now no longer has a Windows OS version.

[36] This draws on the ideas first put forward in Alexander's architectural design book '*A Pattern Language*' [15].

[37] Bernstein suggests spatial hypertext simply is not likely to attract tenure or funding, so in the US at least it is not a promising line of research or new entrants. Eastgate is a private company so not subject to such pressure. Email correspondence with Mark Bernstein, Eastgate System Inc. (Oct 2015)

[38] VITE's name explained (glossary).

There were a few ongoing supporters of spatial hypertext such as Kolb [201] and Bernstein [48] keeping the subject active. Happily active interest in spatial has now been revived by Atzenbeck and collaborators at Hof.

Schedel and Atzenbeck's 2016 paper [307] on spatial parsing has fed into ongoing Tinderbox [43] and Storyspace [49] design. Atzenbeck *et al.* followed this with 'Revisiting Hypertext Infrastructure' [27] that presented the project Open Data INnovation (ODIN). Rewarding here is a return of a computational aspect to the spatial mapping, i.e. more than simply a visualisation of (versions of) the hypertext graph.

Atzenbeck, Roßner and Tzagarakis' 'Mother: An Integrated Approach to Hypertext Domains' [26] is a continuation of ODIN investigating at how the Dexter model [158] may be updated and to reflect a modern 'component-based open hypermedia system' (CB-OHS).

Another aspect of spatial hypertext is its use, as opposed to its design, i.e. the user's perspective and understanding of this medium. This is explored by a small number of papers. Marshall & Shipman's (1997) 'Spatial Hypertext and the Practice of Information Triage' [232] looks at the use of spatial hypertext to investigate and understand information, as discrete from using hypertext tools to write non-linear text. A hope at the time was that spatial hypertext could be used to explore argumentation and furthermore extract it as an output post the triage phase. The authors' VIKI system [233] is used as a point of reference. Neumüller's (2000) 'Mind the Eye: On the Relevance of Composition in Spatial Organization' [276] also reflects on VIKI as a use of spatial hypertext and how it plays with an author's sense of the composition space. Lastly, Marshall in 'Reading Spatial Hypertext' [230]—as previously described in Section B.2—explores the human experience of reading and absorbing spatial hypertexts.

Appendix C

Transclusion

This appendix expands on the mechanism of transclusion as proposed by Ted Nelson in 1965 (see Section 2.1) and expands on Section 2.9.

C.1 The Idea

In Nelson's concept transclusion achieves a number of benefits:

- (re-)use of a canonical source
- retention of provenance—in the case of Xanadu, via visible links

Xanadu remains unbuilt, and in an age when hypertext is constructed and used primarily using Web-type technology—even if for local use on a personal device—the above ideas are now encountered slightly differently.

Transclusion of content has blurred with software concepts like libraries and server-side includes whereby the viewed end result may be constructed on-demand from a number resources. Because the overall setting is within that of use of software, the latter tens to win out in people's understanding. Transclusion can seem an odd construct, yet it is remarkable in its simplicity.

C.2 The Practical Problem

A benison of transclusion for the hypertext writer maintainer is having a single canonical source for, and point of (re-)edit for, content that may then be re-used in multiple locations. Transclusion may be to a single URL, such as a discrete article, or as an inclusion directly within an article, or into multiple discrete URLs. Need for the latter occurs if the same content is more

usefully presented to the reader in multiple contexts in disparate parts of the hypertext, as in the hypothetical case above.

In a hypertext with fine resolution these problems are lessened as if links point more precisely to relevant content, splitting or moving existing nodes (pages)

C.2.1 An Example

As an example of abstraction of emergent pages, consider a feature (or process, concept, etc.) 'X'. It may initially have been written about solely as part of a description of topic 'A' within the overall hypertext. Subsequent changes in the overall subject matter can result in the feature X now being pertinent within topic 'B'. Initially, this can easily be serviced by adding a link from B to A (or more specifically to anchor within A where X is described). But, if the relevance of X spreads to a greater number of topics, there arises the need to abstract the information discrete to X from its current parent (A) into a discrete hypertext node of its own.

The new node describing X can now link to/from all relevant notes without appearing to simply be an aspect of A. Thus A, and B may interlink independently with X, and in due course new nodes 'C', 'D', etc., may also link to X as pertinent.

On the Web X may either be published as a discrete webpage but an alternative approach, if X is insufficient to warrant a public node (page) of its own, is to abstract X into a new node but only transclude it into A, B, etc. Current content management systems generally support such an approach as they already use inclusion of templates (transcluded content) for page building. Either way, ongoing coherence of the hypertext still involves clean abstraction of X from its original embedded context within A.

As well as reviewing linkage from the standpoint of access and cross-reference, i.e. finding the content, there may be issues of referencing and provenance. Whilst X (above example) was in-line within A, A may have contained outbound (link) references underpinning the relevance and provenance of the information now migrated to the separate page about X. These links also require triage as they need apportioning or duplicating across the re-factored content. This is a less obvious aspect of longitudinal hypertext maintenance. As important are inbound links relevant to X that actually simply link to A as these too must be triaged out form the links pertinent to A alone.

C.2.2 One Source, Two Instances—Or Two Versions?

An aspect where modern (Web-style) hypertext has diverged from Nelson's original ideas, in how the transclude functions is in what is transcluded from source to recipient. In Xanadu, each edit is a discrete version of a document as opposed to a 'diff'[1], or difference between versions,

[1] The term derives from the Unix 'diff' [W20] program which reports version differences between files.

an affordance not easily available when Nelson was defining Xanadu. Thus, when making the content shown in Figure C.1[2], any typo or addition to the source reference involved uploading a complete new copy of source document to the 'Xanaviewer' [388] demo server and then updating the Edit Decision List (EDL).

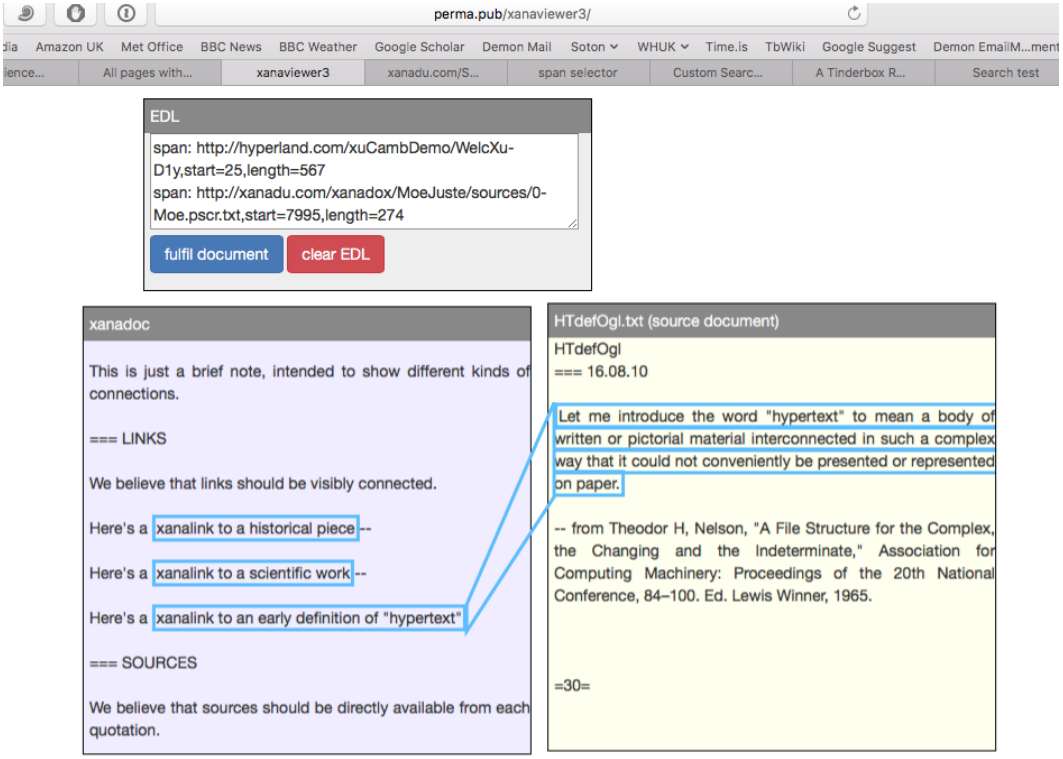

FIGURE C.1: Nelson's Xanadu 'xananviewer3' displaying transcluded text.

The difficulties of updating sources, would doubtless have improved had Xanadu advanced more, but it remains the case that Nelson's transclusion—even if never explicitly documented thus—draws on a particular *version* of source document. In such a scenario a change to the source has no effect on the transclude.

By contrast, the form of transclusion used on the Web is different as it does not, by normal convention, link to a particular version of a document but the current version as dereferenced via the stored URL. Here is how Wikipedia illustrates the process[3]:

[2] Note: this is the same image as shown in Figure 2.1

[3] From Wikipedia's 'Transclusion/How Transclusion Works' [W57]. This is a repeat of 4.1.

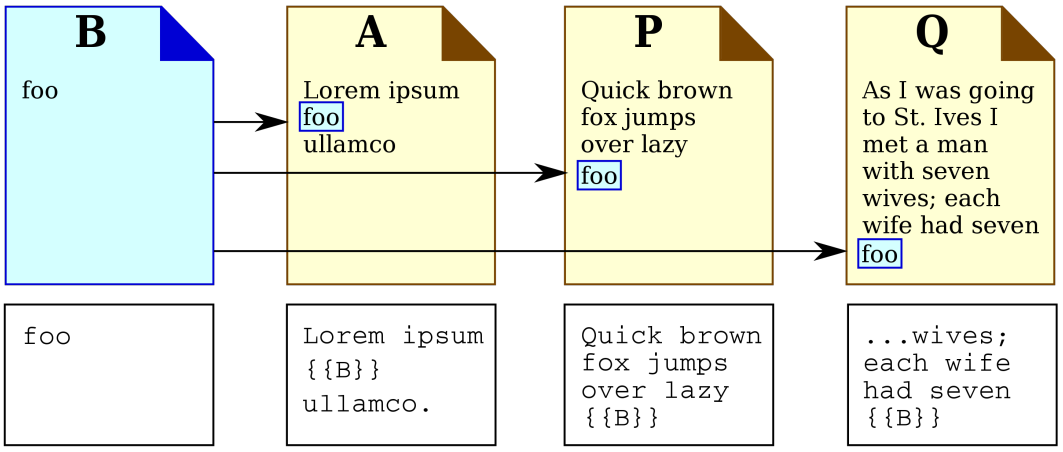

FIGURE C.2: Wikipedia's conceptualisation of transclusion (Image source [W25]).

Thus the transclusion *could* link to a particular resource, but generally does not. In most cases, the latter works well, as it means corrections can be applied by simply replacing the target resource. However, it does undermine Nelson's other aim of showing provenance. Without some other mechanism—even if only using a version-dependent URL—a Web-rule transclude/include does not maintain provenance as it only points to the current version of the target resource.

C.3 The Ongoing Coherence Challenge

Whether by direct authoring or via transclusion, the new nodes (or web pages) resulting from any re-factoring of content may themselves now be of sufficient significance to need adding to any tagging or indexing system. Consideration must then be given to the new node's visibility in the affordances offered to the visitor at the hypertext boundary.

Simply adding a new page (or transcluding content) and then adding/moving a few links is rarely sufficient maintenance action. This observation is based on personal experience of maintaining more than one continuously updated web hypertext for continuous periods exceeding 10 years. Such durations cover significant changes in the subject matter and the external contexts from which they are then accessed. It is especially important for changing hypertexts to pay regard to their readers. For anonymously-accessed informational Web hypertexts, as distinct from artistic works (where the reader is expected to service their own needs), maintenance of the hypertext has to allow for more than the replacement or editing of source text or other hypermedia assets[4].

Such hypertexts require continuous consideration as to whether the content is not only correct but both findable by the intended reader and presented in terms they will comprehend (terminology and taxonomy may drift over time). Those unused to indexing or search, will tend to underestimate this latter task. Search based on (key)word match, even with some machine learning support, still only tends to work well if the search term is distinct in terms of spelling and semantics.

[4] This is a point easily overlooked by authors writing essentially for themselves rather than others.

Transclusion is not always a direct benefit, if considering issues such as differing writing styles or use of tense. This is even more the case if content is to transclude inline, i.e. seamlessly within the body of other text. By comparison, if an article is a discrete webpage or node in a hypertext, then despite issues of writing style the problem of coherence within a broader linear text is removed.

When using hypertext to record knowledge over time about an organisation or process, this will lead to a need to split or merge nodes as well extract sub-sections as described above. Such work offers scope for link breakage unless all affected in and outbound links are properly apportioned or duplicated where necessary.

Appendix D

Roles Within Wikipedia

In keeping with its open and non-didactic nature, Wikipedia has a very limited hierarchy of membership 'user groups' [W699] which define groups with differing scope of editing and configuration control.

D.1 Local Groups

These are present in each discrete wiki within Wikipedia overall. They are listed below, in order of ascending degree of control:

- User – the ordinary editor (contributor)
- Administrator, also known as 'sysops'. The main group of content controllers within Wikipedia.
- Bureaucrat. Primarily, these give or take rights (separately) for 'Administrator' and 'Bot' rights. A Bureaucrat's rights are taken/given by the Stewards group at global Wikipedia level (discussed below in Section D.2 below).

Members of these groups can apply for, or be awarded, one or more of 74 different permissions[1]. Thus individual users may have powerful rights within niche[2] areas where they are adjudged to

[1] The table of 74 discrete permissions [W701] includes 19 relating to essentially niche functions: 17 (prefixed 'ep-') relate solely to the 'Education Program' [W644] and a further 2 relate to the 'Books' feature of Wikipedia [W441]. The education programme is designed to allow educators a pedagogic environment to organise engagement with and contribution to Wikipedia.

[2] Many of the sub-groups defined by these permissions have only single or double-digit memberships.

have expertise[3]. Administrators receive all but 9 of these 74 permissions as part of their office[4]. The result is a complex, non-mutually exclusive and fluid set of user permission groups.

The extra permission flags cover a range of issues and are essentially task-based. Some duplication occurs because the same generic task may be defined as a discrete permission for working on an article (page) and a file (images, sound, etc.). The broad areas covered are:

- creation: 8
- edits: 6
- remove-revert-redirect: 14
- user vetting: 22
- bots: 2
- rate limit: 3
- miscellaneous:
  - wikipedia book collections: 2
  - educational programme: 17

The 36 items representing the 'remove-revert-redirect' and 'user vetting' in total are 48.6% of the total and 66.4% of the non-miscellaneous permissions. These areas deal with observing and controlling (mis)conduct by editors. Vandalism and misbehaviour remain a concern for those running Wikipedia, though with an open system action has to be about detection and repair rather than protection, as might be possible in a closed system.

Within the User group, there is some further classification:

- **Unregistered**, or not logged-in (i.e. anonymous editing).
- **New**. Registered but yet to make noticeable contributions.
- **'Autoconfirmed'** and Confirmed. Currently, Autoconfirmed status implies a user account is both over 4 days old and has at least 10 edits. The Confirmed status is a bypass flag for this restriction and used in special cases.

[3] For instance, the 'file mover' permission requires that '*Users are expected to have at least 6 months of editing history and at least 3,000 edits*'. [W665].

[4] The 'User Access Levels' [W699] describes Administrators as having these permissions 'granted' but counts in statistics listing permissions groups [W153] indicate administrators are not included (i.e. they are smaller than Administrator membership count). It is unclear if Administrators are essentially pre-approved for these roles but do not necessarily take these up as list of per-permission users include some who are marked as Administrators: Special: Users filtered for 'Importers' [W160] lists only 2 'importers' (both of whom are administrators). Administrators also have access to 'Administrator tools' [W430].

- **Extended Confirmed**. Currently requires 30 days' tenure and at least 500 edits.

The differing levels of confirmation within the 'Users' group for use are partly a means to avoid excessive editing by very new users before they have had time to understand Wikipedia's style and the implications of edits. This is done by monitoring both how long a user account has been active and the number of edited the user has made.

The 'Autoconfirmed' level qualification is set very low, but only around 5%[5] of registered users[6] actually make the 10 edits or more to achieve this status. The 'Extended Confirmation' is even more selective, at less than 1%[7]of registered users. The 'Extended Confirmation' group are usually the only users other than sysops able to make live edits to the most controversial articles where the danger of edit wars[8] require additional protections.

Ordinary users wishing to attain Administrator status are required to make a 'Request for adminship' (RfA) [W668] and submit to public examination within Wikimedia and therefore must be in the good standing of their peers[9]. A similar process controls ascension to the Bureaucrat group.

As a discrete group of users, the 'bot' group [W580] merits close control due to bots' ability to effect unattended mass edits and thus have scope to cause serious damage if poorly configured or misused. Control of Bot permissions are retained at Bureaucrat level. Activities of Bot accounts are further controlled by the 'Bot Approvals Group' composed of a mix of Bureaucrats and Administrators (mainly the latter). Although bot activities are BAG-approved configuration of bot accounts is very inconsistent and in many cases accounts gives no clear statement as to their purpose[10].

D.2 Global Groups

The local user groups above (except 'Founder') are local to a given wiki within Wikipedia and are further subordinate to a set of global groups [W651]. The latter are pan-Wikipedia and refer to all wikis within Wikipedia. These groups comprise:

---

[5] The (unarchived) Wikipedia Statistics page of 7 September 2016 [W153] showed a registered user count of just over 29M users. The User Access Level [W699] page reports c.1.4M 'Autoconfirmed' users at 1 August 2016 (no source reference is given), making them 4.83% of registered users. There appears to be no regular report of Autoconfirmed user levels, unlike other user groupings.

[6] Wikipedia knows when someone first joins but generally cannot tell if someone 'leaves' Wikipedia other than by looking at their last edit or log-on. Active users—deemed to have performed an 'action' in the last 30 days—totalled 114,163 (0.39% of registered users at the time) [W153].

[7] Statistics [W153] lists these as 27,172.

[8] '*An edit war occurs when editors who disagree about the content of a page repeatedly override each other's contributions.*' from 'Wikipedia:Edit warring' [W642]

[9] '*This is not a vote but a deliberation in which registered users are invited to comment and support or oppose. In reality this can amount to a small poll of those most engaged with Wikipedia and others such a friends of those seeking elevation. RfAs are not without accusations of bloc voting occurring that subvert the premise of a free and open process*' Wikipedia: Requests for adminship, section 'Expressing opinions' [W669].

[10] It is to be presumed the bot operators either consider such information self-evident or no one else's business.

- Staff. For paid members of the Wikimedia Foundation.
- Stewards. Only Stewards can set/remove 'Bureaucrat' permission for a user within local wikis.
- Sysadmins. Allows sysop access to Wikipedia functions across all wikis.
- Ombudsmen. These for the ombudsman commission who investigate complaints to the Wikimedia Foundation about breaches of privacy policies.
- OTRS-member[11]. Members of the Volunteer Response Team (VRT) who assist with issues of copyright.
- Global bots.
- Global rollbackers.
- *Global sysops (group defined under mediawiki but not implemented in Wikipedia).*
- Global interface editors.
- *Founder*. This is a special role created solely for Wikipedia's founder Jimmy Wales and inherited as a local group by each discrete wiki.

Membership of these groups is extremely small. The Stewards control of Bureaucrat rights is the primary link with local (pre-wiki) users.

D.3 Reputation

This is not a formal role but a user's activities are a matter of record as all edits are recorded and reputation—good or bad—can accrue from that. Bad behaviour is subject to reversion of edits and ultimately being banned from editing (in extreme cases underpinned by IP-based blocking of access to Wikipedia).

More interesting is how positive editing behaviour is reflected. All users have a user page which includes a link to a list of their edits; these may also be searched for by username [W144]. Special pages record 'Featured' and 'Good' articles [W151]. Users may record and note significant contribution to such articles. There is no formal measure of the kudos arising from such works but it is clear in interactions on the wikis that significant contributors will gain standing amongst their peers. Such recognition may be quite local, in relation to the areas of content to which they make contributions.

[11] Member of OTRS

D.4 Content Maintenance

None of the 74 extra permission flags described above, in Section D.1, relate specifically to the longitudinal care of hypertext, although such a task would involve use of some of them. For instance, the 'move' permission is described as 'Change the title of a page by moving it'. This can both restrict ability for accidental or malicious re-titling or be a necessary task in content maintenance.

However, the lack of any explicit acknowledgement of the need to maintain the *hypertextual* nature of Wikipedia content is striking. The assumption drawn is that the ongoing health of the project is simply an emergent property of the aggregated individual editing of multiple users. WikiProjects [W715], which number over 2,400 [W717] in the 'en' wiki are informal aggregations of users around ad hoc areas of common interest[12], or by generic task, such as building missing pages for 'red' links[13]. In fact, the numbers are lower than that when looking at active projects by edit[14] [W638] or contributors[15] [W637]; some projects exist in name only.

The most active WikiProject, 'Military History', is notable for a large number of 'feature' and 'good' articles. Those Wikipedia accolades indicate articles adjudged to be of high standard and reflect the best of the encyclopædia but it should be noted the criteria for such articles makes no mention of the articles' context within the larger corpus. It is another indication that Wikipedia is essentially a set of independent (but interlinked articles); the degree of coverage and completeness is an emergent property.

Wikipedia does have a category system that may, at first sight, appear like some sort of an index or hierarchy of content. In fact, it is a voluntary folksonomy with seemingly light oversight. Using categories to navigate Wikipedia is no guarantee of finding content relevant to the task at hand; search is generally better, but not fool-proof if the searcher is unfamiliar with the article title, or keywords for the subject. It is interesting to note that studies on Wikipedia will sample using categories, as boundary condition, though this may be less precise than imagined. In better maintained areas, such as those with the purview of wiki projects, the categories may be more authoritative but this cannot be safely assumed across all subject areas.

D.5 Summary

Despite the complexity of the features described above, the control mechanisms experienced by most users of a collaborative hypertext in Wikipedia with be the local user groups. As the main role of the (few) Bureaucrats is to control the giving and taking of Administrator and Bot status,

[12] For instance, the 'Military History' project [W718].

[13] For instance, the 'Women in Red' project [W721] which aims to fix links to deleted to uncreated pages and in this case especially those relating to biographies of women.

[14] 631 projects have no recorded edits (via that project), 596 projects have over 100 edits and 89 have over 1,00 edits.

[15] Only 934 recorded projects have 30 watchers (interested editors). Only 291 of these have over 100 watchers.

there is thus essentially a two-level hierarchy of user groups: general editors (Users) and sysops (Administrator) albeit with an overlaying complex skein of niche roles.

This amounts to a technocracy where a user's position is largely dictated by their permissions as these control the actions they may undertake. Separately to this there is (unquantified) kudos attached to the degree and quality of edits made. In both cases the association is one of users to articles via edits. Effort is essentially vertically siloed, with a very flat technical control hierarchy atop it. There is no clear sign of deliberate pan-wiki consideration of content structure; this does not mean it may not exist, but that it is not made evident by Wikipedia's output or documentation.

Appendix E

Practical Suggestions For Sustaining Organisational Hypertexts

Below, I suggest a number of actions it is possible to take today. Contextualised by my experience from a secondment to the UK's Cabinet Office (CO) during my thesis studies, it also draws from my various previous careers before commencing my doctoral studies (q.v. Section 1.2).

I recommend the reader to consider them in the context of their own organisations. There is increasingly a blur between what has hitherto been regarded a 'Web content' and on organisation's business correspondence, and the whole may usefully be considered the organisation's docuverse, if only to avoid silo-ed data: that does not preclude limitations on general access to sensitive areas of data. The exact technical detail of execution may vary with the software in use with the organisation, but the broad strokes should remain the same:

- *Increase use of built-in document headings*. The underlying issue is less the style of the visual rendering but the implicit structure being added. Informal conversations with colleagues suggest few staff make regular and deliberate use of these unaware that, regardless of visual style, consistent styled headings give an important semantic alignment of text and structure that improves understanding of a document's structure by software processes. Also, for all but short text documents, use a Table of Contents. Not only does this give a user direct deep access into the document but it adds extra machine readable data. However, both benefits rest on using headings rather than ad hoc styling.
- *Make use of Templates*. Most major repeating activities should be assumed to need template support and provisioned for accordingly. Templates offer a double benison. Firstly, the template helps to create consistent semantic structure and style (as above). Secondly, the template offers a standardised method for addition of appropriate metadata with, or without, the user's active input (i.e. reducing potential user process error).
- *Encourage More Linking Between Documents*. I rarely encounter departmental documents containing links to other in-cloud documents and, from discussions, an indication of poor

awareness of this functionality amongst those in general office roles. This indicates a failure to register the change from a paper base to a digital one. As well as making it easier to locate the right document, links will help inform Machine Learning (ML) systems about connections and patterns it might otherwise not detect. It is too simple to blame office software training as this is normally conducted early in post before a new user really understands their tools. Linking should be part of the community's normal practice and those who persistently fail to do so should be targeted with remedial training regardless of position.

- *Link to canonical in-cloud sources rather than copy*. Although Google/Microsoft office tools do not explicitly advertise support for transclusions, it is possible to embed sections of original sources rather than recreate or copy them. The links generated will, again, help inform ML activities looking at document patterns.

- *Start to codify organisational process*. Place the canonical copy of documentation on the intra/internet and link to it; avoid local duplication. This both helps cement common consistent process and decreases the chance of documentation being allowed to drift out of date due to neglect. Control of intranet (intra-organisation) content should be removed from control by comms/PR/marketing staff and passed to personnel who understand hypertext (and so not necessarily defaulting to Information Technology (IT) staff).

- *Ensure proper integration of style guides and spelling/grammar support files*. Examples to avoid are style guides produced by graphical design staff using technique entirely for pre-digital era print use (no selectable text, no working links for URLs, etc.), spell-check dictionaries installed for the wrong language (e.g. American English vs. British English), not checking if software-installed spelling or grammar checking files contradict existing house style rules. Having achieved consistency, and a digital-first implementation, ensure this guidance is well promulgated to all staff.

- *Encourage Emails That Contain Mainly/Only Links*. How So? Consider the fate of marginalia and the problem of reconstructing the record (presuming there is no deliberate author's intent to obfuscate). What was once minuted in a file or on a document itself is now scattered through a myriad emails. A better discipline for the public record is for annotations to be added as metadata to the document in question.

- *Do Not Mistake Edit/Diff-based Storage for A **Meaningful** Archive*. Modern pervasive 'back-up' systems now auto-save, freeing software users of the need to deliberately save state in a document. The system may save all user-saved edits and on a timed basis in between. Yet it also conditions them to the presumption that a machine knows all and knows best (i.e. the user need not pay attention to versioning). Yet that same process cannot tell you if edit (saved state) X or edit Y is the important one for the record, merely that they are different and that, in extremis, the document can be rolled back—or restored—to a given edit state X or Y. But, which are the canonical versions of key documents (such as

policy) and who of the many working on them should decide when a document? For now that remains a human function, albeit poorly supported by the design of office software.

- *Include Versioning of Documents in Tool Training*. Early digital 'versioning' experience tended to involve making a new copy of a document. However, in a cloud based system that may not be necessary and the public record better served by the saved edit stream of a document holding successive versions with key edits noted in the metadata. Such behaviour is not common and uptake should allow for the need to train these new methods of work.

- *Use Scripts To Seed Metadata*. New documents should be generated with appropriate default filenames that include information, such as a YYYY-MM-DD date or project code. By supplying a default (stub) name there is more chance of a meaningful long term filename being given as most users will lack the interest to alter or replace the default. In significant contexts, where it is predicted documents will go to the national record, give consideration to some additional scripting to seed embedded metadata captured via an input dialog. This is especially true for repeating events like minutes of committee meetings or performance and progress reviews. It is easier to assess the Return On Investment (ROI) of work attempting to open a file if there is some idea what lies within it. Depending on the organisation's capability to apply customisation to their content creation systems, more precise seeding of metadata can, and should, be used.

- *Understand Your Docuverse*. Having a 'Knowledge and Information Management (KIM)' function does not necessarily yield an understanding of the organisation's knowledge, but may simply reflect high-level management accounting metadata about that knowledge. Whilst the latter activity is not without value it does not remove the need for the organisation to understand its knowledge, the structure of that knowledge, and how the knowledge was created and how it alters over time. Do not simply assume roles with 'knowledge' in the title are the real maintainers of the organisation's knowledge.

- *Monitor And Report on Support of Process*. The above interventions are for naught if their effect is not tracked and reported.

- *Managerial Representation*. Those implementing the above changes need agency at senior level: a manager must speak for the knowledge itself. Simply making the work an adjunct to the IT role will likely lead to stagnation of effort as the task is not core to that role.

- *Leadership by Example*. Senior leaders may need close coaching to improve their digital literacy and use of the above techniques. All levels of the team use the same docuverse so the suggestions above relate to all staff and junior staff will take their cue from their seniors as to appropriate use of and engagement with digital knowledge resources.

I would note that none of the above depends on more and newer technology (e.g. better search, faster/better ML and such) but rather upon better use of that technology which is already in use.

Appendix F

The Wikipedia Social Machine in Books

In considering the Wikipedia social machine for Chapter 6 it was useful to review some of the books pertaining to Wikipedia that I have gathered during my research. Note, I have discussed the origins of Wikipedia in section 2.8.

Although books-sellers lists are full of books keyworded to Wikipedia, few have any relation to Wikipedia beyond that word in their title. So I will confine this discussion to three categories, as below: I have already discussed books on the origins of Wikipedia in section 2.8.

F.1 Technology

I have already discussed Leuf & Cunningham's 'The Wiki Way' [215] in Section 2.8. From 2001, it was one of the first detailed implementation guides as to creating wikis. It is worth bearing in mind that outside academic or technical industries, high-speed access was not as prevalent as today and connection speeds meant distribution of large amounts of code on the enclosed CD still were of value to some readers. Today, such code would simply be indicated via a download URL.

It was from Cunningham's wiki code, freely shared, that the 'mediawiki' [W3] fork of wiki code arose and which now underpins Wikipedia, and many other wikis. Barrett 2008 *MediaWiki: Wikipedia and Beyond* [31] is one of the few books to cover this mediawiki software.

F.2 Origins

The most useful account of the genesis of Wikipedia, without significant digression into any one discipline's perspective is Lih's *The Wikipedia Revolution* [217] from 2009. The book covers the

early origins within early internet bubble of the late 90s, through to initial origins of the cite culminating, as time of publishing with "10 million articles in 250 languages". A remarkable story. Lih's work is referenced, but as written for a general audience it therefore offers an approachable overview which is a useful base to further reading.

F.3 Books on Practical Use of Wikipedia

Surprisingly, there are very few books on how to edit and engage in Wikipedia. The two mentioned below both date from 2008. It might be argued from the present perspective that the Web, wikis and Wikipedia are now so well known that the subject is no longer worthy of print. In 2008, Wikipedia was already into its major growth spurt and these books aimed to help the initial user.

As books, they have more space to engage the reader than might be found in short Web articles or how-to videos. 'Youtube.com' [394], now the hub for much instructional material was only founded in 2005 and not the ubiquitous video source of today. Also, reading books today show how mediawiki has continued to evolve both in the visual style of the site and more importantly here, in the range of features available to editors.

These two books mentioned are Ayers, Matthews & Yates' *How Wikipedia Works: And How You Can Be a Part of It* [28] and Broughton's *Wikipedia: The Missing Manual* [62]. Though both have some coverage of Wikipedia origins, they major on the functional encyclopaedic aspect. Interestingly, neither mentions hypertext in the context of it being a medium: the same with use of links. There is no sense of the whole site as a single hypertext. Rather, it is just a set of web page articles and supporting resources. External links, as in references, are not given a positive gloss: "Though some external links are welcome…" [28, p.153]. Such an omission might be put down to the fact that such things were either known, or not of direct interest, to the books' primary audience of would-be Wikipedia editors.

Both books mention transclusion but without explaining it and they do so only in the context of template use [28, p.272, 62, p.425]. Bots are given cursory mention, but in a negative light as creators of errors [28, p.291] or as needing to be excluded from edit logs [62, p.108]. Neither book a clear sense that Wikipedia is a large hypertext or how links work within its structure—other than giving encouragement to make simple associative links and to link internally where possible. That is not to say the books are unhelpful for the new Wikipedia editor (social machine contributor) but it is concerning to see no acknowledgement of the bigger (hypertextual picture)—the sort of insight that will help editors move beyond the most basic work. As has been argued in the main thesis such an omission might said to be true of the Wikipedia site itself. Its many 'essays' tend to cover behaviour, style and provenance rather than how the sum of the parts are drawn together.

There have been two further books at a more trivial level, *Wikis for Dummies* and *Sams Teach Yourself Wikipedia in 10 Minutes* but these hardly scratch the surface of the subject.

Whilst the books are arguably now out of date as a 'how to', they should not be overlooked as a means to see the sort of issues the authors believed would be faced by someone wanting to edit or learn about Wikipedia. Arguably, those issues have not changed much and of these books *How Wikipedia Works: And How You Can Be a Part of It* was the most informative.

F.4 Books About Wikipedia As An Entity

Books about Wikipedia as an encyclopaedia are also few in number. Examples (in published date order) are: Darlby *The World And Wikipedia: How We Are Editing Reality* [95] (2009), O'Sullivan *Wikipedia: A New Community of Practice?* (2009) [282], Reagle *Good Faith Collaboration: The Culture of Wikipedia* (2012) [298], Tkacz *Wikipedia and the Politics of Openness* (2015) [349], Fichman & Hara *Global Wikipedia: International and Cross-cultural Issues in Online Collaboration* [122] (2014), Jemielniak *Common Knowledge?: An Ethnography of Wikipedia*, (2015) [188], Kennedy's *Textual Curation* [197] (2016), and Lund *Wikipedia, Work and Capitalism: A Realm of Freedom?* [221] (2017). All use a sociological perspective, Wikipedia as a stage for human endeavours and politicking, albeit upon a technical substrate. In doing so, there is scant regard to its hypertextual nature.

Absent is any reference to the hypertext medium, of linking strategy, or of content transclusion[1]. For these authors it seems the socio-technical balance is firmly towards the human, even if not directly stated. Only Tkacz mentions bots—other than in passing—discussing the case of HagermanBot [W260], the bot's signing of users' unsigned edits [349, pp.111–119], and the upset it caused in the user community. Again, human agency is privileged over other actors in the system.

A general theme is how Wikipedia continues to thrive given the free-wheeling libertarian culture within it and the way it is a form of social network. As a result issues like vandalism and consensus making consume much of the authors interests. Fichman & Hara [122], and Darlby [95] cast their net wider than others in considering activities across the many differing wikis rather than focus on the pre-eminent English-language 'en' wiki.The authors also take up matters of gender and representation, essentially those of diversity of voices and facts (not) represented. However, those speak to wider social issues and to the quality of content but not to the functional aspects of the Wikipedia docuverse.

Darlby's book is of note in that is written from the perspective of an active user, having contributed to the 'en' wiki and was an administrator on the Latin ('ln') wiki 'Vicipaedia' [W389]. Darbly write from the perspective of a wiki editor and the entire book is woven around evens occuring on or relating to the wiki. Despite publishing in 2009, there is no mention of bots[2]. Thus, even three years after the start of the introduction of bots, an active editor writing about

[1] The terms do turn up in the historical preambles but still with no evident interest in the medium in which the encyclopædia exists.

[2] There is passing reference to early work by 'Ram-man' [W332] seen at [95, p.118], but even this manages to mention the word bot.

Wikipedia at large ('en' and other languages) see no need to mention this facet of wiki work. The narrative is about edits and edit styles, and the engagement with these shows a more visceral engagement than that of academics writing with an external view. Darlby's exegesis reinforces the perspective that, even if informally, edit count is parsed as a form of standing in the community in a way missed by many external writers.

O'Sullivan's book [282] is in the form of a series of short essays, with an opening historical section giving context to Wikipedia's genesis. The book also has some more practical chapters on the use of the wikis but these lack depth. Lund [221], as the title implies, is taking an economic perspective and the book is somewhat removed for it's subject as a result. There is one chapter directly on Wikipedia that hurries through its subject, taking a back seat to the book's primary lens of economic and political considerations.

Despite different styles, Reagle, Takacz, and Jemielniak [298, 349, 188] all concentrate mainly on governance, behaviour and consensus. There is some mention of bots, but more as a reflection of their interaction with human actors than any investigation of bots in their own right. Fichman & Hara [122], by contrast do consider bots more directly and in a more global scope.

The general lack of writing about Wikipedia other than from the socio-political human perspective is another oblique suggestion as to the general lack of understanding of hypertext as a medium. O'Sullivan's conclusions on Wikipedia as a 'new community of practice' completely omits the agency of the technical and hypertextual aspects of the environment [282, pp.183–186].

Most interesting was Kennedy's *Textual Curation* [197] (2016) which takes a rhetorician's perspective. She contrasts encyclopædias old and new, using Ephraim Chambers' 1728 *Cyclopædia* and today's Wikipedia. In the latter case the use of non-human contributors (bots and other such automated processes) is acknowledged and use to reflect on the changing role of (traditional) curation. Sadly, the work does not engage with that issue at a more practical level, it is there that higher level abstractions are manifested in actions within the social machine.

F.5 Summary

For such a now well-established and high profile public resource, I have been surprised at how little long-form writing there has been about Wikipedia, outside academic journals. Indeed, some of the books above are essentially collections of essays. The nature and convention of academic publishing and its tendency to narrow intra-disciplinary focus means that it can be hard for a wider audience to get a sensible overview of the organisation. All too often, Wikipedia ia either a study substrate (i.e. mainly supplying data, e.g. to test an algorithm) or as a canvas upon which to paint a more abstracted picture, as in some of the books above.

It is telling that much of the discussion about how Wikipedia manages to keep going centres around human behaviour and prejudices, missing out the broader socio-technical setting. In some

sense, even discussion of bots is a reflection on their human operators thus missing the opportunity to look at that broader interaction.

Appendix G

Bot Tasks—Metadata Collected

This Appendix lists custom attributes (fields) added to my Tinderbox research files to hold metadata generated whilst researching RQ2-1.

TABLE G.1: Table of Custom Data Gathered About Wikipedia Bot Accounts.

Attribute Name	**Purpose**
BotDenied	Number of denied Bots: Requests for Approval (BRFA)s.
BotDenied	Number of denied BRFAs.
BotExpired	Number of expired BRFAs.
BotWithdrawn	Number of withdrawn BRFAs.
BotRevoked	Number of revoked BRFAs.
TaskCount	Number of approved BRFAs.
BotFlagStart	Date bot flag was granted (or un-flagged approval given). Date inferred for some early bots.
BotFlagEnd	Date bot had its flag removed.
FirstApprovedDate	Date of first approval, or if unclear date of first article-space edit.
LastEdit	The date of the last known edit of the bot (up to 30 June 2017 in most cases).
UnapprovalDate	Date the user page was marked as unapproved/inactive. Not necessarily removal of bot flag.
URL	User page URL
GlobalCountsURL	Links to user's list of wiki accounts, status and edit counts.
BRFAURL	URL of the first approved BRFA (or appropriate link for older bots).
LogsURL	URL showing block/permission changes for the bot.
SULURL	URL showing apportionment of activity by bot. (Sluggish response if bot has over 1M edits).
OperatorURL	URL of the user account operating the bot.

(continued on next page)

TABLE G.1: (continued from previous page)

Attribute Name	Purpose
APIURL1	Mediawiki API call to get rights, groups, ID and current count.
WikiAccCount	Number of wiki accounts held by the bot (not necessarily all with a bot flag).
WikiAccUsedCount	Number of accounts with 1 or more edits.
WikiMembership	Country codes of wikis in which bot has account(s).
EditCount	Overall number of edits by bot.
EnEditCount	Overall number of edits by bot within 'en' wiki.
HasBotInfoBox	User page has an infobox
HasDescription	User page has a description of the bot's activity.
HasFlagNoBRFA	Bot has a flag granted but no apparent BRFA.
HasGlobalFlag	Bot has a global bot flag across Wikipedia.
HasLinkedDescription	User page links to a description of the bot's tasks, e.g. on user talk page, bot-op's page, BRFA, etc.
HasNoUserPage	No user page exists for this account.
HasSULRename	This user was renamed with '~enwiki' suffix as part of Single User Login (SUL) (Single Unified Log-on).
IsActive2018	Bot was still active in 2018.
IsAdminBot	Bot has admin status.
IsCurrentBot	Bot is flagged as at end of June 2017 and has made edits within the last two years.
IsDeactivated	Has user page 'Deactivated' banner.
IsListedActive	Indicated if listed as an active bot at end of June 2017.
IsRenamed	This bot's username has been renamed.
IsReplaced	This bot is replaced by another bot.
IsReplacement	This bot is a replacement for an existing bot.
IsRetired	Has user page 'Retired' banner.
IsSULRename	Account renamed with ~enwiki suffix as part of SUL process.
UsesAWB	Bot uses AutoWikiBrowser for all/some of its work.
AutoOrMan	If known, does the bot run manually, fully automatically or in some other mode?
EncodedName	Bot user name encoded for URL use.
IconStatus	Type of status icon for the bot, as shown on the user page.
Operator	Name of the user operating the bot.
PrimaryTask	The most significant/largest task, from TaskSet, undertaken by the bot (N.B. some have multiple tasks).
RenamedTo	New name of renamed bot.
ReplacedBy	Name of replacement/superseding bot(s).
TaskSet	Nature of task(s) undertaken by the bot.

(continued on next page)

TABLE G.1: (continued from previous page)

Attribute Name	**Purpose**
UnapprovalDelay	Interval from first approval to cessation of approval.
UnapprovalReason	Reason for bot being marked unapproved, e.g. inactive for over X period of time.

Appendix H

Bot Tasks—Full Classification of Primary Task

H.1 Task Data from RQ2.1

This Appendix holds the full listing of tasks coded for bot account tasks researched as part of RQ2.1. Data source is the 'en' Wikipedia, sampling records from December 2002 (first acknowledge bot account) to June 2017.

H.1.1 Table of Tasks

The tasks listed in the table below are explained in section H.1.2. The top 20 tasks by count of Bots is shown at Table 5.1 in section 5.2.3

TABLE H.1: 'en' Wikipedia Bots, by Primary Task: Number & % of Discrete Bots; Number & % Aggregate Edits; Average Edits per Bot per Task. (Sort Order: Discrete Bots)

Task	Bots	%	**Cum. Edits**	%	**Av. Edits/Bot**
Interwiki	310	27.903	13,711,169	16.961	44,230
Clerking	103	9.271	5,448,985	6.740	52,903
Templates	91	8.191	3,733,545	4.618	41,028
WikiProjects	79	7.111	4,418,379	5.466	55,929
Maintenance	77	6.931	2,796,232	3.459	36,315
Images	52	4.680	3,448,490	4.266	66,317
Categories	44	3.960	2,433,374	3.010	55,304
Vandalism	35	3.150	8,845,214	10.942	252,720
Link Repair	23	2.070	4,069,398	5.034	176,930
List Creation	23	2.070	1,011,859	1.252	43,994
Redirects	22	1.980	873,361	1.080	39,698
Newsletter	21	1.890	841,741	1.041	40,083
Disambiguation	20	1.800	221,485	0.274	11,074

(continued on next page)

TABLE H.1: (continued from previous page)

Task	Bots	%	**Cum. Edits**	%	**Av. Edits/Bot**
Data Input	19	1.710	144,621	0.179	7,612
Code Cleanup	18	1.620	2,844,034	3.518	158,002
Archiving	17	1.530	1,524,954	1.886	89,703
General Tasks	14	1.260	14,181,512	17.543	1,012,965
Infoboxes	13	1.170	487,708	0.603	37,516
Stub Creation	10	0.900	519,357	0.642	51,936
Data Extraction	8	0.720	1,056	0.001	132
Spelling	7	0.630	27,607	0.034	3,944
Permissions	6	0.540	33,104	0.041	5,517
Sockpuppets	6	0.540	115,121	0.142	19,187
Taxonomy	6	0.540	48,147	0.060	8,025
Copyright	5	0.450	82,157	0.102	16,431
Messages	5	0.450	1,683,110	2.082	336,622
Sandbox	5	0.450	112,159	0.139	22,432
AfD	4	0.360	25,737	0.032	6,434
References	4	0.360	60,194	0.074	15,049
TASK NOT KNOWN	4	0.360	9,082	0.011	2,271
Welcome	4	0.360	321,474	0.398	80,369
AfC	3	0.270	121,626	0.150	40,542
Assessment	3	0.270	11,221	0.014	3,740
Formatting	3	0.270	46,248	0.057	15,416
Geo Data	3	0.270	947,377	1.172	315,792
Page Renaming	3	0.270	325,594	0.403	108,531
Research	3	0.270	4	0.000	1
Status	3	0.270	11,032	0.014	3,677
Typos	3	0.270	9,543	0.012	3,181
De-bugging	2	0.180	38	0.000	19
Signatures	2	0.180	2,217,475	2.743	1,108,738
Emailer	2	0.180	1,213	0.002	607
Code Encoding	2	0.180	40,754	0.050	20,377
Link Creation	2	0.180	27,541	0.034	13,771
One-off Change	2	0.180	925	0.001	463
Plagiarism	2	0.180	255,392	0.316	127,696
Portal	2	0.180	3,288	0.004	1,644
Warnings	2	0.180	47,984	0.059	23,992
BLP	1	0.090	122,638	0.152	122,638
COI	1	0.090	1,018,151	1.259	1,018,151
Data Precision	1	0.090	6,712	0.008	6,712
Deletions	1	0.090	5	0.000	5
Grammar	1	0.090	3,795	0.005	3,795
IRC	1	0.090	45	0.000	45
Joke	1	0.090	68	0.000	68
Link Suggestion	1	0.090	636	0.001	636
Page Moves	1	0.090	235	0.000	235
Patrol	1	0.090	944	0.001	944
Personal	1	0.090	82	0.000	82
Read-only	1	0.090	0	0.000	0
Wikidata	1	0.090	1,545,421	1.912	1,545,421

(continued on next page)

TABLE H.1: (continued from previous page)

Task	Bots	%	Cum. Edits	%	Av. Edits/Bot
Wikivoyage	1	0.090	209	0.000	209
Totals	**1,111**		**80,840,562**		

H.1.2 Description of Terms

Given below are short descriptions of the task types listed in Table H.1.

AfC. Work relating to supporting activities relating to the Articles for Creation [W436] page. This describes the considerations to be applied by those creating pages, especially new editors with little experience of article creation.

AfD. Work relating to supporting activities relating to Articles for Deletion [W437] page. Articles considered worthy of deletion, are listed here for seven days and then deleted if there is community consensus for their removal. Generally these are new articles of insufficient noteworthiness or in violations of various Wikipedia rules. They may also be existing page pages that are no longer required, for example if articles are merged.

Archive. Archival of talk pages, logs and various administrative and arbitration pages.

Assessment. Work relating to the assessment, and marking of, article quality. The Content assessment [W634] page lists the various quality levels from Stub through to Featured Article, Wikipedia's highest quality award.

BLP. Wikipedia has strict rules surrounding 'biographies of living persons' [W440] for which attention must be given to 'to verifiability, neutrality, and avoidance of original research' as 'all quotations and any material challenged or likely to be challenged must be supported by an inline citation to a reliable, published source'. This task involves BLP-related matters.

Categories. Work relating to application or management of categories.

Clerking. Work involving a number of discrete but related tasks. For instance, querying to make a list then acting on items in a list, or watching an administrative page for new entries, informing watchers and archiving old entries.

Code Cleanup. Edits to the source code of a page rather the rendered screen content. This may involve replacing old code or improving code layouts—although strictly cosmetic code edits are deprecated.

Code Encoding. Some less usual charters can cause encoding issues either in source code and or rendering.

COI. Issues of **C**onflict **o**f **I**nterest. All Wikipedia editors are bidden to avoid COIs [W632].

Copyright. Checking submitted materials are not bound by copyright and acting on inappropriate use of such files.

Data Input. This may involve uploading new material or placing new/altered criteria into existing source code.

Data Extraction. The collection and downloading of data for off-Wikipedia use.

Data Precision. Issues surrounding the accuracy of statistics, geo-coded data and general numbers.

De-bugging. Bots de-bugging other bot or support code. For example, de-bugging the pywikipediabot framework

Deletions. Work related to deleting articles.

Disambiguation. Work on disambiguation pages, which hold list of links to similarly named items in different subject areas.

Edit Signing. Some editors forget to sign their posts in talk and admin pages. Bots can check edit logs and insert missing signatures to save editors having to read page histories to figure this out.

Emailer. Work relating to sending information out to approved preselected groups, often relating to WikiProjects.

Formatting. Changing article layouts (though not trivial edits). For instance moving section of articles in the same subject area to use a common layout.

General Tasks. This describes a bot whose operator is happy to take on a variety of task such that it is impossible to define a primary type of task undertaken.

Geo Data. Work relating to geographical and geo-coded data.

Grammar. Work related to correcting grammar in articles. Such tasking is no longer approved as it can lead to edit wars.

Images. Work related checking image sources, formats and upload/transfer to the wiki Commons, which is the intended space for all such assets.

Infoboxes. Work related to the insertion of infoboxes or editing parameters therein.

Interwiki. Creating and maintaining inter-wiki links, i.e. cross-connecting articles on the same topic in different language wikis.

IRC. Monitoring **I**nternet **R**elay **C**hat channels.

Joke. A single case where a joke bot (‘Skynet‘) was proposed and accepted.

Link Creation. Actions to create new links between articles or between project pages and relevant article talk pages.

Link Repair. Work to resolve broken links. This may occur if externally linked sites go dark, or due to internal reorganisation (renaming/moving, deletion, merging).

Link Suggestion. A task to suggest possible links, but where human review is still needed.

List Creation. The creation of lists to help editors with their work. If also part of a series of tasks by the same bot, these are instead classed as part of a clerking task.

Maintenance. Generally some form of list maintenance, with other miscellaneous maintenance activities.

Messages. Sending in-wiki messages to editors. For instance, contacting all members of a particular wiki task or interest area.

Newsletter. Sending newsletters to subscribed editors.

One-off Change. Rare occasions of a bot required for a task to be undertaken once, but at a scale precluding human editing.

Page Moves and **Page Renaming**. Broadly a similar task, as a page move normally involves renaming the page. Actual task BRFAs are not always stated clearly enough to make the detail plain.

Patrol. Work relating to 'patrolling' of pages to monitor edit behaviour.

Permissions. Work relating to setting page permissions, either to block a page from access to prohibit edits by some or all classes of editors.

Personal. An editor may wish to use a bot in their own user space but think it appropriate to make a formal request.

Plagiarism. Monitoring edits for plagiarised content. Unapproved re-use of off-wiki content is not allowed (without appropriate permission).

Portal. Activity related to specific wiki Portals (essentially organisation/landing pages for particular subjects).

Read-only. A small group of bots only read data but make no edits, often extracting data for purposes such as research.

Redirects. When pages are renamed/moved it can be necessary to trawl through previously linked pages and fix the link but generally the 'old' page is made into a re-direct. Such pages automatically push a request onto the correct target page. A 'soft' redirect occurs where the redirect page loads but with an indication as to the page now in use.

References. Maintenance of references and citations.

Research. Bots whose work is only research related.

Sandbox. The wiki allows for many pages, including all user pages, to have a 'sandbox' for experimentation such as trying out unfamiliar mark-up code or making API queries. Bots can be used to automatically clear out sandboxes to avoid the build-up of broken/unwanted/abandoned code.

Sockpuppets. Occasionally, editors may attempt to create a bot that is then used as a (sockpuppet), often as a way to get around bans.

Spelling. Bots working to correct spelling errors, generally in real names. General grammar and spelling bots are now deprecated.

Status. Bots that monitor or report status, such as where a user is currently logged in or the number of pending requests on some admin pages.

Stub Creation. The task of creating new article pages which little or no information, ready for some other activity to the populate them.

TASK NOT KNOWN. Unknown tasking occurs if a bot has been banned and the user page has been blocked or deleted and there is no record elsewhere of the bot's activities.

Taxonomy. In some areas of science taxonomic information needs to be applied consistently and bots have assisted in such tasks.

Templates. Many bot tasks require the addition or removal of templates from a page's source code (this is also sometimes referred to as 'tagging').

Typos. Now generally deprecated, this task relates to content errors other than simple spelling errors or grammar.

Vandalism. Vandalism is a constant threat for an open system like Wikipedia and is now policed by the most complex and powerful bots in Wikipedia.

Warnings. A technical task pf substituting warning templates on user pages. The substitution replaces the transcluded template content with fixed content. This task was undertaken to ease loading on servers from heavily transcluded templates.

Welcome. A bot to welcome new users. Generally a deprecated task outside a few approved contexts.

Wikidata. Work related to Wikidata use in Wikipedia.

WikiProjects. Tasks associated with a WikiProject, often a category or template 'tagging' task. Where these relate to WikiProject(s) the bot has been allocated this primary task.

Wikivoyage. Work related to the Wikivoyage project.

Appendix I

Bot Requests—Metadata Collected

This Appendix lists custom attributes (fields) added to my Tinderbox research files to hold metadata generated whilst researching RQ2-2.

TABLE I.1: Table of Custom Data Gathered About Wikipedia Bot Request Threads.

Attribute Name	**Purpose**
AnswerCount	Number of request threads in which answer made.
ArticleCount	Articles per archive page.
BRFACount	Number of Bots: Requests for Approval (BRFA)s arising from request.
BRFAURL1	URL of first BRFA arising from request.
BRFAURL2	URL of second BRFA arising from request.
BRFAURL3	URL of third BRFA arising from request.
BRFAdBots	Bots submitting BRFAs resulting from request.
DoneByUser	Request actioned by user.
EditCount	Number of discrete contributions to request.
ElapsedDays	Number of whole days between request initiation and answer/closure of thread.
FirstUser	Username raising the request.
HasAnon	Thread includes edits from IP-based (not logged in) users.
HasApprovedBRFA	Thread resulted in a (new) approved BRFA.
HasFix	User has fixed one or more requests.
HasResolution	User has provided resolution to one or more requests.
IconSet	Status/progress icon(s) are used in the thread.
MoverEditCount	Number of edits that are related to (partial) thread move/redirect.
MoverNameSet	User(s) only editing to make admin points, e.g. moving content to/-from other pages.
OverallSequence	Sequence number within the overall set of threads.

(continued on next page)

TABLE I.1: (continued from previous page)

Attribute Name	Purpose
RefdBots	Bot accounts referenced by name in the thread.
ReqOutcome	Outcome(s) of the request.
Resolution	Primary outcome of the request.
TaskSet	Set of task(s) for this request.
TaskSize	If given, number of items the task must process.
ThreadMonth	Month of year in which thread started (used for totalling).
ThreadYear	Year in which thread started (used for totalling).
URL	URL of request thread.
URLStub	Wikipedia URL-encoded stub of topic name.
UserCount	Number of discrete users participating in the thread.
UserSet	Set of discrete users in the thread. Excludes users only making admin edits.
WasMoved	Topic moved from a start elsewhere.

TABLE I.2: Table of Custom Data Gathered About Users Participating in Wikipedia Request Threads.

Attribute Name	Purpose
HasFirstEdit	User has started one of more requests.
StartCount	Number of requests threads started by user.
HasReply	User has replied in one or more requests.
FixCount	Number of requests where editor noted as fixing.
StartDate	First input to threads.
EndDatee	Last input to threads.
ActiveDayCount	Number of days from first thread to last.

Appendix J

Bot Requests—Full Classification of Tasks

This Appendix holds the full listing of categories coded for the bot requests researched as part of RQ2.2. Data source is the 'en' Wikipedia's 'Bot requests' [W455], sampling in the period August 2005 (first request) to July 2017.

J.1 Task Data from RQ2.2

The tasks listed in the table below are explained in section J.1.1. The top 20 categories, by count are showing Table 5.2 in Section 5.3.2

TABLE J.1: Categorisation of 4,276 Requests for Bot Assistance

Category	Count	%
Tagging	565	13.213
Link management	389	9.097
Bot how-to	362	8.466
Clerking	258	6.034
Image management	233	5.449
List creation	223	5.215
Redirects	184	4.303
Infoboxes	181	4.233
Categories	168	3.929
Templates	142	3.321
Bot repair	137	3.204
Subediting	125	2.923
Tag parameters	110	2.572
Notifications	108	2.526
References	84	1.964
Vandalism	60	1.403
Article move	58	1.356

(continued on next page)

TABLE J.1: (continued from previous page)

Category	Count	%
Archiving	56	1.310
Assessment	55	1.286
Interwiki	53	1.239
Citations	52	1.216
Code cleanup	46	1.076
List order	41	0.959
Formatting article text	40	0.935
Data input	39	0.912
Spelling	35	0.819
Article content	28	0.655
Disambiguation	27	0.631
Geo-data	27	0.631
Statistics	25	0.585
Article deletion	24	0.561
Stub generation	24	0.561
Article names	23	0.538
Listing update	20	0.468
Data extraction	18	0.421
Navboxes	17	0.398
Grammar	16	0.374
Null edits	16	0.374
Welcome	16	0.374
Copyright	15	0.351
Scripting	15	0.351
Blocking	12	0.281
Page Protection	11	0.257
Sandbox	11	0.257
Spam	10	0.234
Userboxes	10	0.234
Signatures	9	0.210
Bot-Flag requests	8	0.187
File management	8	0.187
Merge	7	0.164
User activity	7	0.164
Transwikification	6	0.140
BLP	6	0.140
COI	6	0.140
Character correction	5	0.117
Sockpuppets	5	0.117
Code encoding	4	0.094
Research	4	0.094
Translation	4	0.094
Cache purge	3	0.070
New article stubs	3	0.070
Accessibility	2	0.047
Bad language	2	0.047
Edit requests	2	0.047
Reporting	2	0.047

(continued on next page)

TABLE J.1: (continued from previous page)

Category	Count	%
Reversions	2	0.047
Search	2	0.047
TASK NOT KNOWN	2	0.047
User names	2	0.047
Interaction bans	1	0.023
Mediawiki interface	1	0.023
Read-only bot	1	0.023
Stub removal	1	0.023
Thanks	1	0.023
User status	1	0.023
Total	**4,276**	

J.1.1 Description of Terms

Accessibility. Tasks related to ensuring ongoing accessibility of the site and thus access for more people of all abilities.

Edit requests. A user may request a bot's help to do specific edits. In general this is an issue of scale rather the nature of the task.

Article content. Tasks relating to the content of the (readable) page.

Data extraction. Requests involving searching for and downloading data.

Article deletion. Articles may be deleted if no longer needed.

Article move and **Article names**. These are essentially the same thing as an articles name is also its address within the wiki. When harmonising articles within a category it mat prove worth while to ue a bot to automate and speed such changes.

Archiving. Requests for bots to undertake archive of talk pages. Bots exist for this purpose but their existence is not well promulgated.

Assessment. Work relating to the assessment, and marking of, article quality. The Content assessment [W634] page lists the various quality levels from Stub through to Featured Article, Wikipedia's highest quality award.

Bad language. Policing incivility.

Blocking. Blocks generally arise as a result disciplinary action within the wiki and bot assistance might be requested.

BLP. Wikipedia has strict rules surrounding 'biographies of living persons' [W440] for which attention must be given to 'to verifiability, neutrality, and avoidance of original research' as 'all

quotations and any material challenged or likely to be challenged must be supported by an inline citation to a reliable, published source'. This task involves BLP-related matters.

Bot-Flag requests. The bot request page is sometimes misunderstand as the place to ask for a bot flag rather than the place to ask for bot assistance.

Bot how-to. Although not strictly part of the remit of the bot requests page, users still regularly request help 'to make a bot'.

Bot repair. Requests for a broken bot to be fixed, taken over or replaced.

Cache purge. Cache purges are requested to ensure that articles get acted on by the constant background processing 'queue'.

Categories. Requests for allocation or removal to categories from specific groups of articles (or user created lists).

Character correction. This relates to precise fine-tuning of article copy.

Citations. Maintenance of citations.

Clerking. A clerking task relates to requests where the task involves several discrete tasks, q.v..

Code encoding. Requests for help with the correct encoding of article data.

Code cleanup. Requests with help working on the source code of articles.

COI. Issues of **C**onflict **o**f **I**nterest. All Wikipedia editors are bidden to avoid COIs [W632].

Copyright. Requesting assistance with Checking submitted materials are not bound by copyright and acting on inappropriate use of such files.

Data input. Requests with assistance inserting data in Wikipedia at volume, but not stub generation.

Disambiguation. Work on disambiguation pages, which hold list of links to similarly named items in different subject areas.

File management. Requests relating to the processing and general management of image and other files (video, audio, etc.). These are normally being moved from the wiki to the Wikipedia Commons which is the correct repository for such assets. The processes also involve checks for copyright issues.

Formatting article text. An editor may request help to (re-)structure a group of articles, perhaps transposing sections to get a consistent structure.

Grammar. As with spelling, this is task is deprecated for automation but it does not stop people asking.

Geo-data. Geo-located data is using in mapping functions and there can be issues of over- or under- specification of accuracy. It is not entirely clear that this is due to concerns as to accuracy as opposed to personal style.

Image management. Due to issues of copyright, the use of image within Wikipedia is closely monitored. This task tends to involve a range task, including detect (suspect) new images in articles, clearance of rights on images and the transfer of images to the Wikipedia Commons which is the correct repository for approved images.

Infoboxes. Requests related to the insertion of infoboxes or editing parameters therein.

Interaction bans. In cases of uncivil behaviour a sanction can be that named editors should not directly interact on-wiki. this request is in support of monitoring such sanctions.

Interwiki. Creating and maintaining inter-wiki links, i.e. cross-connecting articles on the same topic in different language wikis.

Link management. This involves either re-directing links due to article move or deletion.

List order. Whether lists are created by human editors, scripts or bots, there are requests that bots are used to re-order these lists - especially if used in article page content.

List creation. Requests for a bot to create a list based on some query. Often these are requested so a human editor can work through them as the editors lack the technical ability to run such queries themselves.

Listing update. A request for a list, normally one generated by a script, to be updated. This is either to find new items or to remove now-processed ones.

Mediawiki interface. A small number of task relate to the mediawiki software or to functions on the mediawiki domain.

Merge. As articles mature or a subject fades in importance it may be necessary to merge two or more articles into one.

Navboxes. This type of request is for the expanding navboxes seen at the foot of some article pages. Navboxes can be 'opened' (expanded) to reveal sets of links pertinent to a subject. They are added via a template.

New article stubs. The mass creation of stub articles for later population either by human editors or other bots.

Notifications. Tasks to message editors about a particular issue or subject.

Null edits. Null edits are used to 'touch' a file such that the background 'queue' notices the file and acts on it without having to make a literal edit of the screen content.

Page Protection. Page protection is used with controversial articles to avoid edit warring and vandalism.

Read-only bot. Such requests are generally wanting the bot's allowance of faster/more intensive access to APIs.

Redirects. This involves the need to redirect incoming links when article ceases to have a functions. This may arise through, deletion, renaming or an article being split.

References. Maintenance of references.

Reporting. Requests for bot-generated reports.

Research. Requests for bot-based activity for research purposes.

Reversions. Request relating to reverting edits, either those made in error or due to vandalism.

Sandbox. All accounts have a 'sandbox' page for testing code. Editors may request assistance purging these of stale and abandoned material.

Scripting. Requests for assistance with script related issues, either using them or coding them.

Search. Requests for assistance in finding particular articles or types of articles.

Signatures. A small number of bots specialise in scanning recent edits and adding signatures to unsigned posts.

Spam. Requests may occur relating to spam, as part of anti-vandalism activities.

Spelling. Requesting bots to work on correcting spelling errors is generally now deprecated.

Sockpuppets. Occasionally, editors may attempt to create a bot that is then used as a (sockpuppet), often as a way to get around bans.

Statistics. Bots may be requested to assists with data collection and generation of statistical data.

Stub generation. A request to generate stubs, either as a large single pass or iteratively based supplied lists.

Stub removal. A request to remove (i.e. delete) unwanted article stubs.

Subediting. Requests were assistance is required with editing copy.

Tagging. Addition of template and/or category to page source code. When 'tagging' templates are most usually adding banners indicating the page is part of the certain area of interest. Templates may add categories or category code can be added directly. The category code allows ongoing background processes to identify pages as needing to be added to the listing on the relevant category page(s). Owing to the way tasks are described, it is not always to clearly define which of these activities form part of the task as described.

Tag parameters. This involves that adding or changing of parameters applied to either template or category code in article source code.

TASK NOT KNOWN. Insufficient data was available to define the nature of a few requests.

Templates. Requests for allocation or removal to categories from specific groups of articles (or user created lists).

Thanks. A small number of 'requests' were users simply adding a comment to say thanks for help received.

Translation. Requests with help translating content.

Transwikification. Asking for help connecting between Wikipedia projects, similar to interwiki work but connecting outside the main Wikipedia project.

Userboxes. Requesting assistance to apply userboxes.

User activity. Assistance tracking own or other users work.

User names. Despite numerous pieces of guidance as to (in)appropriate usernames, some policing of this is necessary.

User status. Reporting user status as online or offline.

Vandalism. Requests for help dealing with vandalism.

Welcome. Asking for a welcome bot (which is type publicised as generally denied).

Appendix K

Bot Activities and Requests Compared—Full Classification of Tasks

This Appendix holds the full listing of categories coded for both bot tasks in RQ1.2 (Table H.1) and bot requests researched in RQ2.2 (Table J.1).

K.1 Comparing Bot Activities Against Requested Activities

This table is a cross-comparison of Bot Data from RQs 2.1 & 2.2.

TABLE K.1: Bots primary activity categories matched against bot request categories.

Bot Activities	**#**	**%**	**Bot Requests**	**#**	**%**
Accessibility	2	0.047	—		
AfC	3	0.270	—		
AfD	4	0.360	—		
Archiving	17	1.530	Archiving	56	1.310
—			Article content	28	0.655
—			Article deletion	24	0.561
—			Article move	58	1.356
—			Article names	23	0.538
Assessment	3	0.270	Assessment	55	1.286
—			Bad language	2	0.047
—			Blocking	12	0.281
BLP	1	0.090	BLP	6	0.140
—			Bot how-to	362	8.466
—			Bot repair	137	3.204
—			Bot-Flag requests	8	0.187
—			Cache purge	3	0.070

(continued on next page)

TABLE K.1: (continued from previous page)

Bot Activities	#	%	**Bot Requests**	#	%
Categories	44	3.960	Categories	168	3.929
—			Character correction	5	0.117
Clerking	103	9.271	Clerking	258	6.034
Code Cleanup	18	1.620	Code Cleanup	46	1.076
Code Encoding	2	0.180	Code Encoding	4	0.094
COI	1	0.090	COI	6	0.140
Copyright	5	0.450	Copyright	15	0.351
Data Extraction	8	0.720	Data Extraction	18	0.421
Data Input	19	1.710	Data Input	39	0.912
Data Precision	1	0.090	—		
De-bugging	2	0.180	—		
Deletions	1	0.090	—		
Disambiguation	20	1.800	Disambiguation	27	0.631
—			Edit Requests	2	0.047
—			File Management	8	0.187
Formatting	3	0.270	Formatting	40	0.935
General Tasks	14	1.260	—		
Geo Data	3	0.270	Geo Data	27	0.631
Grammar	1	0.090	Grammar	16	0.374
Images	52	4.680	Image Management	233	5.449
Infoboxes	13	1.170	Infoboxes	181	4.233
—			Interaction Bans	1	0.023
Interwiki	310	27.903	Interwiki	53	1.239
IRC	1	0.090	—		
Joke	1	0.090	—		
Link Creation	2	0.180	—		
Link Repair	23	2.070	—		
Link suggestion	1	0.090	—		
—			Link Management	389	9.097
List Creation	23	2.070	List Creation	223	5.215
—			List Order	41	0.959
—			Listing Update	20	0.468
Maintenance	77	6.931	—		
—			Mediawiki Interface	1	0.023
—			Merge	7	0.164
—			Navboxes	17	0.398
—			New article stubs	3	0.070
Emailer	2	0.180	Notifications}	108	2.526
Messages	5	0.450	}		
Newsletter	21	1.890	}		
One-off Change	2	0.180	—		
Page Moves	1	0.090	—		
Page Renaming	3	0.270	—		
Patrol	1	0.090	—		
Permissions	6	0.540	—		
Personal	1	0.090	—		
Plagiarism	2	0.180	—		
Portal	2	0.180	—		

(continued on next page)

TABLE K.1: (continued from previous page)

Bot Activities	**#**	**%**	**Bot Requests**	**#**	**%**
—			Null Edits	16	0.374
—			Page Protection	11	0.257
Read-only	1	0.090	Read-only	1	0.023
Redirects	22	1.980	Redirects	184	4.303
References}	4	0.360	References	84	1.964
}			Citations	52	1.216
—			Reporting	2	0.047
Research	3	0.270	Research	4	0.094
—			Reversions	2	0.047
Sandbox	5	0.450	Sandbox	11	0.257
—			Scripting	15	0.351
—			Search	2	0.047
Signatures	2	0.180	Signatures	9	0.210
Sockpuppets	6	0.540	Sockpuppets	5	0.117
—			Spam	10	0.234
Spelling	7	0.630	Spelling	35	0.819
—			Statistics	25	0.585
Status	3	0.270	—		
Stub Creation	10	0.900	Stub Generation	24	0.561
—			Stub Removal	1	0.023
—			Subediting	125	2.923
—			Tag Parameters	110	2.572
—			Tagging	565	13.213
TASK NOT KNOWN	4	0.360	TASK NOT KNOWN	2	0.047
Taxonomy	6	0.540	—		
Templates	91	8.191	Templates	142	3.321
Typos	3	0.270	—		
—			Thanks	1	0.023
—			Translation	4	0.094
—			Transwikification	6	0.140
—			User activity	7	0.164
—			User names	2	0.047
—			User status	1	0.023
—			Userboxes	10	0.234
Vandalism	35	3.150	Vandalism	60	1.403
Warnings	2	0.180	—		
Welcome	4	0.360	Welcome	16	0.374
Wikidata	1	0.090	—		
WikiProjects	79	7.111	—		
Wikivoyage	1	0.090	—		
Totals	**1,111**			**4,276**	

Appendix L

Wikipedia Per-language Namespace Localisation

Table of Wikipedia namespaces showing lack of coherence, cross-language.

TABLE L.1: Wikipedia per-language namespace localisation - part 1a (namespaces -2–103): English, Cebuano, German, Spanish and French

Space #	en	ceb	de	es	fr
-2	Media	Medya	Medium	Medio	Média
-1	Special	Espesyal	Spezial	Especial	Spécial
0	(Main namespace)				
1	Talk	Hisgot	Diskussion	Discusión	Discussion
2	User	Gumagamit	Benutzer	Usuario	Utilisateur
3	User talk	Hisgot sa Gumagamit	Benutzer Diskussion	Usuario discusión	Discussion utilisateur
4	Wikipedia	Wikipedia	Wikipedia	Wikipedia	Wikipédia
5	Wikipedia talk	Hisgot sa Wikipedia	Wikipedia Diskussion	Wikipedia discusión	Discussion Wikipédia
6	File	Payl	Datei	Archivo	Fichier
7	File talk	Hisgot sa Payl	Datei Diskussion	Archivo discusión	Discussion fichier
8	MediaWiki	MediaWiki	MediaWiki	MediaWiki	MediaWiki
9	MediaWiki talk	Hisgot sa MediaWiki	MediaWiki Diskussion	MediaWiki discusión	Discussion MediaWiki
10	Template	Plantilya	Vorlage	Plantilla	Modèle
11	Template talk	Hisgot sa Plantilya	Vorlage Diskussion	Plantilla discusión	Discussion modèle
12	Help	Tabang	Hilfe	Ayuda	Aide
13	Help talk	Hisgot sa Tabang	Hilfe Diskussion	Ayuda discusión	Discussion aide
14	Category	Kategoriya	Kategorie	Categoría	Catégorie
15	Category talk	Hisgot sa Kategoriya	Kategorie Diskussion	Categoría discusión	Discussion catégorie
100	Portal		Portal	Portal	Portail
102				Wikiproyecto	Projet
103				Wikiproyecto Discusión	Discussion projet

TABLE L.2: Wikipedia per-language namespace localisation - part 1b (namespaces 104–2600): English, Cebuano, German, Spanish and French

Space #	en	ceb	de	es	fr
104				Anexo	Référence
105				Anexo Discusión	Discussion référence
106					
107					
108	Book				
109	Book talk				
114					
115					
118	Draft				
119	Draft talk				
446	Education Program			Programa educativo	
447	Education Program talk			Programa educativo discusión	
710	TimedText				
711	TimedText talk				
828	Module	Module	Modul	Módulo	Module
829	Module talk	Module talk	Modul Diskussion	Módulo discusión	Discussion module
2300	Gadget	Gadget	Gadget	Accesorio	Gadget
2301	Gadget talk	Gadget talk	Gadget Diskussion	Accesorio discussion	Gadget talk
2302	Gadget definition	Gadget definition	Gadget-Definition	Accesorio definición	Gadget definition
2303	Gadget definition talk	Gadget definition talk	Gadget-Definition Diskussion	Accesorio definición discusión	Gadget definition talk
2600	Topic	Topic	Thema	Tema	Sujet

TABLE L.3: Wikipedia per-language namespace localisation - part 2a (namespaces -2–103): English, Italian, Japanese and Dutch

Space #	en	it	ja	nl
-2	Media	Media	メディア	Media
-1	Special	Spécial	特別	Speciaal
0	(Main namespace)			
1	Talk	Discussione	ノート	Overleg
2	User	Utente	利用者	Gebruiker
3	User talk	Discussioni utente	利用者-会話	Overleg gebruiker
4	Wikipedia	Wikipedia	Wikipedia	
5	Wikipedia talk	Discussioni Wikipedia	Wikipedia-ノート	Overleg Wikipedia
6	File	File	ファイル	Bestand
7	File talk	Discussioni file	ファイル-ノート	Overleg bestand
8	MediaWiki	MediaWiki	MediaWiki	MediaWiki
9	MediaWiki talk	Discussioni MediaWiki	MediaWiki-ノート	Overleg MediaWiki
10	Template	Template	Template	Sjabloon
11	Template talk	Discussioni template	Template-ノート	Overleg sjabloon
12	Help	Aiuto	Help	Help
13	Help talk	Discussioni aiuto	Help-ノート	Overleg help
14	Category	Categoria	Category	Categorie
15	Category talk	Discussioni categoria	Category-ノート	Overleg categorie
100	Portal	Portale	Portal	Portaal
101	Portal talk	Discussioni portale	Portal-ノート	Overleg portaal
102		Progretto	プロジェクト	
103		Discussioni Progretto	プロジェクト-ノート	

TABLE L.4: Wikipedia per-language namespace localisation - part 2b (namespaces 104–2600): English, Italian, Japanese and Dutch

Space #	en	it	ja	nl
104				
105				
106				
107				
108	Book			
109	Book talk			
114				
115				
118	Draft			
119	Draft talk			
446	Education Program			Onderwijsprogramma
447	Education Program talk			Overleg onderwijsprogramma
710	TimedText			
711	TimedText talk			
828	Module	Modulo	モジュール	Module
829	Module talk	Discussione modulo	モジュール‐ノート	Overleg module
2300	Gadget	Accessorio	Gadget	Gadget
2301	Gadget talk	Discussioni accessorio	Gadget talk	Gadget talk
2302	Gadget definition	Definizione accessorio	Gadget definition	Gadget definition
2303	Gadget definition talk	Discussioni definizione accessorio	Gadget definition talk	Gadget definition talk
2600	Topic	Argomento	Topic	Onerwerp

TABLE L.5: Wikipedia per-language namespace localisation - part 3a (namespaces -2–103): English, Russian, Swedish and English WikiSource

Space #	en	ru	sv	ws-en
-2	Media	Медиа	Media	Media
-1	Special	Служебная	Special	Special
0	(Main namespace)			
1	Talk	Обсуждение	Diskussion	Talk
2	User	Участник	Användare	User
3	User talk	Обсуждение участника	Användardiskussion	User talk
4	Wikipedia	Википедия	Wikipedia	Wikisource
5	Wikipedia talk	Обсуждение Википедии	Wikipediadiskussion	Wikisource talk
6	File	Файл	Fil	File
7	File talk	Обсуждение файла	Fildiskussion	File talk
8	MediaWiki	MediaWiki	MediaWiki	MediaWiki
9	MediaWiki talk	Обсуждение MediaWiki	MediaWiki-diskussion	MediaWiki talk
10	Template	Шаблон	Mall	Template
11	Template talk	Обсуждение шаблона	Malldiskussion	Template talk
12	Help	Справка	Hjälp	Help
13	Help talk	Обсуждение справки	Hjälpdiskussion	Help talk
14	Category	Категория	Kategorie	Category
15	Category talk	Обсуждение категории	Kategoridiskussion	Category talk
100	Portal	Портал	Portal	Portal
101	Portal talk	Обсуждение портала	Portaldiskussion	Portal talk
102		Инкубатор		Author
103		Обсуждение Инкубатора		Author talk

TABLE L.6: Wikipedia per-language namespace localisation - part 3b (namespaces 104–2600): English, Russian, Swedish and English WikiSource

Space #	en	ru	sv	ws-en
104		Проект		Page
105		Обсуждение проекта		Page talk
106		Арбитраж		Index
107		Обсуждение арбитража		Index talk
108	Book			
109	Book talk			
114				Translation
115				Translation talk
118	Draft			
119	Draft talk			
446	Education Program	Образовательная программа	Utbildningsprogram	
447	Education Program talk	Обсуждение образовательной программы	Utbildningsprogramsdiskussion	
710	TimedText			
711	TimedText talk			
828	Module	Модуль	Modul	Module
829	Module talk	Обсуждение модуля	Moduldiskussion	Module talk
2300	Gadget	Gadget	Gadget	Gadget
2301	Gadget talk	Gadget talk	Gadget talk	Gadget talk
2302	Gadget definition	Gadget definition	Gadget definition	Gadget definition
2303	Gadget definition talk	Gadget definition talk	Gadget definition talk	Gadget definition talk
2600	Topic	Тема	Ämne	Topic

Appendix M

Wikipedia per-language localisation of ‘Disambiguation’

Wikipedia features numerous pages for disambiguation of content due to homonyms, homophones, or the splitting of large topics into smaller pages. The default mediawiki string value is the English term ‘Disambiguation’.

Wiki Language	Localised String
ceb	Maong
de	Begriffsklärung
en	Disambiguation
es	Desambiguación
fr	Homonymie
it	Disambigua
jp	曖昧さ回避
nl	doorverwijspagina
ru	неоднозначности
sv	olika betydelser

TABLE M.1: Wikipedia per-language string localisations for ‘Disambiguation’.

Appendix N

Extracts from Otlet's *Traité de Documentation*

Otlet's 1934 oeuvre *Traité de Documentation: Le Livre sur le Livre* [284] is not available in translation from the original French, but the text is available on the French wikisource at '*Traité de documentation*' [W724] (under 'CC BY-SA 3.0 Licence (French)' licence [81]). The latter was used, via Google Translate, and light post-editing for sense, to provide the excerpts shown below.

N.1 Bibliology & Biblions

Although not mentioned by Otlet, his description of bibliology as a wider form of bibliography sounds close to that described in Bianchi's 'The New Science Of Bibliology' [52] as first appearing in 1812 in *Traité de Bibliologie* ("Treatise on Bibliology") by French bibliographer Gilbert Peignot [W31][1]. As Bianchi summarises: "Whereas bibliography is concerned primarily with the classification of books and articles, bibliology aims to examine and explain the various manifestations of the book and the written word in general with regard to its production and distribution as well as to its reading. Bibliology is the study of the written word as a tool of communication within society, with all its sociological, cultural and economic implications.". This meaning is not totally aligned with current dictionary definitions, such as the Oxford English Dictionary's 'bibliology [1] entry. Otlet's meaning certainly has no connection with an alternate definition relating to biblical studies.

The most important concept enshrined in Otlet's term 'bibliology' (and Peignot's definition above) was a consideration of knowledge (then enshrined in books) as encompassing more the manner of the classification of their contents. They saw, in bibliology, the wider wider ecosystem

[1] The list of Peignot's works in the Wikipedia article does not list a title match for the work as cited by Bianchi though she might be referring to a subtitle of Peignot's *Répertoire Bibliographique Universel* of 1812.

in which the (then) book existed, and its creators. In doing so, we foresaw the present need for people to maintain the knowledge about the knowledge in a many beyond the norms of librarianship.

In Wright's 'Cataloguing The world' he describes Otlet foreseeing a new science of bibliology (as in Section N.2 below): "As opposed to bibliography—which merely tries to describe the contents of books—bibliology aspires to a higher purpose: guiding the production and distribution of of all kinds of recorded knowledge. The bibliologist would do more than simply gather information; he or she would enable access to every domain of human knowledge, create new works, and 'provoke inventions' to help support the the spread of human knowledge." [387, p.229].

The building blocks of this new knowledge would be Otlet's 'biblions' (in Section N.3 below), which he saw as the atomic particles of knowledge from which new and better resources might be created.

Otlet was very aware that at the time he wrote, knowledge was almost entirely contained within books. But his vision let him imagine a time when all the technical transformations just beginning to happen around him would bear fruit. He regretted having to wait, noting: Despite advances in scientific thought and bibliographical material, the modes of recording knowledge have not made much progress." [284, Ch.4, 411.43(d)].

Otlet did eventually plan a 'mundaneum' machine, which now makes Bush's similarly non-existent memex look modern. But that is to judge with today's eyes. Both saw a machine, howsoever fashioned, as an important means to improve humanity' engagement with its knowledge. Otlet's utopian view also seems to dovetail with Nelson's notion of the (Xanadu) docuverse.

N.2 Chapter 1. '*La Bibliologie ou Documentologie*', Section 11

Bibliology or Documentology

Book and Documentation Sciences

11 CONCEPT. DEFINITION. CHARACTERISTICS

111 Concept.

1. Book (Biblion or Document or Gram) is the conventional term used here to express all kinds of documents. It includes not only the book itself, manuscript or printed, but journals, newspapers, writings and graphic reproductions of all kinds, drawings, engravings, maps, diagrams, diagrams, photographs, etc. Documentation in the broad sense of the term includes: Book, elements used to indicate or reproduce a thought envisaged in any form.

2. The Book thus understood presents a double aspect:
a) it is first and foremost a work of man, the result of his intellectual work;

b) but, multiplied to numerous copies, it also appears as one of the many objects created by civilization and capable of acting on it; it is characteristic of any object having a bodily character and technically arranged.

112 Need for a Bibliology.

There is a common language, a common logic, a common mathematics. We must create a common bibliology: Art of writing, publishing and disseminating scientific data.

We now need not only Bibliography, the description of books, but Bibliology, that is to say a general science and technique of the document. Knowledge relating to the Book, Information, and Documentation has remained too long in the state or was Biology a century ago; there were then many unrelated sciences which, however, all had as their object living beings and life (anatomy, physiology, botany, zoology).

Biology has brought together and coordinated all these particular sciences into a general science. For the book, we now have treatises on rhetoric, library science, bibliography, printing. But we have not yet formed a Bibliology, that is to say a general science embracing the systematic classified set of data relating to the production, conservation, circulation and use of writings and documents. of all kinds. This science would lead minds to reflect more deeply on the very bases which serve as the foundation of the various particular disciplines of the book; it would make it possible to consider further progress, thanks to more general and in-depth definitions,

Efforts must therefore be made to constitute in autonomous science all the theoretical and practical knowledge relating to the Book, Bibliology. This science is called to bring out of empiricism the applications and the realizations.

We have to form huge libraries, we have to develop powerful directories. But, just as after Jussieu[2] and Linné[3], describing thousands of species, arrived, the Darwins and Claude Bernard who created Biology, an explanatory theoretical science, the evolutionary of all living beings, so are times now come where it is necessary to found Bibliology, the theoretical, comparative, genetic and abstract science, embracing all books, all species and all forms of documents. Like sociology, the synthesis of the sciences of society was formed with all the particular social sciences. As there is a general applied mechanics, independent of any particular study of the book, science of all the particular forms bibliology: bibliographic synthesis,

113 Goal.

Bibliology should aim to:

1. Analyse, generalize, classify, synthesize the data acquired in the fields of the book and at the same time promote new research intended above all to deepen the theoretical why of certain practices of experience.

[2] 'Antoine Laurent de Jussieu' [W2]. French botanist who improved upon Linneus' (q.v.) classification of plants.

[3] 'Carl Linnaeus' [W11] (later Carl von Linné). Swedish botanist and creator of the Linnean system for classification and naming of organisms

2. Develop a complete set of “documentary forms” into which data can flow. of scientific or practical thought, from the simple document to the complexes of the great collections and the elevated forms that constitute the Treatise and the Encyclopedia.

3. To make progress us that all can tend to the more methodical and more rational presentation of the data of our knowledge and practical information.

4. To provoke certain inventions which will undoubtedly be able to remain unused for a long time and without application, but which one day will perhaps be the starting point of transformations so profound that they will amount in this matter to real revolutions.

5. Bibliology elaborates the scientific and technical data relating to this quadruple object:

1. the recording of human thought and external reality in elements of a material nature called documents;
2. the conservation, circulation, use, cataloguing, description and analysis of these documents;
3. establishment with the help of simple documents, of more complex documents, and with the help of specific documents, a set of documents;
4. to the ultimate degree, the recording of data more and more complete, exact, precise, simple, direct, rapid, synoptic, in both an analytical and synthetic fashion; following an increasingly integral, encyclopedic, universal and global plan.

6. From a scientific point of view, the fundamental bibliological principle, principle-tendency of optimum publication is expressed in these four desiderata:

1. To say the whole of one thing.
2. Say everything once.
3. The truth about everything.
4. In the form most apt to be understood.

This principle is tempered in four ways:

1. What is still ignored.
2. The thesis of truth, doubt, discussion, various theses.
3. The variety of intelligences: language, degree, age, previous preparations.
4. The plurality of possible forms, of presentation according to the taste of each one, and of the possible progress in the presentation, the accessibility, the price.

114 Conditions of the constitution of Bibliology in science.

Bibliology must meet the following eight conditions which are necessary for there to be complete science:

1. A general or special object (beings, entities, facts).
2. A specific point of view or separate intellectual object to consider these facts and coordinate them.
3. Generalization, general facts, fundamental concepts, laws.
4. Systematization , coordinated results, classification.
5. Method: with what it includes: a) research methods, logical or reasoning procedures; b) classification, terminology; c) system of measurements; d) instruments; e) recording and conservation of acquired data (Sources, Bibliography).
6. Organization of work (division of labour, cooperation, national and international organizations, associations, commissions, congresses, institutes covering the functions of research, discussion, decision on methods, teaching and dissemination).
7. History.
8. Application of the various orders of study and activity.

Purpose of the Bibliology.

1. What is its own in the Book, what is properly bibliographical? We have already said the distinction between:
a) Objective Reality,
b) Subjective Thought or the state of consciousness provoked or the ego by reality,
c) Objective Thought which is the effort of combined and collective reflection on these first data up to impersonal and total science,
d) Language, collective instrument of the expression of Thought.

Collected as a whole, any book contains these four elements associated concretely in itself and that by abstraction only, it is possible to dissociate and study separately. What is specific to the book is the fifth element: thought now fixed by the writing of words or the image of things, visible signs, fixed on a material support.

2. Hence these three consequences:
a) Reality, objective or subjective Thought, Language each have an anterior existence independent of the book. They study in their respective knowledge (Psychology-Science-Linguistics).
b) On the contrary, signs and supports are characteristic of the book and it is a question in the bibliological sciences of studying them in all aspects.

c) But in turn the concrete complex of ideas, words, images as embodied in the book and the document ("biblified" or "documentised") are, with respect to each other, in the position of content and container. Their relations, inter-influences, repercussions, are to be examined, in their turn, and this is a domain common to the sciences of a and those of b.

3. It is necessary to continue studies and realizations of Documentation within the general framework of the body of knowledge and activities, establishing correlations:
a) with the various sciences;
b) with the various techniques and their subject (Universal Science, General Technique);
c) with the various organizational plans (Global Plan).

4. To define Bibliology is to characterize the domain over which this science extends its empire and at the same time indicate the limits which separate this domain from its neighbours.

5. The object of a science should be distinguished from the science of this object. Science is the organization of the knowledge of an object. The object exists outside the knowledge that we have. Geological science, for example, is a recent creation, when the earth existed before. There were books long before there were bibliological sciences.

6. Bibliology has a universal encyclopedic character, due to the fact that documents (its object) refer to the whole of all Things.

Bibliology participates in the same generality as Logic and Linguistics: everything is susceptible, at the same time, of expression, of documentation. Logic, said the Logicians, is a general science in the sense that it regulates the content of all the others and that all must be constituted according to its laws. Its object of extreme simplicity and unlimited extension is the being of reason. Bibliology, in so far as it considers the conditions of the best book done or to be made, does not regulate thought for itself. However, its influence is great on each thought, because, more and more, each one tends to express himself, to communicate to others, to question them, to answer them in a documentary form. Now such a form can either alter or exalt thought itself. Consequently, Bibliology must be considered as a general science, auxiliary to all the others and which imposes its standards on them as soon as they have to cast their results in the form of a “document”. The object of Bibliology, like that of Logic, is of extreme simplicity and unlimited extension. It is here the “being documented”, as the object of Logic is the “being of reason”.

7. The point of view specific to General Bibliology is that of the Book considered as a whole, of the totality of Books. Just as Sociology is concerned not with phenomena which take place in society, but with phenomena which react socially, so Bibliology is concerned with facts which have a general effect on the Book.

8. The proper domain of Bibliology must be determined and explored. In the broad sense, it includes the History of Literature and Criticism. But alongside the history of Books and that of authors, there is at the same time the History of thought

116 Basis.

There is a reality made up of the total and that is what it is. Within this reality, we see Man, Men and their Society at work within Nature. In man, observation if not definition and explanation, we are led to distinguish two elements:

1. the deep, personal, lived self; free qualitative mobility in the long term foreign to him; pure memory plunging into the indivisible movement of the vital impulse;
2. the intelligent self, with practical functions, with a deterministic mechanism. The two elements coexist, producing all the works with their two methods, intuition and direct knowledge for one; logic and discursive knowledge for the other. These two elements are found in the individual, in the life of society (thought, feeling, activity) and we find them in the books which are the manifestation or expression of them.

Intelligence, by competing for it with instinct, by proceeding from the conscious to the unconscious, has become clear, communicative, demonstrative, cooperative in two great creations which are largely specific to it, which are social: systematic Science and Coordinated Civilization. The book is par excellence the work of intelligence, but not exclusively, because Intuition (Instinct, feeling) also plays a large part in it. A bifurcation has been determined among the kinds of books along two major divergent lines: the Book of Science and Reasoned Practice; the Book of Literature which ranges from simple spontaneous notation to written and graphic fixations of the highest mysticism.

N.3 Chapter 1. '*Le Livre et le Document*', Section 211.3

3. **The Biblion**.

There is now a generic term (Biblion or Bibliogramme or Document) which covers all species at the same time: volumes, brochures, reviews, articles, maps, diagrams, photographs, prints, patents, statistics, even phonographic records, glasses or motion pictures.

The "Biblion" will be for us the intellectual and abstract unity but which we can find concretely and really but covered in various modalities. The biblion is conceived in the manner of the atom (ion) in physics, of the cell in Biology, of the mind in psychology, of the human aggregation (the socion) in sociology. The atom has given rise to an increasingly precise representation and on the basis of which all research and discussions have started. (It was Bohr who gave the first figure.)

a) In the cosmos (set of things) the book or Document takes place among bodily (non-incorporeal), artificial (non-natural), and intellectual (non-material) things. Material creations are either productions or means of producing. There are: a) the means of producing useful and consumable things (machines); b) the means to produce natural phenomena, disregarding any concern for

utility (devices); c) the means of measuring the phenomena (the instruments). The Book is a means of producing intellectual utilities.

b) Things have various kinds of relations with documents:

1. Relation of signifying things to insignificant things, which constitutes the very foundation of documentation.
2. The things themselves treated as object of documentation as for specimen and sample they appear in documentary collections (museums, exhibitions).
3. The things created, models and mechanisms for scientific, educational or advertising demonstration.
4. Marks of any kind worn on objects and which serve for their identification and signalling.
5. The application by analogy of the methods of documentation to the administration of things themselves (administrative documentation).

c) The writings have the property called by the adage "*scripta manent verba volant*", i.e. the writings remain if the words fly away. But from the point of view of the rigour of the thought one can propose in Latin terms this other adage, the terms ranging in degree:

- *verba divagantur*. The word can ramble. In other words, Gone with the Wind. Speech being successive can be translated without being anything other than a series of points, the material link of which, simply sound, is so light that it can float in all directions.
- *scripta concentrant*. The writings concentrate the thought of those who establish them. They're on the surface. We read them, being able to go back in the text from front to back. The logical links of truth if they are not real can easily be detected.
- *constructiones coordinant*. The constructions, stereograms in three dimensions, strictly coordinate the ideas. By the voids and the overloads, by the three directions of the idea which must be concordant, which allow an easy control, it is already more difficult to venture into superficial and badly studied developments.
- *mechanica logicant*. Finally, machines are logicians par excellence. They can only enter into motion and remain there through the rigorously exact, concordant and simultaneous play of all their parts.

d) The Document offers an image of Reality at the sixth derivation. We have in fact the following intermediate terms:

1. The World (or Reality itself);
2. The senses of man who perceive the world exactly and completely;

3. Intelligence, which develops sensory data;
4. Language, a social instrument of communication;
5. Science, or collective knowledge;
6. The Document composed by Intelligence and to express Science.

Each of its intermediaries is a cause of deformations and frictions absorbing intellectual energy.

Every effort must therefore be made: a) to eliminate or attenuate intermediate deformations and friction; b) to create means of perceiving or representing reality.

Appendix O

Tools & Colophon

O.1 Research Tools

Hardware. Analysis of data was performed across 3 Mac computers. This multi-computer use occurred simply due to working in a number of discrete locations. Macs used: MacBook Pro (mid 2011), used as a home desktop); MacBook Air (2013), mobile working; Mac Mini (2015), in the University's lab. All Macs were running same current macOS 10.x.x version, 8Mb RAM. A single 2TB USB3 external drive was used for additional data storage (manually) shared/used with all three Macs. All data extraction scripts were run on the MacBook Pro or MacBook Air depending on the location at the time, with no difference in outcome other than slight differences in performance due to underlying hardware.

Software. The following tools/environments were used for analysis/evaluation, all running on Mac OS X:

- *Python v3.4.4*, run via macOS Terminal. The following non-default additional modules were used:
  - `lxml`: specifically, the `etree` method for iterative parsing of nested elements within large XML files.
  - `numpy`: statistical functions.
  - `scipy`: statistical functions.
- *R* and *RStudio*: analysis for, and creation of, some diagrams for this thesis.
- *Sublime Text*: Python script coding.
- *BBEdit*: inspection and cleaning of data from script-exported files.

- *Tinderbox* v6.x-v8.x: working notes; analysis/illustration of transclusion use and pathways; analysis and coding of bot account data. Tinderbox was used for most of the analysis of extracted data as it has strong support for prototypes and incremental formalisation allowing for open-ended exploration of the data.
- *Excel 2016/Excel 365*: storage and manipulation of tabular result data.
- macOS 'bash' shell command line tools: inspection and cleaning of data from script-exported files, decompressing bzip2 archive files.
- *Bookends*: reference management.

O.2 Thesis Production

This thesis was written and produced using the following tools:

- Production method: *LaTeX*, using *BibTeX* (with *biber* back-end) for reference ingest.
- Writing environment: *TeXstudio* (initially *TeXworks*).
- Reference management: *Bookends*. Data exported in UTF-8 BibTeX format.
- BibTex file triage: *BBEdit*.
- SVG chart generation: *R* and *RStudio*.
- Bitmap image editing: *GraphicConverter*.

O.2.1 Spelling

This thesis is written in (British) English, 'BrE', which differs from American English ('AmE') in some spellings: -our word endings in BrE would be -or in AmE, thus humour/humor. AmE spelling has been retained in the titles of references and in quotes. By comparison actual errors are noted [sic] or, in a trivial cases corrected (but not within direct quotes).

O.2.2 Hypertextual Features

Note: this section contains additional notes in relation to section 1.6.

In addition to a bibliography of normal academic literary references, this thesis contains a large number of references to web pages within Wikipedia and Wikimedia Foundation related domains. These pertain mainly to the research in reported in Chapters 4 and 5. Wikipedia's wiki software records each edit with a discrete Unique Identifier (UID). Therefore, where possible and

pertinent, the Uniform Resource Locators (URLs) given are for the Wikipedia pages as would have been displayed at the time the data was sampled[1].

As many of the URLs are extremely long they make the main bibliography unwieldy. Likewise printing them in full within the main text or footnotes would render it very hard to read. Therefore the references are split into two discrete sections. Separate glossary and acronym lists are supplied.

- Blue text implies one of various types of *internal* links within the document:
  - Items in The Table of Contents, List of Figures and List of Tables all link those locations in the text.
  - Headings within the document text are also link back to the appropriate part of the Table of Contents.
  - Hyperlinks to Glossary items. All glossary items also contain a back link to marked instances of that item in the document.
  - Ad hoc in-document hyperlinks. These provide web-like cross-referencing links; back-links are only present if deliberately added.
- Green text indicates an *internal* link to a reference in one of the two bibliographies. Clicking these will move the PDF viewer to the reference in the bibliography. The listing will include a list of (blue) back-links to all instances of the reference.
  - The main reference list uses numbered references in this style: '[1]'.
  - Items from the Wikipedia reference list use the same style but with a 'W' prefix: '[W1]'.
- Red text denotes a web link to an *external* resource. Clicking any of these will open the associated URL in a web browser.

For Web-based references, where practical and pertinent, the text contains both a (red) web link and a (green) link to the bibliography. When using the digital version of the thesis, viewing the bibliography will provide a textual version of the URL without needing to open it.

[1] Due to the live nature of transclusion, transcluded material is always from the current revision even if the containing page viewed is for a past revision. This is done by appending the UID of the active revision desired, allowing the reader to see the page version being referenced. Some pages, such as 'special' pages which are used for reporting status, have no cached data and the data shown will be for access date so the URL cannot be used as a reliable record of note.

Bibliography

General References

This section contains the main bibliography. A separate bibliography follows containing only Wikipedia references, due to the large number of them that are cited.

[1] Oxford English Dictionary (online). *Bibliology, n., a.* 1807. URL: https://www.oed.com/view/Entry/18634 (visited on 01/08/2019) (cit. on p. 311).

[2] Oxford English Dictionary (online). *Curator, n., II.3.* 1893. URL: http://www.oed.com/view/Entry/45960#eid7532846 (visited on 01/08/2019) (cit. on p. 8).

[3] Oxford English Dictionary (online). *Fettle, v. tr., 1a.* 1895. URL: http://www.oed.com/view/Entry/69637#eid4381087 (visited on 01/08/2019) (cit. on p. 8).

[4] *'Distracted Boyfriend' meme*. knowyourmeme.com. 2019. URL: https://knowyourmeme.com/memes/distracted-boyfriend (visited on 30/01/2019) (cit. on p. 6).

[5] *'Linking Open Data' Project*. www.w3.org. 2019. URL: https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData (visited on 30/01/2019) (cit. on p. 48).

[6] *A Half-Century of Hypertext at Brown—Session 1 (@1:34:40)*. Youtube. 2019. URL: https://youtu.be/ZKw5sEVmGgM?t=5677 (visited on 30/01/2019) (cit. on p. 238).

[7] *A Half-Century of Hypertext at Brown: Session 1, at c.1:53:29, FRESS*. Youtube.com. 2019. URL: https://youtu.be/ZKw5sEVmGgM?t=3253 (visited on 30/01/2019) (cit. on p. 235).

[8] *A Half-Century of Hypertext at Brown: Session 1, at c.33:16, HES*. Youtube.com. 2019. URL: https://youtu.be/ZKw5sEVmGgM?t=3253 (visited on 30/01/2019) (cit. on p. 229).

[9] *A Half-Century of Hypertext at Brown: Session 3, at c.41:226*. Youtube.com. 2019. URL: https://youtu.be/Nz_U0bBozLs?list=PLTiEffrOcz_52mjCMy65cWt9Ik26rI1xx&t=2486 (visited on 30/01/2019) (cit. on p. 205).

[10] *A Partial Implementation of Nelson's Zigzag® ideas for Web Browsers*. Southampton University. 2001. URL: http://www.southampton.ac.uk/~lac/zigzag/ (visited on 30/01/2019) (cit. on p. 30).

[11] *Active Navigation*. Active Navigation Limited. 2019. URL: https://activenavigation.com (visited on 30/01/2019) (cit. on p. 243).

[12] AdaptivePath.org. *Ajax: A New Approach to Web Applications*. 2016. URL: http://adaptivepath.org/ideas/ajax-new-approach-web-applications/ (visited on 07/04/2017) (cit. on p. 43).

[13] Robert M. Akscyn and Donald L. McCracken. 'ZOG and the USS Carl Vinson'. In: *Computer Science Research Review*. London: Interaction Design Foundation, 1984, pp. 901–906 (cit. on p. 237).

[14] Robert M. Akscyn, Donald L. McCracken and Elise Yoder. 'KMS: A Distributed Hypermedia System for Managing Knowledge in Organizations'. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 1–20. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317428 (cit. on pp. 239, 398).

[15] Christopher Alexander. *A Pattern Language: Towns, Buildings, Construction*. OUP USA, 1978, p. 1216. ISBN: 0195019199 (cit. on pp. 42, 252).

[16] *Amazon's 'Mechanical Turk'*. Amazon.com. 2019. URL: https://www.mturk.com (visited on 30/01/2019) (cit. on p. 49).

[17] Aubrey Anable. 'The Architecture Machine Group's Aspen Movie Map: Mediating the Urban Crisis in the 1970s'. In: *Television & New Media* 13.6 (2012), pp. 498–519. DOI: 10.1177/1527476411423673 (cit. on p. 237).

[18] G. L. Anderson. 'The McBee Keysort System for Mechanically Sorting Folklore Data'. In: *The Journal of American Folklore* 66.262 (1953), pp. 340–343. DOI: 10.2307/53673 (cit. on p. 20).

[19] Mark W. R. Anderson. *The Itinerary of HM Queen Elizabeth II*. 2012. URL: http://www.shoantel.com/proj/maps2/index.html (visited on 01/08/2019) (cit. on p. 10).

[20] Mark W. R. Anderson. *A Tinderbox Reference File*. 2015. URL: http://www.acrobatfaq.com/atbref8/index.html (visited on 01/08/2019) (cit. on p. 10).

[21] Mark W. R. Anderson, Les A. Carr and David E. Millard. 'There and Here: Patterns of Content Transclusion in Wikipedia'. In: vol. Proceedings of the 28th ACM Conference on Hypertext and Social Media. Prague, Czech Republic: ACM, 2017, pp. 115–124. ISBN: 978-1-4503-4708-2. DOI: 10.1145/3078714.3078726 (cit. on pp. xvii, 31).

[22] Marc Andreessen. *Proposed New Tag: IMG*. 1992. URL: http://1997.webhistory.org/www.lists/www-talk.1993q1/0182.html (visited on 01/08/2019) (cit. on p. 43).

[23] Keith Andrews, Frank Kappe and Hermann A. Maurer. 'The Hyper-G Network Information System'. In: *Special Issue: Selected Proceedings of the Workshop on Distributed Multimedia Systems*. Ed. by Hermann A. Maurer, Cristian Calude and Arto Salomaa. Berlin Heidelberg: Springer, 1996, pp. 206–220. ISBN: 978-3-642-80352-9. DOI: 10.1007/978-3-642-80350-5_20 (cit. on p. 244).

[24] Sally A. Applin and Michael D. Fischer. 'Exploring Cooperation with Social Machines'. In: vol. Proceedings of the 25th International Conference Companion on World Wide Web. Montréal, Québec, Canada Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 765–768. ISBN: 978-1-4503-4144-8. DOI: 10.1145/2872518.2890591 (cit. on p. 51).

[25] Bill Atkinson. *Triangulation 247: Bill Atkinson, part 2*. 2018. URL: https://youtu.be/INdByDjhClU?t=1341 (visited on 01/08/2019) (cit. on p. 240).

[26] Claus Atzenbeck, Daniel Roßner and Manolis Tzagarakis. 'Mother: An Integrated Approach to Hypertext Domains'. In: vol. Proceedings of the 29th on Hypertext and Social Media. Baltimore, MD, USA New York, NY, USA: ACM, 2018, pp. 145–149. ISBN: 978-1-4503-5427-1. DOI: 10.1145/3209542.3209570 (cit. on pp. 245, 253).

[27] Claus Atzenbeck et al. 'Revisiting Hypertext Infrastructure'. In: vol. Proceedings of the 28th ACM Conference on Hypertext and Social Media. Prague, Czech Republic New York, NY, USA: ACM, 2017, pp. 35–44. ISBN: 978-1-4503-4708-2. DOI: 10.1145/3078714.3078718 (cit. on p. 253).

[28] Phoebe Ayers, Charles Matthews and Ben Yates. *How Wikipedia Works: And How You Can Be a Part of It*. No Starch Press, 2008, p. 536. ISBN: 159327176X (cit. on p. 272).

[29] Thierry Bardini. *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Stanford University Press, 2000, p. 320. ISBN: 0804738718 (cit. on pp. 25–26, 398).

[30] Belinda Barnet. *Memory Machines: The Evolution of Hypertext*. London, UK: Anthem Press, 2013, p. 192. ISBN: 978-1-78308-344-2 (cit. on pp. 25, 28, 41, 234–236, 241–243, 246, 251–252).

[31] Daniel J. Barrett. *MediaWiki: Wikipedia and Beyond*. O'Reilly Media, 2008, p. 380. ISBN: 9780596554149 (cit. on p. 271).

[32] Roland Barthes and Richard Howard. *S/Z*. Farrar, Straus and Giroux, 1974, p. 271. ISBN: 9780374521677 (cit. on p. 17).

[33] Steven A. Battle and Matthew Bernius. 'Transquotation in EBooks'. In: vol. Proceedings of the 10th ACM Symposium on Document Engineering. Manchester, United Kingdom New York, NY, USA: ACM, 2010, pp. 69–72. ISBN: 978-1-4503-0231-9. DOI: 10.1145/1860559.1860575 (cit. on p. 247).

[34] David Bawden. 'Bush, Goldberg, Memex and the revision of history'. In: *Journal of Documentation* 66.4 (2010). DOI: 10.1108/jd.2010.27866daa.001 (cit. on p. 24).

[35] Tim Berners-Lee and Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. HarperBusiness, 2000, p. 256. ISBN: 006251587X (cit. on pp. 47, 185, 202, 248).

[36] Tim Berners-Lee, James A. Hendler and Ora Lassila. 'The Semantic Web'. In: *Scientific American* 284.5 (2001), pp. 34–43 (cit. on p. 47).

[37] Tim Berners-Lee and Kieron O'Hara. 'The read–write Linked Data Web'. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013), 5pp. DOI: 10.1098/rsta.2012.0513 (cit. on p. 48).

[38] Larry Bernstein and F. E. Slojkowski. 'An Information System for the Coordination of Program Design'. In: vol. Proceedings of the 1966 21st National Conference. New York, NY, USA: ACM, 1966, pp. 87–96. DOI: 10.1145/800256.810685 (cit. on p. 29).

[39] Mark Bernstein. 'The Bookmark and the Compass: Orientation Tools for Hypertext Users'. In: *SIGOIS Bulletin* 9.4 (1988), pp. 34–45. DOI: 10.1145/51640.51645 (cit. on p. 242).

[40] Mark Bernstein. 'Web Research: The Eastgate Web Squirrel'. In: *SIGWEB Newsletter* 5.1 (1996), pp. 6–6. DOI: 10.1145/232782.232784 (cit. on p. 250).

[41] Mark Bernstein. 'Patterns of Hypertext'. In: vol. Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia : Links, Objects, Time and Space—Structure in Hypermedia Systems : Links, Objects, Time and Space—Structure in Hypermedia Systems. New York, New York, USA: ACM Press, 1998, pp. 21–29. ISBN: 0897919726. DOI: 10.1145/276627.276630 (cit. on p. 252).

[42] Mark Bernstein. 'Storyspace 1'. In: vol. Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia. ACM, 2002, pp. 172–181. ISBN: 1581134770. DOI: 10.1145/513338.513383 (cit. on pp. 241, 246, 250).

[43] Mark Bernstein. *Tinderbox*. 2002. URL: http://www.eastgate.com/Tinderbox/ (visited on 01/08/2019) (cit. on pp. 72, 253).

[44] Mark Bernstein. 'Collage, Composites, Construction'. In: vol. Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia. ACM, 2003, pp. 122–123. ISBN: 1581137044. DOI: 10.1145/900051.900077 (cit. on p. 251).

[45] Mark Bernstein. 'On Hypertext Narrative'. In: vol. Proceedings of the 20th ACM Conference on Hypertext and Hypermedia. ACM, 2009, pp. 5–14. ISBN: 160558486X. DOI: 10.1145/1557914.1557920 (cit. on p. 247).

[46] Mark Bernstein. 'Shadows in The Cave: Hypertext Transformations'. In: *Journal of Digital Information (JoDI)* 10.3 (2009), pp. 1–8 (cit. on p. 251).

[47] Mark Bernstein. *The Tinderbox Way*. Watertown, MA, USA: Eastgate Systems Inc., 2009, p. 215. ISBN: 1884511465 (cit. on p. 251).

[48] Mark Bernstein. 'Can We Talk About Spatial Hypertext?' In: vol. Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia. ACM, 2011, pp. 103–112. ISBN: 1450302564. DOI: 10.1145/1995966.1995983 (cit. on pp. 252–253).

[49] Mark Bernstein. 'Storyspace 3'. In: vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada New York, NY, USA: ACM, 2016, pp. 201–206. ISBN: 978-1-4503-4247-6. DOI: 10.1145/2914586.2914624 (cit. on pp. 241, 246, 252–253).

[50] Mark Bernstein and Diane Greco. *Reading Hypertext*. Watertown, MA, USA: Eastgate Systems Inc., 2009, p. 271. ISBN: 1884511481 (cit. on pp. 243, 246–247).

[51] Mark Bernstein, David E. Millard and Mark J. Weal. 'On Writing Sculptural Hypertext'. In: vol. Proceedings of the Thirteenth ACM Conference on Hypertext and Hypermedia. ACM, 2002, pp. 65–66. ISBN: 1581134770. DOI: 10.1145/513338.513355 (cit. on pp. 247, 252).

[52] Anne-Marie Bianchi. 'The New Science Of Bibliology'. In: *UNESCO Courier* 2 (1985), pp. 33–33. ISSN: 2220-2285 (cit. on p. 311).

[53] Christian Bizer, Tom Heath and Tim Berners-Lee. 'Linked Data: The Story So Far'. In: *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. Ed. by Amit Sheth. IGI Global, 2011, pp. 205–277. ISBN: 9781609605933. DOI: 10.4018/978-1-60960-593-3.ch008 (cit. on p. 48).

[54] Rebecca Blood. *Weblog: A History and Perspective*. 2000. URL: http://www.rebeccablood.net/essays/weblog_history.html (visited on 01/08/2019) (cit. on p. 39).

[55] Jay David Bolter. *Writing Space*. Vol. 2nd ed. Routledge, 2001, p. 246. ISBN: 9780805829181 (cit. on pp. 7, 246).

[56] Jay David Bolter and Michael Joyce. 'Hypertext and Creative Writing'. In: vol. Proceedings of the ACM Conference on Hypertext. ACM, 1987, pp. 41–50. ISBN: 089791340X. DOI: 10.1145/317426.317431 (cit. on p. 241).

[57] Jorge Luis Borges. *The Garden of Forking Paths*. Penguin Classics, 1941, 17pp. ISBN: 9780241339053 (cit. on p. 21).

[58] *Bots vs. Wikipedians—Anons vs. Logged-Ins—Bot Counter*. Thomas Steiner/Google Germany. 2019. URL: http://wikipedia-edits.herokuapp.com (visited on 30/01/2019) (cit. on p. 200).

[59] Philip Boulain, Nigel R. Shadbolt and Nicholas M. Gibbins. 'Studies on Editing Patterns in Large-Scale Wikis'. In: *Weaving Services and People on the World Wide Web*. Ed. by Irwin King and Ricardo Baeza-Yates. Springer, 2009, pp. 325–349. ISBN: 978-3-642-00570-1. DOI: 10.1007/978-3-642-00570-1_16 (cit. on p. 9).

[60] Philip Boulain, Nigel R. Shadbolt and Nicholas M. Gibbins. 'A Model for Open Semantic Hyperwikis'. In: vol. Recent Trends and Developments in Social Software, vol 6045. Springer, 2010, pp. 1–14. ISBN: 978-3-642-16580-1. DOI: 10.1007/978-3-642-16581-8_1 (cit. on p. 48).

[61] Niels Olof Bouvin and Clemens Nylandsted Klokmose. 'Classical Hypermedia Virtues on the Web with Webstrates'. In: vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada New York, NY, USA: ACM, 2016, pp. 207–212. ISBN: 978-1-4503-4247-6. DOI: 10.1145/2914586.2914622 (cit. on p. 44).

[62] John Broughton. *Wikipedia: The Missing Manual*. O'Reilly Media, 2008, p. 512. ISBN: 0596515162 (cit. on p. 272).

[63] Meredith Broussard. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, 2018, p. 248. ISBN: 0262038005 (cit. on p. 9).

[64] John Seely Brown and Paul Duguid. 'Organizing Knowledge'. In: *California Management Review* 40.3 (1998), pp. 90–111. DOI: 10.2307/41165945 (cit. on p. 8).

[65] Peter J. Brown. 'A Simple Mechanism for Authorship of Dynamic Documents'. In: vol. Text Processing and Document Manipulation: Proceedings of the International Conference, University of Nottingham, 14-16 April 1986. Cambridge, UK: Cambridge University Press, 1986, pp. 34–42. ISBN: 9780521110310 (cit. on p. 237).

[66] Peter J. Brown. ‘Turning Ideas into Products: The Guide System’. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 33–40. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317430 (cit. on pp. 236–237).

[67] Susan L. Bryant, Andrea Forte and Amy Bruckman. ‘Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia’. In: vol. Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work. Sanibel Island, Florida, USA New York, NY, USA: ACM, 2005, pp. 1–10. ISBN: 1-59593-223-2. DOI: 10.1145/1099203.1099205 (cit. on p. 49).

[68] Werner Buchholz. ‘The System Design of the IBM Type 701 Computer’. In: *Proceedings of the IRE* 41.10 (1953), pp. 1262–1275. DOI: 10.1109/JRPROC.1953.274300 (cit. on p. 26).

[69] Michael Keeble Buckland. *Emanuel Goldberg and His Knowledge Machine*. Greenwood Publishing Group, 2006, p. 380. ISBN: 9780313313325 (cit. on pp. 20, 24–25).

[70] Roberto Busa. ‘The Annals of Humanities Computing: The Index Thomisticus’. In: *Computers and the Humanities* 14.2 (1980), pp. 83–90. DOI: 10.1007/BF02403798 (cit. on p. 21).

[71] Vannevar Bush. ‘As We May Think’. In: *The Atlantic Monthly* 176.1 (1945), pp. 101–108 (cit. on pp. 2, 10, 22–24, 37, 247).

[72] Vannevar Bush. ‘Memex II’. In: *From Memex To Hypertext: Vannevar Bush and the Mind’s Machine*. Ed. by James M. Nyce and Paul Kahn. Original manuscript dated 1959. Academic Press, 1992, pp. 165–184. ISBN: 9780125232708 (cit. on p. 25).

[73] Vannevar Bush. ‘Memex Revisited’. In: *From Memex To Hypertext: Vannevar Bush and the Mind’s Machine*. Ed. by James M. Nyce and Paul Kahn. Originally published 1967. Academic Press, 1992, pp. 197–216. ISBN: 9780125232708 (cit. on p. 25).

[74] bzip.org. *bzip2*. 2016. URL: http://bzip.org (visited on 01/08/2019) (cit. on p. 68).

[75] *Cabinet Office - GOV.UK*. Gov.uk. 2019. URL: https://www.gov.uk/government/organisations/cabinet-office (visited on 30/07/2019) (cit. on p. 398).

[76] *Card Sharks and Holy Scrollers*. Mark Bernstein. 2000. URL: https://www.markbernstein.org/Oct10/CardSharksandHolyScrollers.html (visited on 30/01/2019) (cit. on p. 33).

[77] Steven Carmody et al. ‘A Hypertext Editing System for the /360’. In: vol. Pertinent Concepts in Computer Graphics. Proceedings of The Second University of Illonois Conference on Computer Graphics, Vol 2. University of Illinois Press, 1969, pp. 291–330 (cit. on p. 234).

[78] Les A. Carr, Wendy Hall and David C. De Roure. ‘The Evolution of Hypertext Link Services’. In: *ACM Computing Surveys (CSUR)* 31.4es (1999), Article 9. DOI: 10.1145/345966.345989 (cit. on p. 248).

[79] Les A. Carr et al. ‘The Microcosm Link Service and its Application to the World Wide Web’. In: vol. Proceedings of the First WWW Conference. Geneva, Switzerland, 1994, pp. 25–34 (cit. on p. 243).

[80] Robert S. Casey and James Whitney Perry. *Punched Cards: Their Applications to Science and Industry*. New York, NY: Reinhold Publushing Corporation, 1951 (cit. on p. 19).

[81] *CC BY-SA 3.0 Licence (French)*. Creativecommons.org. 2019. URL: https://creativecommons.org/licenses/by-sa/3.0/deed.fr (visited on 30/01/2019) (cit. on p. 311).

[82] David Chapman. *There’s a chicken-and-egg issue: no one understands what a hypertext book is, because there are so few of them; there are few because the tools are inadequate; there’s no demand for better tools because no one wants to write hypertext books, because no one can read them.*... Twitter.com, July 2019. URL: https://twitter.com/Meaningness/status/1154827701754007552 (visited on 30/08/2019) (cit. on p. 4).

[83] EunJeong Cheon and Norman Makoto Su. ‘Configuring the User: “Robots Have Needs Too”’. In: vol. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Portland, Oregon, USA New York, NY, USA: ACM, 2017, pp. 191–206. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998329 (cit. on p. 197).

[84] Maxime Clément and Matthieu J. Guitton. 'Interacting with Bots Online: User's Reactions to Actions of Automated Programs in Wikipedia'. In: *Computers in Human Behavior* 50 (2015), pp. 66–75. DOI: 10.1016/j.chb.2015.03.078 (cit. on p. 199).

[85] George H. Collier. 'Thoth-II: Hypertext with Explicit Semantics'. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 269–289. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317446 (cit. on pp. 241–242).

[86] Johann Amos Comenius. *Orbis Sensualium Pictus*. 1659 (cit. on p. 19).

[87] Creative Commons. *Creative Commons Attribution-ShareAlike Licence (CC-BY-SA) v3.0*. 2019. URL: https://creativecommons.org/licenses/by-sa/3.0/ (visited on 01/08/2019) (cit. on p. 397).

[88] Jeff Conklin. 'Hypertext: An Introduction and Survey'. In: *IEEE Computer* 20.9 (1987), pp. 17–41 (cit. on p. 34).

[89] Jeff Conklin and Michael L. Begeman. 'gIBIS: A Hypertext Tool for Exploratory Policy Discussion'. In: *ACM Transactions on Information Systems (TOIS)* 6.4 (1988), pp. 303–331. DOI: 10.1145/58566.59297 (cit. on pp. 244, 249–250).

[90] William Wallace Cook. *Plotto*. Battle Creek, MI, USA: Ellis Publishing Company, 1928, p. 300 (cit. on p. 20).

[91] Ward Cunningham. *C2 wiki: OOPSLA '92 Experience Report*. Cunningham & Cunningham, Inc. 2019. URL: http://c2.com/doc/oopsla92.html (visited on 30/01/2019) (cit. on p. 399).

[92] Cunningham & Cunningham, Inc. *WelcomeVisitors*. 2014. URL: http://wiki.c2.com/?WelcomeVisitors (visited on 01/08/2019) (cit. on p. 40).

[93] Cunningham & Cunningham, Inc. *Wiki Design Principles*. 2014. URL: http://c2.com/cgi/wiki?WikiDesignPrinciples (visited on 01/08/2019) (cit. on p. 40).

[94] Cunningham & Cunningham, Inc. *Wiki Wiki Hypercard*. 2014. URL: http://wiki.c2.com/?WikiWikiHypercard (visited on 01/08/2019) (cit. on p. 40).

[95] Andrew Dalby. *The World And Wikipedia: How We Are Editing Reality*. Siduri Books, 2009, p. 256. ISBN: 9780956205209 (cit. on pp. 9, 195, 273).

[96] Andries van Dam. 'Hypertext '87: Keynote Address'. In: *Communications of the ACM* 31.7 (1988), pp. 887–895. DOI: 10.1145/48511.48519 (cit. on pp. 35, 238).

[97] Anil Dash. *The Missing Building Blocks of the Web*. Medium.com. 2018. URL: https://medium.com/@anildash/the-missing-building-blocks-of-the-web-3fa490ae5cbc (visited on 30/01/2019) (cit. on p. 3).

[98] Paul De Bra, Geert-Jan Houben and Hongjing Wu. 'AHAM: A Dexter-Based Reference Model for Adaptive Hypermedia'. In: vol. Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia: Returning to Our Diverse Roots: Returning to Our Diverse Roots. New York, NY, USA Darmstadt, Germany: Association for Computing Machinery, 1999, pp. 147–156. ISBN: 1581130643. DOI: 10.1145/294469.294508 (cit. on p. 245).

[99] Paul M. E. De Bra et al. 'AHA! Meets AHAM'. In: vol. Adaptive Hypermedia and Adaptive Web-Based Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 388–391. ISBN: 978-3-540-47952-9. DOI: 10.1007/3-540-47952-X_44 (cit. on p. 245).

[100] Paul M. E. De Bra et al. 'AHA! The Adaptive Hypermedia Architecture'. In: vol. Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia. ACM, 2003, pp. 81–84. ISBN: 1581137044. DOI: 10.1145/900051.900068 (cit. on p. 245).

[101] David C. De Roure et al. 'Observing Social Machines Part 1: What to Observe?' In: vol. Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil New York, NY, USA: ACM, 2013, pp. 901–904. ISBN: 978-1-4503-2038-2. DOI: 10.1145/2487788.2488077 (cit. on p. 49).

[102] David C. De Roure et al. ‘Observing Social Machines Part 2: How to Observe?’ In: vol. Proceedings of the ACM Web Science Conference. Oxford, United Kingdom New York, NY, USA: ACM, 2015, 13:1–13:5. ISBN: 978-1-4503-3672-7. DOI: 10.1145/2786451.2786475 (cit. on p. 50).

[103] Norman Delisle and Mayer Schwartz. ‘Neptune: A Hypertext System for CAD Applications’. In: vol. Proceedings of the 1986 ACM SIGMOD International Conference on Management of Data. Washington, D.C., USA New York, NY, USA: ACM, 1986, pp. 132–143. ISBN: 0-89791-191-1. DOI: 10.1145/16894.16867 (cit. on p. 240).

[104] Steven J. DeRose and Andries van Dam. ‘Document Structure and Markup in the FRESS Hypertext System’. In: *Markup Language* 1 (1999), pp. 7–32. DOI: 10.1162/109966299751940814 (cit. on pp. 19, 236).

[105] Angelo Di Iorio and John Lumley. ‘From XML Inclusions to XML Transclusions’. In: vol. Proceedings of the 20th ACM Conference on Hypertext and Hypermedia. Torino, Italy New York, NY, USA: ACM, 2009, pp. 147–156. ISBN: 978-1-60558-486-7. DOI: 10.1145/1557914.1557942 (cit. on pp. 43–44).

[106] *Donald Rumsfeld, Defence Department Briefing. 12 February 2002: 37:30–39:00*. C-SPAN. 2002. URL: https://www.c-span.org/video/?c4813731/unknown-unknowns (visited on 30/01/2019) (cit. on p. 399).

[107] Bob Doyle. ‘Wikiful Thinking’. In: *EContent* 37.1 (2008), p. 37 (cit. on p. 200).

[108] Eastgate Systems, Inc. *Tinderbox Backstage*. 2018. URL: http://backstage.eastgate.com (visited on 01/08/2019) (cit. on p. 10).

[109] Eastgate Systems, Inc. *Tinderbox Forum*. 2018. URL: http://forum.eastgate.com (visited on 01/08/2019) (cit. on p. 10).

[110] Eastgate Systems, Inc. *Tinderbox: The Tool For Notes*. 2018. URL: http://www.eastgate.com/Tinderbox/ (visited on 01/08/2019) (cit. on p. 10).

[111] Elizabeth L. Eisenstein. *The Printing Press as an Agent of Change (Volumes 1 and 2)*. Cambridge University Press, 1982, p. 818. ISBN: 0521299551 (cit. on p. 7).

[112] Douglas Carl Engelbart. ‘Augmenting Human Intellect: A Conceptual Framework’. In: 1 (1962), p. 2007 (cit. on p. 26).

[113] Douglas Carl Engelbart. ‘NLS teleconferencing features: The journal, and shared-screen telephoning’. In: *Fall COMPCON 75 Digest of Papers* (1975), pp. 173–177 (cit. on p. 27).

[114] Douglas Carl Engelbart. ‘The Augmented Knowledge Workshop’. In: vol. Proceedings of the 1986 ACM Conference on the History of Personal Workstations. ACM, 1986, p. 11. ISBN: 0-89791-176-8. DOI: 12178.12184 (cit. on p. 28).

[115] Douglas Carl Engelbart. ‘The Augmented Knowledge Workshop’. In: *A History of Personal Workstations*. Ed. by Adele Goldberg. New York, NY, USA: ACM, 1988, pp. 185–248. ISBN: 0-201-11259-0. DOI: 10.1145/61975.66918 (cit. on p. 26).

[116] Douglas Carl Engelbart. ‘Toward High-performance Knowledge Workers’. In: *Computer-supported Cooperative Work: A Book of Readings*. Ed. by Irene Greif. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988, pp. 67–78. ISBN: 0-934613-57-5 (cit. on p. 28).

[117] Astrid Ensslin. *Canonizing Hypertext: Explorations and Constructions*. A&C Black, 2007, p. 197. ISBN: 9780826495587 (cit. on p. 251).

[118] Oren Etzioni, Neal Lesh and Richard Segal. ‘Building Softbots for UNIX’. In: vol. Software Agents—Papers from the 1994 Spring Symposium (Technical Report SS–94–03). AAAI Press, 1994, pp. 9–16 (cit. on p. 45).

[119] Oren Etzioni and Daniel Weld. ‘A Softbot-Based Interface to the Internet’. In: *Communications of the ACM* 37.7 (1994), pp. 72–76. DOI: 10.1145/176789.176797 (cit. on p. 45).

[120] Samer Faraj, Sirkka L. Jarvenpaa and Ann Majchrzak. 'Knowledge Collaboration in Online Communities'. In: *Organization Science* 22.5 (2011), pp. 1224–1239. DOI: 10.1287/orsc.1100.0614 (cit. on p. 212).

[121] Steven Feiner. 'Authoring Large Hypermedia Documents with IGD.' In: *Electronic Publishing* 3.1 (1990), pp. 29–46 (cit. on p. 238).

[122] Pnina Fichman and Noriko Hara. *Global Wikipedia: International and Cross-cultural Issues in Online Collaboration*. Rowman & Littlefield, 2014, p. 178. ISBN: 9780810891029 (cit. on pp. 273–274).

[123] Kathleen M. Fisher. 'SemNet: A Tool for Personal Knowledge Construction'. In: vol. Cognitive Tools for Learning. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 63–75. ISBN: 978-3-642-77222-1 (cit. on p. 238).

[124] Uwe Flick. *An Introduction to Qualitative Research*. SAGE Publications Ltd, 2014, p. 616. ISBN: 1446267792 (cit. on p. 62).

[125] Uwe Flick, Ernst Kardoff von and Ines Steinke. *A Companion to Qualitative Research*. Ed. by Uwe Flick. SAGE Publications Ltd, 2004, p. 446. ISBN: 9780761973751 (cit. on p. 62).

[126] Andrew M. Fountain et al. 'MICROCOSM: An Open Model for Hypermedia with Dynamic Linking.' In: vol. Proceedings of the First European Conference on Hypertext. The Pitt Building, Trumpington Street, Cambridge: Cambridge University Press, 1990, pp. 298–311. ISBN: 9780521405171 (cit. on p. 243).

[127] Stan Franklin and Art Graesser. 'Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents'. In: vol. Intelligent Agents III Agent Theories, Architectures, and Languages. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 21–35. ISBN: 978-3-540-68057-4. DOI: 10.1007/BFb0013570 (cit. on pp. 45, 122).

[128] Richard Furuta. 'Halasz's "Seven Issues" in Context'. In: *ACM Journal of Computer Documentation (JCD)* 25.3 (2001), pp. 92–95. DOI: 10.1145/507317.507324 (cit. on p. 35).

[129] Richard Furuta and David Stotts. 'Separating Hypertext Content from Structure in Trellis'. In: *Hypertext: State of The Art*. Ed. by Ray McAleese and Catherine Green. Oxford, UK: Intellect Limited, 1990, pp. 205–213. ISBN: 1-871516-08-0 (cit. on p. 243).

[130] Richard Furuta et al. 'Hypertext Paths and the World-Wide Web: Experiences with Walden's Paths'. In: vol. Proceedings of the Eighth ACM Conference on Hypertext. Southampton, United Kingdom New York, NY, USA: ACM, 1997, pp. 167–176. ISBN: 0-89791-866-5. DOI: 10.1145/267437.267455 (cit. on p. 38).

[131] *Galaxy Zoo*. Zooniverse.org. 2019. URL: https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/ (visited on 30/01/2019) (cit. on pp. 49, 203).

[132] R. Stuart Geiger. 'The Social Roles of Bots and Assisted Editing Programs'. In: vol. Proceedings of the 5th International Symposium on Wikis and Open Collaboration. Orlando, Florida New York, NY, USA: ACM, 2009, 30:1–30:2. ISBN: 978-1-60558-730-1. DOI: 10.1145/1641309.1641351 (cit. on p. 47).

[133] R. Stuart Geiger. 'The Lives of Bots'. In: *Critical Point of View: A Wikipedia Reader*. Ed. by Geert Lovink and Nathaniel Tkacz. Institute of Network Cultures, 2011, pp. 78–93. ISBN: 978-90-78146-13-1 (cit. on pp. 45, 58).

[134] R. Stuart Geiger. 'Are Computers Merely "Supporting" Cooperative Work: Towards an Ethnography of Bot Development'. In: vol. Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion. San Antonio, Texas, USA New York, NY, USA: ACM, 2013, pp. 51–56. ISBN: 978-1-4503-1332-2. DOI: 10.1145/2441955.2441970 (cit. on pp. 45, 195).

[135] R. Stuart Geiger. 'Design by Bot: Power and Resistance in the Development of Automated Software Agents'. In: *ir14* AoIR Selected Papers of Internet Research (2013), 4 pp. (Cit. on pp. 45–46).

[136] R. Stuart Geiger. 'Bots, Bespoke, Code and the Materiality of Software Platforms'. In: *Information, Communication & Society* 17.3 (2014), pp. 342–356. DOI: 10.1080/1369118X.2013.873069 (cit. on pp. 124–125, 199).

[137] R. Stuart Geiger and Aaron Halfaker. 'When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes?' In: vol. Proceedings of the 9th International Symposium on Open Collaboration. Hong Kong, China New York, NY, USA: ACM, 2013, 6:1–6:6. ISBN: 978-1-4503-1852-5. DOI: 10.1145/2491055.2491061 (cit. on p. 199).

[138] R. Stuart Geiger and Aaron Halfaker. 'Operationalizing Conflict and Cooperation Between Automated Software Agents in Wikipedia: A Replication and Expansion of 'Even Good Bots Fight''. In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW (2017), 49:1–49:33. DOI: 10.1145/3134684 (cit. on pp. 199–200).

[139] Jim Gemmell et al. 'MyLifeBits: Fulfilling the Memex Vision'. In: vol. Proceedings of the Tenth ACM International Conference on Multimedia. Juan-les-Pins, France New York, NY, USA: ACM, 2002, pp. 235–238. ISBN: 1-58113-620-X. DOI: 10.1145/641007.641053 (cit. on p. 32).

[140] R. S. Gilyarevskii and Martin M. Subbotin. 'Russian Experience in Hypertext: Automatic Compiling of Coherent Texts'. In: *Journal of the American Society for Information Science* 44.4 (1993), pp. 185–193. DOI: 10.1002/(SICI)1097-4571(199305)44:4\%3C185::AID-ASI1\%3E3.0.CO;2-2/abstract (cit. on p. 34).

[141] Robert J. Glushko. *The Discipline of Organizing*. MIT Press, 2013, p. 752. ISBN: 0262518503 (cit. on p. 42).

[142] gnu.org. *The GNU General Public License*. 2016. URL: http://www.gnu.org/licenses/licenses.html#GPL (visited on 01/08/2019) (cit. on p. 41).

[143] gnu.org. *The GNU Free Documentation License*. 2019. URL: https://www.gnu.org/licenses/fdl-1.3.html (visited on 01/08/2019) (cit. on p. 398).

[144] Danny Goodman et al. *The Complete Hypercard Handbook*. New York, NY: Bantam Books, 1988, 912 pp. ISBN: 9780553343915 (cit. on pp. 37, 240).

[145] H. Ted Goranson. *Ted Goranson - Personal Blog*. 2016. URL: http://blog.tedgoranson.com/ (visited on 01/08/2019) (cit. on pp. 10, 29).

[146] Robert Gorwa and Douglas Guilbeault. 'Unpacking the Social Media Bot: A Typology to Guide Research and Policy'. In: *arXiv* (2018), 30pp (cit. on pp. 46, 196).

[147] James O. Goulding, Timothy J. Brailsford and Helen L. Ashman. 'Hyperorders and Transclusion: Understanding Dimensional Hypertext'. In: vol. Proceedings of the 21st ACM Conference on Hypertext and Hypermedia. ACM, 2010, pp. 201–210. ISBN: 1450300413. DOI: 10.1145/1810617.1810652 (cit. on p. 30).

[148] T. A. Grinchenko. 'Hypertext: A new information technology'. In: *Cybernetics and Systems Analysis* 28.5 (1992), pp. 751–765. DOI: 10.1007/BF01131855 (cit. on p. 34).

[149] Kaj Grønbæk and Randall H. Trigg. 'Design Issues for a Dexter-Based Hypermedia System'. In: *Commun. ACM* 37.2 (1994), pp. 40–49. DOI: 10.1145/175235.175238 (cit. on p. 245).

[150] Kaj Grønbæk and Randall H. Trigg. 'Toward a Dexter-Based Model for Open Hypermedia: Unifying Embedded References and Link Objects'. In: vol. Proceedings of the the Seventh ACM Conference on Hypertext. ACM, 1996, pp. 149–160. ISBN: 0897917782. DOI: 10.1145/234828.234843 (cit. on p. 245).

[151] Bernard J. Haan et al. 'IRIS Hypermedia Services'. In: *Communications of the ACM (CACM)* 35.1 (1992), pp. 36–51. DOI: 10.1145/129617.129618 (cit. on p. 239).

[152] Katie Hafner and Matthew Lyon. *Where Wizards Stay Up Late: The Origins of the Internet*. Simon & Schuster, 1998, p. 304. ISBN: 0684832674 (cit. on p. 397).

[153] Frank Halasz and Schwartz Mayer. 'The Dexter Hypertext Reference Model'. In: vol. Proceedings of the Second Hypertext Standarization Workshop (NIST). Gaithersburg, MD: National Institute of Standards and Technology, 1990, pp. 95–133 (cit. on p. 245).

[154] Frank G. Halasz. 'Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems'. In: *Communications of the ACM (CACM)* 31.7 (1988), pp. 836–852. DOI: 10.1145/48511.48514 (cit. on pp. 5, 35–36, 236, 239, 249–250).

[155] Frank G. Halasz. '"Seven Issues": Revisited'. In: vol. Proceedings of ACM Hypertext'91 Conference. San Antonio, Texas: ACM, 1991, pp. 1–18 (cit. on pp. 35–37, 243, 249).

[156] Frank G. Halasz. 'Reflections on "Seven Issues": Hypertext in the Era of the Web'. In: *ACM Journal of Computer Documentation (JCD)* 25.3 (2001), pp. 109–114. DOI: 10.1145/507317.507328 (cit. on pp. 35–36).

[157] Frank G. Halasz, Thomas P. Moran and Randall H. Trigg. 'Notecards in a Nutshell'. In: *ACM SIGCHI Bulletin* 17.SI (1986), pp. 45–52. DOI: 10.1145/29933.30859 (cit. on pp. 35, 239, 249–250).

[158] Frank G. Halasz et al. 'The Dexter Hypertext Reference Model'. In: *Communications of the ACM (CACM)* 37.2 (1994), pp. 30–39. DOI: 10.1145/175235.175237 (cit. on pp. 245, 248, 253).

[159] Aaron Halfaker and John Riedl. 'Bots and Cyborgs: Wikipedia's Immune System'. In: *Computer* 45.3 (2012), pp. 79–82. DOI: 10.1109/MC.2012.82 (cit. on pp. 123, 198).

[160] Andrew Hall, Loren Terveen and Aaron Halfaker. 'Bot Detection in Wikidata Using Behavioral and Other Informal Cues'. In: *Proceedings of ACM Human-Computer Interaction - CSCW (CSCW)* 2.CSCW (2018), 64:1–64:18. DOI: 10.1145/3274333 (cit. on p. 198).

[161] Wendy Hall, Hugh C. Davis and Gerard Hutchings. *Rethinking Hypermedia: The Microcosm Approach*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1996. ISBN: 9780792396796 (cit. on p. 399).

[162] Wendy Hall and Thanassis Tiropanis. 'Web Evolution and Web Science'. In: *Computer Networks* 56.18 (2012), pp. 3859–3865. DOI: 10.1016/j.comnet.2012.10.004 (cit. on p. 50).

[163] Nick V. Hammond and Lesley Allinson. 'Travel Around a Learning Support Environment: Rambling, Orienteering or Touring?' In: vol. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Washington, D.C., USA New York, NY, USA: ACM, 1988, pp. 269–273. ISBN: 0-201-14237-6. DOI: 10.1145/57167.57212 (cit. on p. 37).

[164] Charlie Hargood, Fred Charles and David E. Millard. 'Intelligent Generative Locative Hyperstructure'. In: vol. Proceedings of the 29th on Hypertext and Social Media. Baltimore, MD, USA New York, NY, USA: ACM, 2018, pp. 238–241. ISBN: 978-1-4503-5427-1. DOI: 10.1145/3209542.3210574 (cit. on p. 248).

[165] Charlie Hargood, Mark J. Weal and David E. Millard. 'The StoryPlaces Platform: Building a Web-Based Locative Hypertext System'. In: vol. Proceedings of the 29th on Hypertext and Social Media. Baltimore, MD, USA New York, NY, USA: ACM, 2018, pp. 128–135. ISBN: 978-1-4503-5427-1. DOI: 10.1145/3209542.3209559 (cit. on p. 248).

[166] Charlie Hargood et al. 'Exploring (the Poetics of) Strange (and Fractal) Hypertexts'. In: vol. Proceedings of the 23rd ACM Conference on Hypertext and Social Media. ACM, 2012, pp. 181–186. ISBN: 1450313353. DOI: 10.1145/2309996.2310027 (cit. on p. 247).

[167] Charlie Hargood et al. 'Patterns of Sculptural Hypertext in Location Based Narratives'. In: vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada New York, NY, USA: ACM, 2016, pp. 61–70. ISBN: 978-1-4503-4247-6. DOI: 10.1145/2914586.2914595 (cit. on p. 247).

[168] L.P. Hartley. *The Go-Between*. London, UK: Hamish Hamilton, 1953, p. 296 (cit. on p. 219).

[169] Angela Hass. 'Wampum as Hypertext: An American Indian Intellectual Tradition of Multimedia Theory and Practice.' In: *Studies in American Indian Literatures* 19.4 (2007), pp. 77–100. DOI: 10.1353/ail.2008.0005 (cit. on p. 22).

[170] Matt Haughey. *Caroline Wishes Outloud For*. 2000. URL: http://a.wholelottanothing.org/2000/03/caroline-wishes-outloud-for.html (visited on 01/08/2019) (cit. on p. 39).

[171] Tom Heath and Christian Bizer. 'Linked Data: Evolving the Web into a Global Data Space'. In: *Synthesis Lectures on the Semantic Web: Theory and Technology* 1 (2011), pp. 1–136. DOI: 10.2200/S00334ED1V01Y201102WBE001 (cit. on p. 48).

[172] James A. Hendler and Tim Berners-Lee. 'From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web'. In: *Artificial Intelligence* 174.2 (2010), pp. 156–161. DOI: 10.1016/j.artint.2009.11.010 (cit. on p. 50).

[173] James A. Hendler and Alice M. Mulvehill. *Social Machines: The Coming Collision of Artificial Intelligence, Social Networking, and Humanity*. Apress, 2016, p. 192. ISBN: 148421157X (cit. on p. 50).

[174] Michael Hiltzik. *Dealers of Lightning: Xerox Parc and the Dawn of the Computer Age*. HarperBusiness, 2000, p. 480. ISBN: 0887309895 (cit. on p. 27).

[175] Yutaka Hirakawa, Shuichi Uchida and Michihiro Monden. 'Hypermedia As A Communication Design Software Support'. In: *Hypertext: State of the Art* (1990), pp. 244–251 (cit. on p. 34).

[176] *Holistic, adj., a*. Oxford English Dictionary (online). 1926. URL: https://www.oed.com/view/Entry/104170#eid1458658 (visited on 30/01/2019) (cit. on p. 9).

[177] *Houghton Mifflin Harcourt*. Hmhco.com. 2019. URL: https://www.hmhco.com (visited on 30/01/2019) (cit. on p. 229).

[178] Ronald D. Houston and Glynn Harmon. 'Vannevar Bush and Memex'. In: *Annual Review of Information Science and Technology* 41.1 (2007), pp. 55–92. DOI: 10.1002/aris.2007.1440410109 (cit. on pp. 23, 25).

[179] Jennifer Howard. *Computers in English Class, Circa 1974*. 2016. URL: https://www.neh.gov/humanities/2016/spring/feature/computers-in-english-class-circa-1974 (visited on 01/08/2019) (cit. on p. 20).

[180] Haowei Hsieh and Frank M. Shipman III. 'VITE: A Visual Interface Supporting the Direct Manipulation of Structured Data Using Two-way Mappings'. In: vol. Proceedings of the 5th International Conference on Intelligent User Interfaces. New Orleans, Louisiana, USA New York, NY, USA: ACM, 2000, pp. 141–148. ISBN: 1-58113-134-8. DOI: 10.1145/325737.325810 (cit. on p. 252).

[181] *HypeDyn*. Narrativeandplay.org. 2019. URL: http://www.narrativeandplay.org/hypedyn/ (visited on 30/01/2019) (cit. on p. 252).

[182] *Hypertext Markup Language - 2.0*. Internet Engineering Task Force (IETF). 1995. URL: https://tools.ietf.org/html/rfc1866 (visited on 30/01/2019) (cit. on p. 43).

[183] *Hypertext: An Educational Experiment in English and Computer Science at Brown University*. Youtube.com. 1976. URL: https://www.youtube.com/watch?v=xOFM5_n8KeI (visited on 30/01/2019) (cit. on p. 236).

[184] *Hypertext'89 Trip Report*. Nielsen Norman Group. 1990. URL: https://www.nngroup.com/articles/trip-report-hypertext-89/ (visited on 30/01/2019) (cit. on p. 238).

[185] *Hyperwave*. Hyperwave GmbH. 2019. URL: https://www.hyperwave.com (visited on 30/01/2019) (cit. on p. 244).

[186] *I Can't Stop Thinking! #4*. Scott McCloud. 2000. URL: https://scottmccloud.com/1-webcomics/icst/icst-4/icst-4.html (visited on 30/01/2019) (cit. on p. 38).

[187] *INDEX THOMISTICUS*. Enrique Alarcón & Eduardo Bernot. 2005. URL: http://www.corpusthomisticum.org/it/index.age (visited on 30/01/2019) (cit. on p. 21).

[188] Dariusz Jemielniak. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press, 2015, p. 312. ISBN: 0804797234 (cit. on pp. 273–274).

[189] E. E. Jervell. *For This Author, 10,000 Wikipedia Articles Is a Good Day's Work*. 2014. URL: http://www.wsj.com/articles/for-%20this-author-10-000-wikipedia-articles-is-a-good-days-work-1405305001 (visited on 01/08/2019) (cit. on p. 70).

[190] Daniel S. Jordan et al. ‘Facilitating the Development of Representations in Hypertext with IDE’. In: vol. Proceedings of the Second Annual ACM Conference on Hypertext. Pittsburgh, Pennsylvania, USA New York, NY, USA: ACM, 1989, pp. 93–104. ISBN: 0-89791-339-6. DOI: 10.1145/74224.74232 (cit. on p. 244).

[191] Michael Joyce. *afternoon: a story*. Watertown, MA: Eastgate Systems, Inc., 1987. ISBN: 1-884511-01-5 (cit. on p. 241).

[192] Nicolas Jullien. ‘What We Know About Wikipedia: A Review of the Literature Analyzing the Project(s).’ In: (2012), pp.86 (cit. on p. 200).

[193] Lucie-Aimée Kaffee and Elena Simperl. ‘Analysis of Editors’ Languages in Wikidata’. In: vol. Proceedings of the 14th International Symposium on Open Collaboration. New York, NY, USA Paris, France: Association for Computing Machinery, 2018, pp. 1–5. ISBN: 9781450359368. DOI: 10.1145/3233391.3233965 (cit. on p. 48).

[194] Lucie-Aimée Kaffee et al. ‘A Glimpse into Babel: An Analysis of Multilinguality in Wikidata’. In: vol. Proceedings of the 13th International Symposium on Open Collaboration. Galway, Ireland New York, NY, USA: ACM, 2017, 14:1–14:5. ISBN: 978-1-4503-5187-4. DOI: 10.1145/3125433.3125465 (cit. on p. 48).

[195] Alan Curtis Kay. ‘The Early History of Smalltalk’. In: *History of Programming languages—II*. Ed. by Thomas J. Bergin Jr. and Richard G. Gibson Jr. New York, NY, USA: ACM, 1996, pp. 511–598. ISBN: 0-201-89502-1. DOI: 10.1145/234286.1057828 (cit. on p. 211).

[196] Krista Kennedy. ‘Textual machinery: Authorial agency and bot-written texts in Wikipedia’. In: vol. The Responsibilities of Rhetoric: Proceedings of the 2008 Rhetoric Society of America Conference. 2009 (cit. on p. 198).

[197] Krista Kennedy. ‘Textual Curation’. In: *Computers and Composition* 40 (2016), pp. 175–189. DOI: 10.1016/j.compcom.2016.03.005 (cit. on pp. 8, 58, 273–274).

[198] Sofia Kitromili, James Jordan and David E. Millard. ‘What is Hypertext Authoring?’ In: vol. Proceedings of ACM Hypertext and Social Media. Accepted paper, pre-conference draft. 2019 (cit. on p. 248).

[199] Aniket Kittur and Robert E. Kraut. ‘Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination’. In: vol. Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. San Diego, CA, USA New York, NY, USA: ACM, 2008, pp. 37–46. ISBN: 978-1-60558-007-4. DOI: 10.1145/1460563.1460572 (cit. on p. 187).

[200] *Knowledge, n., III.9.b*. Oxford English Dictionary (online). 1534. URL: https://www.oed.com/view/Entry/104170#eid40032629 (visited on 30/01/2019) (cit. on pp. 6, 16).

[201] David Kolb. ‘Other Spaces for Spatial Hypertext’. In: *Journal of Digital Information (JoDI)* 10.3 (2009) (cit. on p. 253).

[202] David Kolb. *Sprawling Places*. University of Georgia Press, 2010, p. 280. ISBN: 9780820336626 (cit. on p. 247).

[203] David Kolb. ‘Story/Story’. In: vol. Proceedings of the 23rd ACM Conference on Hypertext and Social Media. ACM, 2012, pp. 99–102. ISBN: 1450313353. DOI: 10.1145/2309996.2310013 (cit. on p. 247).

[204] David Kolb. ‘Narrative Hypertext, On The Level’. In: vol. Proceedings of the 3rd Narrative and Hypertext Workshop. ACM, 2013, p. 3. ISBN: 1450320058. DOI: 10.1145/2462216.2462219 (cit. on p. 247).

[205] Jason Kottke. *Finally. Did You Notice The*. 2000. URL: http://www.kottke.org/00/03/finally-did-you-notice-the (visited on 01/08/2019) (cit. on p. 39).

[206] Peter M. Krafft, Michael Macy and Alex “Sandy” Pentland. ‘Bots As Virtual Confederates: Design and Ethics’. In: vol. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Portland, Oregon, USA New York, NY, USA: ACM, 2017, pp. 183–190. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998354 (cit. on p. 196).

[207] Paul B. de Laat. 'The Use of Software Tools and Autonomous Bots Against Vandalism: Eroding Wikipedia's Moral Order?' In: *Ethics and Information Technology* 17.3 (2015), pp. 175–188. DOI: 10.1007/s10676-015-9366-9 (cit. on pp. 58, 199).

[208] Anatoli. S. Lakayev and Martin. M. Subbotin. 'Research and Development of Hypertext and Multimedia Systems in Russia'. In: *IEEE Transactions on Professional Communication* 37.2 (1994), pp. 73–74. DOI: 10.1109/47.291571 (cit. on p. 34).

[209] George P. Landow. *Cyberspace, Hypertext, & Critical Theory*. 1995. URL: http://cyberartsweb.org/cpace/ (visited on 01/08/2019) (cit. on p. 246).

[210] George P. Landow. *The Postcolonial and Postimperial Web*. 1995. URL: http://www.postcolonialweb.org/index.html (visited on 01/08/2019) (cit. on p. 246).

[211] George P. Landow. *The Victorian Web*. 1995. URL: http://www.victorianweb.org/misc/credits.html (visited on 01/08/2019) (cit. on p. 246).

[212] George P. Landow. *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization*. Baltimore, MD, USA: The Johns Hopkins University Press, 2006, p. 456. ISBN: 0801882567 (cit. on pp. 3, 17, 247).

[213] K. Faith Lawrence and Anna Jordanous. 'Gnome on the Range: Finding the Hypertextual Narratives in Ancient Wisdom Texts'. In: vol. Proceedings of the 3rd Narrative and Hypertext Workshop. Paris, France New York, NY, USA: ACM, 2013, 4:1–4:4. ISBN: 978-1-4503-2005-4. DOI: 10.1145/2462216.2462220 (cit. on p. 22).

[214] Andrew Leonard. *Bots: The Origin of New Species*. San Francisco, CA, USA: HardWired, 1997, 218 pp. ISBN: 0140275665 (cit. on p. 45).

[215] Bo Leuf and Ward Cunningham. *The Wiki Way: Quick Collaboration on The Web*. Addison-Wesley Boston, MA, 2001, p. 435. ISBN: 0-201-71499-X (cit. on pp. 40, 271).

[216] Steven Levy. *Insanely Great: The Life And Times of Macintosh the Computer That Changed Everything*. Viking, 1995, p. 304. ISBN: 978-0670852444 (cit. on p. 27).

[217] Andrew Lih. *The Wikipedia Revolution*. Hyperion, 2009, p. 272. ISBN: 9781401303716 (cit. on p. 271).

[218] Randall M. Livingstone. 'Network of Knowledge: Wikipedia as a Sociotechnical System of Intelligence'. PhD thesis. 2012 (cit. on p. 198).

[219] Kiel Long et al. '"Could You Define That in Bot Terms"?: Requesting, Creating and Using Bots on Reddit'. In: vol. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Denver, Colorado, USA New York, NY, USA: ACM, 2017, pp. 3488–3500. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025830 (cit. on pp. 46, 196–197).

[220] Joseph Luft and Harry Ingham. 'The Johari Window'. In: *Human Relations Training News* 5.1 (1961), pp. 6–7 (cit. on p. 399).

[221] Arwid Lund. *Wikipedia, Work and Capitalism: A Realm of Freedom?* Springer, 2017, p. 373. ISBN: 9783319506906 (cit. on pp. 273–274).

[222] Ms. Maiwald. *Hypertext is an unavoidable format when reading digitally. We should be teaching students how to interact with hypertext and to use hypertext to increase their comprehension, not distract. What strategies can we teach students to support their digital reading success?* Twitter.com, Feb. 2019. URL: https://twitter.com/AbbyMaiwald/status/1092208941092749313 (visited on 30/08/2019) (cit. on p. 5).

[223] Judy Malloy. 'Uncle Roger, an online narrabase'. In: *Leonardo* 24.2 (1991), pp. 195–202 (cit. on p. 244).

[224] Bob Mankoff. *The New Yorker Encyclopedia of Cartoons*. Vol. 1. Hachette UK, 2018, p. 1536. ISBN: 9780316484770 (cit. on p. 53).

[225] Michael Marcotty. *Full Text of All Questions Submitted*. Ed. by Richard L. Wexelblat. Vol. History of Programming Languages I. New York, NY, USA: ACM, 1981, pp. 270–276. ISBN: 0-12-745040-8. DOI: 10.1145/800025.1198371 (cit. on p. 42).

[226] *Mariana De Oliveira*. Facebook.com. 2019. URL: https://www.facebook.com/rosenbaren (visited on 30/01/2019) (cit. on p. 228).

[227] Annette Markham and Nancy Baym. *Internet Inquiry: Conversations About Method*. Ed. by Annette N. Markham. Thousand Oaks, CA, USA: Sage Publications, 2008, p. 264. ISBN: 1412910013 (cit. on p. 62).

[228] Catherine C. Marshall. 'NoteCards in the Age of the Web: Practice Meets Perfect'. In: *ACM Journal of Computer Documentation (JCD)* 25.3 (2001), pp. 96–103. DOI: 10.1145/507317.507325 (cit. on p. 249).

[229] Catherine C. Marshall. *Reading and Writing the Electronic Book*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2009, p. 200. ISBN: 1598299050 (cit. on p. 247).

[230] Catherine C. Marshall. 'Reading Spatial Hypertext'. In: *Reading Hypertext*. Ed. by Mark Bernstein and diane Grecco. Watertown, MA: Eastgate Systems, Inc., 2009, pp. 211–220. ISBN: 1-884511-48-1 (cit. on p. 253).

[231] Catherine C. Marshall and Russell A. Rogers. 'Two Years before the Mist: Experiences with Aquanet'. In: vol. Proceedings of the ACM Conference on Hypertext. ACM, 1992, pp. 53–62. ISBN: 089791547X. DOI: 10.1145/168466.168490 (cit. on pp. 250, 397).

[232] Catherine C. Marshall and Frank M. Shipman III. 'Spatial Hypertext and the Practice of Information Triage'. In: vol. Proceedings of the Eighth ACM Conference on Hypertext. ACM, 1997, pp. 124–133. ISBN: 0897918665. DOI: 10.1145/267437.267451 (cit. on p. 253).

[233] Catherine C. Marshall, Frank M. Shipman III and James H. Coombs. 'VIKI: Spatial Hypertext Supporting Emergent Structure'. In: vol. Proceedings of the 1994 ACM European Conference on Hypermedia technology. ACM, 1994, pp. 13–23. ISBN: 0897916409. DOI: 10.1145/192757.192759 (cit. on pp. 250, 253, 400).

[234] Catherine C. Marshall et al. 'Aquanet: A Hypertext Tool to Hold Your Knowledge in Place'. In: vol. Proceedings of the Third Annual ACM Conference on Hypertext. ACM, 1991, pp. 261–275. ISBN: 0897914619. DOI: 10.1145/122974.123000 (cit. on pp. 250, 397).

[235] J. Nathan Matias and Frederick Cheung. 'Emberlight: Share & Publish Spatial Hypertext to the Web'. In: vol. Proceedings of the 21st ACM Conference on Hypertext and Hypermedia. ACM, 2010, pp. 311–312. ISBN: 1450300413. DOI: 10.1145/1810617.1810691 (cit. on pp. 10, 251).

[236] J. Nathan Matias and David P. Williams. 'Comparing Spatial Hypertext Collections'. In: vol. Proceedings of the 20th ACM Conference on Hypertext and Hypermedia. ACM, 2009, pp. 45–50. ISBN: 160558486X. DOI: 10.1145/1557914.1557925 (cit. on p. 251).

[237] Hermann A. Maurer. *The Hyper-G Now Hyperwave: The Next Generation Web Solution*. Addison-Wesley Longman, 1996. ISBN: 0201403463 (cit. on p. 244).

[238] Scott McCloud. *Understanding Comics: The Invisible Art*. New York, NY, USA: William Morrow (HarperCollins), 1993, p. 224. ISBN: 006097625X (cit. on pp. 38, 246).

[239] Scott McCloud. *Reinventing Comics*. New York, NY, USA: William Morrow (HarperCollins), 2000, p. 256. ISBN: 0060953500 (cit. on p. 38).

[240] Donald L. McCracken and Robert M. Akscyn. 'Experience with the ZOG Human-Computer Interface System'. In: *International Journal of Man-Machine Studies* 21.4 (1984), pp. 293–310 (cit. on p. 237).

[241] Marshall McLuhan. *Understanding Media: The Extensions of Man*. Reprint of 1964 original. MIT Press, 1994, p. 390. ISBN: 0262631598 (cit. on p. 7).

[242] Marshall McLuhan. *The Gutenberg Galaxy*. Reprint of 1962 original. University of Toronto Press, 2017, p. 336. ISBN: 1487522339 (cit. on p. 7).

[243] Mediawiki.org. *MediaWiki*. 2016. URL: http://mediawiki.org/ (visited on 01/08/2019) (cit. on p. 41).

[244] Luis Meneses et al. 'Analyzing the Perceptions of Change in a Distributed Collection of Web Documents'. In: vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada New York, NY, USA: ACM, 2016, pp. 273–278. ISBN: 978-1-4503-4247-6. DOI: 10.1145/2914586.2914628 (cit. on p. 214).

[245] Katrice Metaluna. *Web browsers make use of a concept called hypertext, which was a thing developed in the 80s and at one time touted as the next big thing. While it made the Internet more accessible to more people, it's still just one protocol among many.* Twitter.com, May 2019. URL: https://twitter.com/magickalslave/status/1001605033744465920 (visited on 30/01/2019) (cit. on p. 4).

[246] N. K. Meyrowitz. 'The Missing Link: Why We're All Doing Hypertext Wrong'. In: *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. Ed. by Edward Barrett. Cambridge, MA, USA: MIT Press, 1989, pp. 107–114. ISBN: 0-262-02291-5 (cit. on p. 239).

[247] Norman K. Meyrowitz. 'Intermedia: The Architecture and Construction of an Object-Oriented Hypermedia System and Applications Framework'. In: vol. Conference Proceedings on Object-Oriented Programming Systems, Languages and Applications. Portland, Oregon, USA New York, NY, USA: ACM, 1986, pp. 186–201. ISBN: 0-89791-204-7. DOI: 10.1145/28697.28716 (cit. on p. 239).

[248] Microsoft Corporation. *Microsoft Delivers World's Fastest Modern Browser Available Today*. 1999. URL: https://web.archive.org/web/20071213175948/http://www.microsoft.com/presspass/press/1999/mar99/ieavlpr.mspx (visited on 01/08/2019) (cit. on p. 43).

[249] David E. Millard and Charlie Hargood. 'Tiree Tales: A Co-operative Inquiry into the Poetics of Location-Based Narrative'. In: vol. Proceedings of the 28th ACM Conference on Hypertext and Social Media. Prague, Czech Republic New York, NY, USA: ACM, 2017, pp. 15–24. ISBN: 978-1-4503-4708-2. DOI: 10.1145/3078714.3078716 (cit. on p. 248).

[250] David E. Millard et al. 'FOHM: A Fundamental Open Hypertext Model for Investigating Interoperability between Hypertext Domains'. In: vol. Proceedings of the Eleventh ACM on Hypertext and hypermedia. ACM, 2000, pp. 93–102. ISBN: 1581132271. DOI: 10.1145/336296.336334 (cit. on p. 245).

[251] David E. Millard et al. 'Canyons, Deltas and Plains: Towards a Unified Sculptural Model of Location-Based Hypertext'. In: vol. Proceedings of the 24th ACM Conference on Hypertext and Social Media. ACM, 2013, pp. 109–118. ISBN: 145031967X. DOI: 10.1145/2481492.2481504 (cit. on p. 247).

[252] Alex Mitchell and Kevin McGee. 'The HypeDyn Hypertext Fiction Authoring Tool'. In: vol. Proceedings of the 2Nd Workshop on Narrative and Hypertext. Milwaukee, Wisconsin, USA New York, NY, USA: ACM, 2012, pp. 19–22. ISBN: 978-1-4503-1408-4. DOI: 10.1145/2310076.2310081 (cit. on p. 252).

[253] Claudia Müller-Birn, Leonhard Dobusch and James D. Herbsleb. 'Work-to-rule: The Emergence of Algorithmic Governance in Wikipedia'. In: vol. Proceedings of the 6th International Conference on Communities and Technologies. Munich, Germany New York, NY, USA: ACM, 2013, pp. 80–89. ISBN: 978-1-4503-2104-4. DOI: 10.1145/2482991.2482999 (cit. on p. 196).

[254] Matthias Müller-Prove. 'Vision and reality of hypertext and graphical user interfaces'. PhD thesis. Hamburg Germany: Fachbereich Informatik, Universität Hamburg, 2002 (cit. on p. 34).

[255] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge, MA, USA: MIT Press, 1998, p. 336. ISBN: 0262631873 (cit. on p. 246).

[256] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. MIT Press, 2017, p. 440. ISBN: 0262533480 (cit. on p. 246).

[257] Kumiyo Nakakoji and Yasuhiro Yamamoto. 'Position Statement for Spatial Hypertext Workshop at Hypertext2002'. In: vol. Second Workshop on Spatial Hypertext. College Park, Maryland: ACM, 2002, p. 2 (cit. on p. 34).

[258] Kumiyo Nakakoji et al. 'Interaction Design for Scholarly Writing: Hypertext Representations as a Means for Creative Knowledge Work'. In: *New Review of Hypermedia and Multimedia* 11.1 (2005), pp. 39–67. DOI: 10.1080/13614560500191238 (cit. on p. 34).

[259] Theodor Holm Nelson. 'Complex Information Processing: A File Structure for the Complex, the Changing and the Indeterminate'. In: vol. Proceedings of the 1965 20th National Conference. Cleveland, Ohio, USA New York, NY, USA: ACM, 1965, pp. 84–100. DOI: 10.1145/800197.806036 (cit. on pp. 17, 28, 41–42).

[260] Theodor Holm Nelson. 'The Hypertext'. In: vol. Congress of the International Federation for Documentation. Washington, D.C.: International Federation for Documentation, 1965 (cit. on p. 28).

[261] Theodor Holm Nelson. *Xanadu Manifesto*. 1966. (Visited on 01/08/2019) (cit. on pp. 28, 229).

[262] Theodor Holm Nelson. *Hypertext Implementation Notes*. Digital transcription at: https://archive.org/stream/hin68/Transcription,hin68#mode/2up. 1968. URL: https://archive.org/details/hin68 (visited on 01/08/2019) (cit. on pp. 29, 234).

[263] Theodor Holm Nelson. 'No More Teachers' Dirty Looks'. In: *Computer Decisions* 9.8 (1970), pp. 16–23 (cit. on p. 29).

[264] Theodor Holm Nelson. *Computer Lib/Dream Machines*. Vol. (1987 Reprint). Tempus Books of Microsoft Press, 1974, p. 336. ISBN: 0914845497 (cit. on pp. 2, 29).

[265] Theodor Holm Nelson. *Selected Papers 1965–77*. 1977. URL: https://archive.org/details/SelectedPapers1977 (visited on 01/08/2019) (cit. on p. 29).

[266] Theodor Holm Nelson. *Literary Machines*. Vol. Edition 87.1. 1987 Revised Edition. First published 1980. Editions: 1980, 1981, 1982, 1983, 1984, 1987 (revised Ed.) Theodor Holm Nelson, 1987, p. 264. ISBN: 0893470562 (cit. on pp. 18, 30, 41–42, 216).

[267] Theodor Holm Nelson. 'As We Will Think'. In: *From Memex to Hypertext*. Ed. by James M. Nyce and Paul Kahn. Originally presented at the International Conference on Online Interactive Computing, at Brunel University, UK (4–7 September 1972), re-published as part of Nyce and Khan's *From Memex To Hypertext* [280]. San Diego, CA, USA: Academic Press Professional, Inc., 1991, pp. 245–260. ISBN: 0-12-523270-5 (cit. on p. 30).

[268] Theodor Holm Nelson. *Literary Machines*. Vol. Edition 93.1. Frontispiece marked 91-1 Ed., but containing 1993 preface. First published 1980. Editions: 1980, 1981, 1982, 1983, 1984, 1987 (revised Ed.), 1990, 1991, 1992, 1993. Mindful Press, 1993, p. 292. ISBN: 978-0893470562 (cit. on pp. 41, 75).

[269] Theodor Holm Nelson. 'The Heart of Connection: Hypermedia Unified by Transclusion'. In: *Communications of the ACM (CACM)* 38.8 (1995), pp. 31–33. DOI: 10.1145/208344.208353 (cit. on pp. 30, 41).

[270] Theodor Holm Nelson. 'Xanalogical Structure, Needed Now More Than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-use'. In: *ACM Computing Surveys (CSUR)* 31.4es (1999), Article 33 (32pp). DOI: 10.1145/345966.346033 (cit. on pp. 30, 41).

[271] Theodor Holm Nelson. *Geeks Bearing Gifts*. 2008, p. 199. ISBN: 9780578004389 (cit. on pp. 234–235).

[272] Theodor Holm Nelson. *Possiplex*. Mindful Press, 2011, p. 382. ISBN: 089347004X (cit. on pp. 19, 26, 28, 205, 234–235).

[273] Theodor Holm Nelson. *Two Cheers for Now*. 2016. URL: https://youtu.be/PtHbfkgKNjQ?t=2630 (visited on 01/08/2019) (cit. on p. 41).

[274] Theodor Holm Nelson. *Xanadu Basics 1a - VISIBLE CONNECTION*. 2018. URL: https://youtu.be/hMKy52Intac (visited on 01/08/2019) (cit. on p. 41).

[275] Netscape Communication Corporation. *Server-Side JavaScript Guide; Ch.2 Getting Started*. 1998. (Visited on 01/08/2019) (cit. on p. 43).

[276] Moritz Neumüller. 'Mind The Eye: On The Relevance of Composition in Spatial Organization'. In: *New Review of Hypermedia and Multimedia* 6.1 (2000), pp. 197–202. DOI: 10.1080/13614560008914723 (cit. on p. 253).

[277] Allen Newell et al. *Report on a workshop in new techniques in cognitive research*. Pittsburgh, PA: Department of Computer Science, Carnegie-Mellon University, 1972, p. 54 (cit. on p. 236).

[278] Sabine Niederer and José van Dijck. 'Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System'. In: *New Media & Society* 12.8 (2010), pp. 1368–1387. DOI: 10.1177/1461444810365297 (cit. on p. 193).

[279] James M. Nyce and Paul Kahn. 'A Machine for The Mind: Vannevar Bush's Memex'. In: *From Memex To Hypertext: Vannevar Bush and the Mind's Machine*. Ed. by James M. Nyce and Paul Kahn. Academic Press, 1992, pp. 37–66. ISBN: 9780125232708 (cit. on pp. 23–24, 38).

[280] James M. Nyce and Paul Kahn. *From Memex To Hypertext: Vannevar Bush and the Mind's Machine*. Academic Press, 1992, p. 367. ISBN: 9780125232708 (cit. on pp. 18, 25, 340).

[281] Kieron O'Hara. 'Trust in Social Machines: The Challenges'. In: AISB/IACAP World Congress 2012: Social Computing, Social Cognition, Social Networks and Multiagent Systems.SOCIAL TURN/SNAMAS (2012) (cit. on p. 193).

[282] Dan O'Sullivan. *Wikipedia: A New Community of Practice?* Routledge, 2009, p. 200. ISBN: 0754674339 (cit. on pp. 273–274).

[283] George Orwell. *Animal Farm: A Fairy*. London, England: Secker and Warburg, 1945 (cit. on p. 54).

[284] Paul Otlet. *Le Livre sur le Livre: Traité de Documentation*. Les Impressions nouvelles, 2015. ISBN: 287449299X (cit. on pp. 20, 207, 311–312).

[285] Andrew Pam and Arnould Vermeer. 'A comparison of WWW and Hyper-G'. In: *J. UCS The Journal of Universal Computer Science* 1.11 (1995), pp. 744–750. DOI: 10.3217/jucs-001-11-0744 (cit. on p. 244).

[286] Amy Pearl. 'Sun's Link Service: A Protocol for Open Linking'. In: vol. Proceedings of the Second Annual ACM Conference on Hypertext. New York, NY, USA Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 1989, pp. 137–146. ISBN: 0897913396. DOI: 10.1145/74224.74236 (cit. on p. 245).

[287] Benjamin Peters. *How Not to Network a Nation: The Uneasy History of the Soviet Internet*. MIT Press, 2017, p. 312. ISBN: 0262534665 (cit. on p. 34).

[288] Michael Polanyi. *The Tacit Dimension*. London, UK: Routledge & Kegan Paul, 1967, p. 108. ISBN: 978-0226672984 (cit. on pp. 8, 250).

[289] Neil Postman. *Amusing Ourselves to Death*. Penguin, 2005, p. 208. ISBN: 9781101042625 (cit. on p. 7).

[290] *Project MAC*. mulicians.org. 1963. URL: https://multicians.org/project-mac.html (visited on 30/01/2019) (cit. on p. 29).

[291] *Project Xanadu*. Xanadu.com. 2019. URL: http://xanadu.com/xuDemoPage.html (visited on 30/01/2019) (cit. on p. 31).

[292] *Project Xanadu*. Xanadu.com. 2019. URL: http://xanadu.com/xuEDL.html (visited on 30/01/2019) (cit. on p. 398).

[293] Sebastian Rahtz, Les A. Carr and Wendy Hall. 'Creating multimedia documents: Hypertext processing'. In: *Hypertext: State of The Art* (1990) (cit. on p. 245).

[294] Sebastian Rahtz, Les A. Carr and Wendy Hall. 'New Designs for Archaeological Reports'. In: *Science and Archaeology* 31 (1990), pp. 20–34 (cit. on p. 245).

[295] Jef Raskin. 'The Hype in Hypertext: A Critique'. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 325–330. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317449 (cit. on p. 33).

[296] Eric S. Raymond. *The Cathedral & the Bazaar*. O'Reilly Media, 2001, p. 258. ISBN: 0596001088 (cit. on pp. 9, 56, 76, 109, 135).

[297] *Raytheon*. Raytheon Company. 2019. URL: https://www.raytheon.com (visited on 30/01/2019) (cit. on p. 22).

[298] Joseph Michael Reagle. *Good Faith Collaboration: The Culture of Wikipedia*. MIT Press, 2012, p. 264. ISBN: 0262518201 (cit. on pp. 197, 273–274).

[299] *reCAPTCHA*. Google.com. 2019. URL: https://www.google.com/recaptcha/about/ (visited on 30/01/2019) (cit. on p. 49).

[300] *Reddit*. Reddit.com. 2019. URL: https://www.reddit.com (visited on 30/01/2019) (cit. on p. 197).

[301] Joel R. Remde, Louis M. Gomez and Thomas K. Landauer. 'SuperBook: An Automatic Tool for Information Exploration & Hypertext?' In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 175–188. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317440 (cit. on p. 243).

[302] George Robertson, Donald L. McCracken and Allen Newell. 'The ZOG Approach to Man-Machine Communication'. In: *International Journal of Man-Machine Studies* 14.4 (1981), pp. 461–488 (cit. on pp. 236, 400).

[303] George Robertson, Allen Newell and K. Ramakrishna. *ZOG: A man-machine communication philosophy*. Pittsburgh PA: Dept. of Computer Science, Carnegie-Mellon University, 1977, p. 60 (cit. on p. 236).

[304] James Ryan '@xfol'. *Images from 'Consdier The Consequences!'* 2017. URL: https://twitter.com/xfoml/status/904043273022349312 (visited on 01/08/2019) (cit. on p. 20).

[305] Marc Saporta. *Composition No. 1*. Éditions du Seuil, 1962, p. 155 (cit. on p. 22).

[306] Simon Schaffer. 'Enlightened Automata'. In: *The Sciences in Enlightened Europe*. Ed. by William Clark, Jan Golinski and Simon Schaffer. The (Mechanical) Turk is described on pp.154–163. Chicago, IL: The Chicago University Press, 1999, pp. 129–135. ISBN: 978-0226109404 (cit. on p. 181).

[307] Thomas Schedel and Claus Atzenbeck. 'Spatio-Temporal Parsing in Spatial Hypermedia'. In: vol. Proceedings of the 27th ACM Conference on Hypertext and Social Media. Halifax, Nova Scotia, Canada New York, NY, USA: ACM, 2016, pp. 149–157. ISBN: 978-1-4503-4247-6. DOI: 10.1145/2914586.2914596 (cit. on p. 253).

[308] Jan Schultz. 'A History of the PROMIS Technology: An Effective Human Interface'. In: vol. Proceedings of the ACM Conference on The History of Personal Workstations. Palo Alto, California, USA New York, NY, USA: ACM, 1986, pp. 159–182. ISBN: 0-89791-176-8. DOI: 10.1145/12178.12188 (cit. on p. 235).

[309] Nigel R. Shadbolt, Tim Berners-Lee and Wendy Hall. 'The Semantic Web Revisited'. In: *IEEE Intelligent Systems* 21 (2006), pp. 96–101. DOI: 10.1109/MIS.2006.62 (cit. on pp. 47–48).

[310] Nigel R. Shadbolt et al. 'The Future(s) of Social Machines: The Research Agenda'. In: *The Theory and Practice of Social Machines*. Ed. by Nigel R. Shadbolt, Kieron O'Hara and Wendy Hall. Cham: Springer International Publishing, 2019, pp. 201–217. ISBN: 978-3-030-10889-2. DOI: 10.1007/978-3-030-10889-2_5 (cit. on p. 47).

[311] Nigel R. Shadbolt et al. *The Theory and Practice of Social Machines*. Cham: Springer International Publishing, 2019, p. 260. ISBN: 978-3-030-10889-2. DOI: 10.1007/978-3-030-10889-2 (cit. on p. 186).

[312] Jason Shiga. *Meanwhile: Pick Any Path*. Amulet Books, 2010, p. 80. ISBN: 9780810984233 (cit. on p. 38).

[313] Frank M. Shipman III and Catherine C. Marshall. 'Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems'. In: *Computer Supported Cooperative Work (CSCW)* 8.4 (1999), pp. 333–352. DOI: 10.1023/A:1008716330212 (cit. on pp. 62, 250).

[314] Frank M. Shipman III et al. 'The Visual Knowledge Builder: A Second Generation Spatial Hypertext'. In: vol. Proceedings of the 12th ACM Conference on Hypertext and Hypermedia. ACM, 2001, pp. 113–122. ISBN: 1581134207. DOI: 10.1145/504216.504245 (cit. on p. 250).

[315] Ben Shneiderman. 'User Interface Design for the Hyperties Electronic Encyclopedia'. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 189–194. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317441 (cit. on p. 238).

[316] Ben Shneiderman. *Hypertext on Hypertext: IBM PC & Compatibles*. Ed. by Ben Shneiderman. New York, NY, USA: ACM, 1988. ISBN: 0-89791-280-2 (cit. on pp. 33, 238).

[317] *SHOWING CONNECTIONS– the general problem*. Youtube.com. 2019. URL: https://www.youtube.com/watch?v=kjKZUTtAo1Y&frags=pl%2Cwn (visited on 30/01/2019) (cit. on p. 30).

[318] SIMILE Project. *SIMILE Timeline*. 2006. URL: http://www.simile-widgets.org/timeline/ (visited on 01/08/2019) (cit. on p. 10).

[319] Upton Sinclair. 'I, Candidate for Governor'. In: *Oakland Tribune* (1934), p.19 (cit. on p. 67).

[320] Paul R. Smart. 'Knowledge Machines'. In: *The Knowledge Engineering Review* (2017). DOI: 10.1017/s0269888918000139 (cit. on p. 51).

[321] Paul R. Smart and Nigel R. Shadbolt. 'Social Machines'. In: *Encyclopedia of Information Science and Technology Social Machines (Third Edition)*. Ed. by Khosrow-Pour Mehdi. IGI Global, 2014, pp. 6855–6862. ISBN: 9781466658882. DOI: 10.4018/978-1-4666-5888-2.ch675 (cit. on p. 49).

[322] John B. Smith et al. 'WE: A Writing Environment for Professionals'. In: vol. National Computer Conference 1987. Chicago, Ill: American Federation of Information Processing Societies, 1987, pp. 725–736. ISBN: 9780882830513 (cit. on p. 242).

[323] *SOCIAM: The Theory and Practice of Social Machines*. Engineering and Physical research Council. 2015. URL: https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/J017728/ (visited on 30/01/2019) (cit. on p. 51).

[324] *Softer Software: an excerpt from the Maryland State of Mind program*. Youtube.com. 1999. URL: https://youtu.be/b28dGXAvFhE?t=117 (visited on 30/01/2019) (cit. on p. 238).

[325] Nick Sousanis. *Unflattening*. Harvard University Press; edition (2015-04-28), 2015, p. 208. ISBN: 978-0674744431 (cit. on p. 38).

[326] Thomas Steiner. 'Bots vs. Wikipedians, Anons vs. Logged-ins'. In: vol. Proceedings of the 23rd International Conference on World Wide Web. Seoul, Korea New York, NY, USA: ACM, 2014, pp. 547–548. ISBN: 978-1-4503-2745-9. DOI: 10.1145/2567948.2576948 (cit. on pp. 58, 200).

[327] Thomas Steiner. 'Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata'. In: vol. Proceedings of The International Symposium on Open Collaboration. Berlin, Germany New York, NY, USA: ACM, 2014, 25:1–25:7. ISBN: 978-1-4503-3016-9. DOI: 10.1145/2641580.2641613 (cit. on p. 200).

[328] *Storyspace*. Eastgate.com. 2019. URL: http://www.eastgate.com/storyspace/ (visited on 30/01/2019) (cit. on p. 10).

[329] P. David Stotts and Richard Furuta. 'Adding Browsing Semantics to the Hypertext Model'. In: vol. Proceedings of the ACM Conference on Document Processing Systems. Santa Fe, New Mexico, USA New York, NY, USA: ACM, 1988, pp. 43–50. ISBN: 0-89791-291-8. DOI: 10.1145/62506.62516 (cit. on p. 243).

[330] Silvia M. Stoyanova. 'Fragmentary Narrative and the Formation of Pre-Digital Scholarly Hypertextuality: G. Leopardi's Zibaldone and Its Hypertext Rendition'. In: vol. Proceedings of the 3rd Narrative and Hypertext Workshop. Paris, France New York, NY, USA: ACM, 2013, 2:1–2:6. ISBN: 978-1-4503-2005-4. DOI: 10.1145/2462216.2462218 (cit. on p. 19).

[331] *Strange Hypertexts*. University of southampton. 2011. URL: http://www.strangehypertexts.org (visited on 30/01/2019) (cit. on p. 247).

[332] Alex Strick van Linschoten. *Kandahar Timeline 1979-2010*. 2010. URL: http://www.alexstrick.com/blog/2010/06/kandahar-timeline-1979-2010 (visited on 01/08/2019) (cit. on p. 10).

[333] Georgios D. P. Styliaras and Sotiris P. Christodoulou. 'HyperSea: Towards a Spatial Hypertext Environment for Web 2.0 Content'. In: vol. Proceedings of the 20th ACM Conference on Hypertext and Hypermedia. Torino, Italy New York, NY, USA: ACM, 2009, pp. 35–44. ISBN: 978-1-60558-486-7. DOI: 10.1145/1557914.1557924 (cit. on p. 252).

[334] Georgios D. P. Styliaras and Sotiris P. Christodoulou. 'Organizing Personal Web 2.0 Content with Hypersea'. In: vol. Proceedings of the 6th Euro American Conference on Telematics and Information Systems. Valencia, Spain New York, NY, USA: ACM, 2012, pp. 223–230. ISBN: 978-1-4503-1012-3. DOI: 10.1145/2261605.2261638 (cit. on p. 252).

[335] Ed Summers and Ricardo Punzalan. 'Bots, Seeds and People: Web Archives As Infrastructure'. In: vol. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Portland, Oregon, USA New York, NY, USA: ACM, 2017, pp. 821–834. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998345 (cit. on p. 194).

[336] Olof Sundin. 'Janitors of Knowledge: Constructing Knowledge in The Everyday Life of Wikipedia Editor'. In: *Journal of Documentation* 65.5 (2011), pp. 840–862. DOI: 10.1108/00220411111164709 (cit. on p. 190).

[337] Aaron Swartz. *Who Writes Wikipedia*. 2006. URL: http://www.aaronsw.com/weblog/whowriteswikipedia (visited on 01/08/2019) (cit. on p. 9).

[338] *Symbolics Document Examiner and Concordia*. Vimeo.com: Rainer Joswig. 2014. URL: https://vimeo.com/83886950 (visited on 30/01/2019) (cit. on p. 240).

[339] Philip Tchernavskij. 'Transclusion in Software for the Complex, the Changing, and the Indeterminate'. PhD thesis. Aarhus Universitet, Datalogisk Institut, 2016, p. 99 (cit. on p. 44).

[340] *Ted Nelson demos his ZigZag database system*. Youtube.com. 2010. URL: https://www.youtube.com/watch?v=6yFcJXBMGQQ (visited on 30/01/2019) (cit. on p. 30).

[341] *TEDtalk by TED at the TED2 conference, 1990: at 00:00:30*. Nelson cites his aphorism "Power corrupts and obsolete power corrupts obsoletely". Youtube.com. 2019. URL: https://youtu.be/PUTmmwbkN6Y?t=30 (visited on 30/01/2019) (cit. on pp. 185, 226).

[342] Jennifer Thom-Santelli. 'Expressing Territoriality in Collaborative Activity'. In: vol. Proceedings of the ACM 2009 International Conference on Supporting Group Work. Sanibel Island, Florida, USA New York, NY, USA: ACM, 2009, pp. 389–390. ISBN: 978-1-60558-500-0. DOI: 10.1145/1531674.1531738 (cit. on p. 189).

[343] Jennifer Thom-Santelli. 'Expressing territoriality in online collaborative environments'. PhD thesis. Cornell University, 2010, p. 211 (cit. on p. 189).

[344] Jennifer Thom-Santelli, Dan Cosley and Geri Gay. 'What Do You Know?: Experts, Novices and Territoriality in Collaborative Systems'. In: vol. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Atlanta, Georgia, USA New York, NY, USA: ACM, 2010, pp. 1685–1694. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753578 (cit. on p. 190).

[345] Jennifer Thom-Santelli, Dan R. Cosley and Geri Gay. 'What's Mine is Mine: Territoriality in Collaborative Authoring'. In: vol. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, MA, USA New York, NY, USA: ACM, 2009, pp. 1481–1484. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518925 (cit. on p. 189).

[346] Jennifer Thom-Santelli, Michael J. Muller and David R. Millen. 'Social Tagging Roles: Publishers, Evangelists, Leaders'. In: vol. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Florence, Italy New York, NY, USA: ACM, 2008, pp. 1041–1044. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357215 (cit. on p. 189).

[347] Melissa Thomas-Hunt, Tonya Y. Ogden and Margaret A. Neale. 'Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange within Groups'. In: *Management Science* 49.4 (2003), pp. 464–477. DOI: 10.1287/mnsc.49.4.464.14425 (cit. on p. 192).

[348] tiddlywiki.com. *TiddlyWiki*. 2016. URL: http://tiddlywiki.com (visited on 01/08/2019) (cit. on p. 40).

[349] Nathaniel Tkacz. *Wikipedia and the Politics of Openness*. University of Chicago Press, 2014, p. 224. ISBN: 022619230X (cit. on pp. 58, 273–274).

[350] Robert Tolksdorf and Elena Paslaru Bontas Simperl. 'Towards Wikis As Semantic Hypermedia'. In: vol. Proceedings of the 2006 International Symposium on Wikis. Odense, Denmark New York, NY, USA: ACM, 2006, pp. 79–88. ISBN: 1-59593-413-8. DOI: 10.1145/1149453.1149470 (cit. on p. 47).

[351] trailblazer.io. *Trailblazer*. Site is now dark. 2014. URL: http://www.trailblazer.io (visited on 15/10/2015) (cit. on p. 38).

[352] Jesús Tramullas, Piedad Garrido-Picazo and Ana I. Sánchez-Casabón. 'Use of Wikipedia Categories on Information Retrieval Research: A Brief Review'. In: vol. Proceedings of the 5th Spanish Conference on Information Retrieval. Zaragoza, Spain New York, NY, USA: ACM, 2018, 17:1–17:4. ISBN: 978-1-4503-6543-7. DOI: 10.1145/3230599.3230617 (cit. on p. 201).

[353] Randall H. Trigg. 'A Network-Based Approach to Text Handling for the Online Scientific Community'. PhD thesis. College Park, MD, USA: University of Maryland at College Park, 1983 (cit. on pp. 35, 248).

[354] Randall H. Trigg. 'Guided Tours and Tabletops: Tools for Communicating in a Hypertext Environment'. In: *ACM Transactions on Information Systems (TOIS)* 6.4 (1988), pp. 398–414. DOI: 10.1145/58566.59299 (cit. on pp. 37, 249).

[355] Randall H. Trigg. 'From Trailblazing to Guided Tours: The Legacy of Vannevar Bush's Vision of Hypertext Use'. In: *From Memex to Hypertext*. Ed. by James M. Nyce and Paul Kahn. San Diego, CA: Academic Press Professional, Inc., 1991, pp. 353–367. ISBN: 0-12-523270-5 (cit. on p. 38).

[356] Randall H. Trigg and Mark Weiser. 'TEXTNET: A Network-Based Approach to Text Handling'. In: *ACM Transactions on Information Systems (TOIS)* 4.1 (1986), pp. 1–23 (cit. on pp. 37, 239, 245).

[357] *True Structure: ZigZag*. Xanadu.com. 2019. URL: http://xanadu.com/zigzag/ (visited on 30/01/2019) (cit. on p. 30).

[358] Milena Tsvetkova et al. 'Even Good Bots Fight: The Case of Wikipedia'. In: *PLoS One* 12.2 (2017). DOI: 10.1371/journal.pone.0171774 (cit. on pp. 58, 199).

[359] Alan M. Turing. 'Computing Machinery and Intelligence'. In: *Mind* LIX.236 (1950), pp. 433–460. DOI: 10.1093/mind/LIX.236.433 (cit. on p. 44).

[360] *Two Basic Hypertext Presentation Models*. Nielsen Norman Group. 1995. URL: https://www.nngroup.com/articles/two-basic-hypertext-presentation-models/ (visited on 30/01/2019) (cit. on p. 33).

[361] The W.E.B. Du Bois Center at the University of Massachusetts Amherst. *W. E. B. Du Bois's Data Portraits: Visualizing Black America*. Ed. by Whitney Battle-Baptiste and Britt Rusert. Princeton Architectural Press, 2018, p. 144. ISBN: 9781616897062 (cit. on p. 21).

[362] *Ushahidi*. Ushahidi.com. 2019. URL: https://www.ushahidi.com (visited on 30/01/2019) (cit. on p. 49).

[363] William Van Ornum. *The Priest Who Invented Hypertext*. 2013. URL: https://aleteia.org/2013/09/26/the-priest-who-invented-hypertext/ (visited on 01/08/2019) (cit. on p. 21).

[364] Nader Vossoughian. *Otto Neurath*. Nai Uitgevers Publishers, 2011, p. 175. ISBN: 9789056627980 (cit. on p. 21).

[365] Mitchell Waldrop. *The Dream Machine*. Viking, 2001, p. 512. ISBN: 0670899763 (cit. on p. 397).

[366] Janet H. Walker. 'Document Examiner: Delivery Interface for Hypertext Documents'. In: vol. Proceedings of the ACM Conference on Hypertext. Chapel Hill, North Carolina, USA New York, NY, USA: ACM, 1987, pp. 307–323. ISBN: 0-89791-340-X. DOI: 10.1145/317426.317448 (cit. on p. 240).

[367] QianYing Wang et al. 'The Role of Psychological Ownership and Ownership Markers in Collaborative Working Environment'. In: vol. Proceedings of the 8th International Conference on Multimodal Interfaces. Banff, Alberta, Canada New York, NY, USA: ACM, 2006, pp. 225–232. ISBN: 1-59593-541-X. DOI: 10.1145/1180995.1181041 (cit. on p. 188).

[368] Noah Wardrip-Fruin. *The New Media Reader*. Cambridge, MA, USA: MIT Press, 2003, p. 824. ISBN: 0262232278 (cit. on p. 246).

[369] Noah Wardrip-Fruin. 'What Hypertext Is'. In: vol. Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia. ACM, 2004, pp. 126–127. ISBN: 1581138482. DOI: 10.1145/1012807.1012844 (cit. on p. 249).

[370] Marc Weber. 'A Common Language'. In: *Internet Histories* 1.1-2 (2017), pp. 26–38. DOI: 10.1080/24701475.2017.1317118 (cit. on p. 51).

[371] Doris Webster and Mary Alden Hopkins. *Consider the Consequences*. The Century Company, 1930, p. 186 (cit. on p. 20).

[372] Joseph Weizenbaum. *Computer Power and Human Reason: From Judgment to Calculation*. Oxford, England: W. H. Freeman & Co, 1976, p. 300. ISBN: 0716704633 (cit. on p. 44).

[373] H. G. Wells. *World Brain*. H. G. Wells Library, 1937, p. 146. ISBN: 147333375X (cit. on p. 21).

[374] Stephen A. Weyer. 'The design of a dynamic book for information search'. In: *International Journal of Man-Machine Studies* 17.1 (1982), pp. 87–107. DOI: 10.1016/S0020-7373(82)80011-4 (cit. on p. 240).

[375] *Where are the hypertexts?* Eastgate.com. 2019. URL: https://www.eastgate.com/ht99/slides/sld001.htm (visited on 30/01/2019) (cit. on p. 15).

[376] *Why Hyperwave?* Jaschke.net. 2000. URL: https://jaschke.net/hyperwave.html (visited on 30/01/2019) (cit. on p. 244).

[377] Wikimedia Foundation. *Wikimedia Foundation*. 2016. URL: https://wikimediafoundation.org/ (visited on 01/08/2019) (cit. on p. 62).

[378] Michael Williamson. *I don't see it*. Comment to Anil Dash's article 'The Missing Building Blocks of the Web'. Medium.com. 2019. URL: https://medium.com/@mawcs/i-dont-see-it-9f157f2badbb (visited on 30/01/2019) (cit. on p. 4).

[379] Michael Wooldridge and Nicholas R. Jennings. 'Intelligent Agents: Theory And Practice'. In: *The Knowledge Engineering Review* 10.2 (1995), pp. 115–152. DOI: 10.1017/S0269888900008122 (cit. on p. 45).

[380] World Wide Web Consortium (W3C). *Hypertext Markup Language - 2.0*. 1995. URL: http://www.w3.org/MarkUp/html-spec/html-spec_toc.html#SEC5.10 (visited on 01/08/2019) (cit. on p. 43).

[381] World Wide Web Consortium (W3C). *Inline frames: The IFRAME element*. 1997. URL: https://www.w3.org/TR/REC-html40-971218/present/frames.html#h-16.5 (visited on 01/08/2019) (cit. on p. 43).

[382] World Wide Web Consortium (W3C). *A history of HTML (Ch.2)*. 1998. URL: http://www.w3.org/People/Raggett/book4/ch02.html (visited on 01/08/2019) (cit. on p. 43).

[383] World Wide Web Consortium (W3C). *Document Object Model (DOM)*. 2005. URL: http://www.w3.org/DOM/ (visited on 01/08/2019) (cit. on p. 249).

[384] World Wide Web Consortium (W3C). *HTML 5 Working Draft 22*. 2008. URL: https://www.w3.org/TR/2008/WD-html5-20080122/ (visited on 01/08/2019) (cit. on p. 82).

[385] World Wide Web Consortium (W3C). *HTML 5*. 2014. URL: http://www.w3.org/TR/html5/ (visited on 01/08/2019) (cit. on p. 82).

[386] Alex Wright. *Glut*. National Academies Press, 2007, p. 294. ISBN: 9780309133876 (cit. on p. 235).

[387] Alex Wright. *Cataloging the World: Paul Otlet and the Birth of the Information Age*. OUP USA, 2014, p. 360. ISBN: 0199931410 (cit. on pp. 20–22, 312).

[388] *xanaviewer3*. Xanadu.com. 2019. URL: http://perma.pub/xanaviewer3/ (visited on 30/01/2019) (cit. on pp. 31, 257).

[389] *XLink Specification, section 'Resources in Relation to the Physical Location of a Linking Element'*. W3.org. 2019. URL: https://www.w3.org/TR/xlink11/#dt-linkbase (visited on 30/01/2019) (cit. on p. 399).

[390] Nicole Yankelovich et al. ‘Intermedia: The Concept and the Construction of a Seamless Information Environment’. In: *Computer* 21.1 (1988), pp. 81–96. DOI: 10.1109/2.222120 (cit. on p. 246).

[391] Dave Yates, Christian Wagner and Ann Majchrzak. ‘Factors Affecting Shapers of Organizational Wikis’. In: *Journal of the American Society for Information Science and Technology* 61.3 (2010), pp. 543–554. DOI: 10.1002/asi.21266 (cit. on p. 211).

[392] Amber Young, Ariel Wigdor and Gerald Kane. ‘Theorizing Human And Bot Co-production Effects on Information Quality’. In: *CIS2018* Thirty-ninth International Conference on Information Systems: Social Media And Digital Collaboration (2018) (cit. on p. 197).

[393] Jeffrey S. Young. ‘Hypermedia’. In: *Macworld* 3.3 (1986), pp. 116–121 (cit. on p. 34).

[394] *YouTube*. Youtube.com. 2019. URL: https://www.youtube.com (visited on 30/01/2019) (cit. on p. 272).

[395] V. L. Zefirova and V.M. Chelnokov. ‘HYPERLOG: Hypertext System with Logicosemantic Navigation’. In: *Scientific and Technical Information Series 2 - Information Processes And Systems* 4 (1990), pp. 10–14 (cit. on p. 34).

[396] Polle T. Zellweger. ‘Active Paths Through Multimedia Documents’. In: vol. Proceedings of the International Conference on Electronic Publishing on Document Manipulation and Typography. Nice, France New York, NY, USA: Cambridge University Press, 1988, pp. 19–34. ISBN: 0-521-36294-6 (cit. on p. 37).

[397] Polle T. Zellweger. ‘Scripted Documents: A Hypermedia Path Mechanism’. In: vol. Proceedings of the Second Annual ACM Conference on Hypertext. Pittsburgh, Pennsylvania, USA New York, NY, USA: ACM, 1989, pp. 1–14. ISBN: 0-89791-339-6. DOI: 10.1145/74224.74225 (cit. on pp. 37, 244).

[398] *Zooniverse*. Zooniverse.org. 2019. URL: https://www.zooniverse.org (visited on 30/01/2019) (cit. on p. 186).

Wikipedia References

This section contains only Wikipedia URLs. Many refer to particular edits of a given article page.

[W1] *Amazon Mechanical Turk*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Amazon_Mechanical_Turk (visited on 30/01/2019) (cit. on pp. 186, 203).

[W2] *Antoine Laurent de Jussieu (Wikipedia)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Antoine_Laurent_de_Jussieu (visited on 30/01/2019) (cit. on p. 313).

[W3] *API: Main page*. Wikipedia.org. 2019. URL: https://www.mediawiki.org/wiki/API:Main_page (visited on 30/01/2019) (cit. on pp. 59, 155, 271).

[W4] *Autaugaville, Alabama*. Edit as at 04:21, 9 December 2002. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Autaugaville,_Alabama&oldid=471793 (visited on 30/01/2019) (cit. on p. 125).

[W5] *Big Data*. Edit as at 16:16, 20 July 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Big_data&oldid=730677809 (visited on 20/07/2016) (cit. on p. 7).

[W6] *Biogenic amine*. Edit as at 09:43, 26 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Biogenic_amine&oldid=696845226 (visited on 07/09/2016) (cit. on p. 93).

[W7] *Blind men and an elephant*. Wikimedia.org. 2019. URL: https://en.wikipedia.org/wiki/Blind_men_and_an_elephant (visited on 30/01/2019) (cit. on p. 193).

[W8] *Boston Bruins*. Edit as at 04:06, 11 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=699243091 (visited on 07/09/2016) (cit. on p. 100).

[W9] *Boston Bruins roster*. Edit as at 21:52, 9 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Boston_Bruins_roster&oldid=699038842 (visited on 07/09/2016) (cit. on p. 100).

[W10] *Bots*. Wikipedia.org. 2018. URL: https://en.wikipedia.org/wiki/Bot (visited on 07/09/2018) (cit. on p. 130).

[W11] *Carl Linnaeus (Wikipedia)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Carl_Linnaeus (visited on 30/01/2019) (cit. on p. 313).

[W12] *Category talk: Artists from New Hampshire*. Edit difference: edit as at 21:03, 12 December 2010 vs. edit as at 16:52, 16 December 2010. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Category_talk:Artists_from_New_Hampshire&diff=prev&oldid=466182531&diffmode=source (visited on 30/01/2019) (cit. on p. 137).

[W13] *Category: Revoked Wikipedia bot requests for approval*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Category:Revoked_Wikipedia_bot_requests_for_approval (visited on 30/01/2019) (cit. on p. 137).

[W14] *Cebuano wiki: Xanthopimpla punctate*. Edit as at 04:15, 16 March 2015. Wikipedia.org. 2016. URL: https://ceb.wikipedia.org/w/index.php?title=Xanthopimpla_punctata&oldid=5716074 (visited on 07/09/2016) (cit. on p. 90).

[W15] *Climate change*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Climate_change (visited on 30/01/2019) (cit. on p. 188).

[W16] *Criminal Minds*. Edit as at 08:26, 10 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Criminal_Minds&oldid=699108166 (visited on 07/09/2016) (cit. on p. 103).

[W17] *Criminal Minds*. Edit as at 01:58, 1 May 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Criminal_Minds&oldid=718016840 (visited on 07/09/2016) (cit. on p. 103).

[W18] *Criminal Minds (season 1)*. Edit as at 20:16, 27 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Criminal_Minds_(season_1)&oldid=697036415 (visited on 07/09/2016) (cit. on p. 103).

[W19] *Criminal Minds, section at 'Episodes'*. Edit as at 08:26, 10 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Criminal_Minds&oldid=699108166#Episodes (visited on 07/09/2016) (cit. on p. 103).

[W20] *Diff*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Diff (visited on 30/01/2019) (cit. on p. 256).

[W21] *Digital Methods Initiative: DMI Tools Pertaining to Wikipedia*. Digital Methods Initiative. 2019. URL: https://wiki.digitalmethods.net/Dmi/ToolDatabase?cat=DeviceCentric&subcat=Wikipedia (visited on 30/01/2019) (cit. on p. 63).

[W22] *Dutch wiki: Boston Bruins*. Edit as at 12:29, 30 August 2014. Wikipedia.org. 2016. URL: https://nl.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=41975389 (visited on 07/09/2016) (cit. on p. 100).

[W23] *Extension talk: Labeled Section Transclusion*. Mediawiki.org. 2019. URL: https://www.mediawiki.org/wiki/Extension_talk:Labeled_Section_Transclusion (visited on 30/01/2019) (cit. on p. 82).

[W24] *Federal Standard 1037C*. Edit as at 05:18, 16 February 2002. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Federal_Standard_1037C&oldid=11864 (visited on 30/01/2019) (cit. on p. 125).

[W25] *File: 'Transclusion-multiple-foo.svg'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Transclusion#/media/File:Transclusion-multiple-foo.svg (visited on 30/01/2019) (cit. on pp. 77, 258).

[W26] *Fox cartoons*. Edit as at 08:54, 4 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Fox_cartoons&oldid=698151923 (visited on 07/09/2016) (cit. on p. 105).

[W27] *French wiki: Bruins de Boston*. Edit as at 13:01, 7 January 2016. Wikipedia.org. 2016. URL: https://fr.wikipedia.org/w/index.php?title=Bruins_de_Boston&oldid=122040311 (visited on 07/09/2016) (cit. on p. 100).

[W28] *French wiki: Effectif actuel des Bruins de Boston*. Edit as at 23:54, 25 November 2015. Wikipedia.org. 2016. URL: https://fr.wikipedia.org/w/index.php?title=Effectif_actuel_des_Bruins_de_Boston&oldid=120808444 (visited on 07/09/2016) (cit. on p. 100).

[W29] *Frenchwiki: Manchester United Football Club*. Edit as at 12:04, 10 January 2016. Wikipedia.org. 2016. URL: https://fr.wikipedia.org/w/index.php?title=Manchester_United_Football_Club&oldid=122150406 (visited on 07/09/2016) (cit. on p. 100).

[W30] *Fress*. Wiktionary.org. 2019. URL: https://en.wiktionary.org/wiki/fress (visited on 01/08/2019) (cit. on p. 235).

[W31] *Gabriel Peignot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Gabriel_Peignot (visited on 30/01/2019) (cit. on p. 311).

[W32] *Gamergate controversy*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Gamergate_controversy (visited on 30/01/2019) (cit. on p. 188).

[W33] *Gerald Ford*. Edit as at 07:23, 1 February 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Gerald_Ford&oldid=702716371 (visited on 07/09/2016) (cit. on p. 94).

[W34] *Gerald Ford section 'Pardon of Nixon'*. Edit as at 07:23, 1 February 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Gerald_Ford&oldid=702716371#Pardon_of_Nixon (visited on 07/09/2016) (cit. on p. 94).

[W35] *Gerald Ford section 'Pardon of Nixon'*. Sampled 27 July 2016, when most recent edit was as at 20:22, 23 July 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Gerald_Ford&oldid=731209982#Pardon_of_Nixon (visited on 07/09/2016) (cit. on p. 94).

[W36] *German wiki: Albert Schmid*. Edit as at 19:02, 5 September 2015. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Albert_Schmid&oldid=145751815 (visited on 07/09/2016) (cit. on p. 99).

[W37] *German wiki: Boston Bruins*. Edit as at 12:27, 19 October 2015. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=147160434 (visited on 07/09/2016) (cit. on p. 100).

[W38] *German wiki: Manchester United*. Edit as at 16:38, 7 January 2016. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Manchester_United&oldid=149907098 (visited on 07/09/2016) (cit. on p. 100).

[W39] *German wiki: Marianne Faithfull*. Edit as at 22:05, 29 December 2015. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Marianne_Faithfull&oldid=149584107 (visited on 07/09/2016) (cit. on p. 105).

[W40] *German wiki: Marianne Faithfull/Diskografie*. Edit as at 17:56, 6 December 2015. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Marianne_Faithfull/Diskografie&oldid=148809069 (visited on 07/09/2016) (cit. on p. 105).

[W41] *German wiki: Schmid*. Edit as at 10:10, 25 December 2015. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Schmid&oldid=149411372 (visited on 07/09/2016) (cit. on p. 98).

[W42] *German wiki: Vorlage:Navigationsleiste Kader der Boston Bruins*. Edit as at 19:15, 7 January 2016. Wikipedia.org. 2016. URL: https://de.wikipedia.org/w/index.php?title=Vorlage:Navigationsleiste_Kader_der_Boston_Bruins&oldid=149912322 (visited on 07/09/2016) (cit. on p. 100).

[W43] *Gmina*. Edit as at 13:17, 14 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Gmina&oldid=685701569 (visited on 07/09/2016) (cit. on p. 106).

[W44] *Help: Category*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Category (visited on 30/01/2019) (cit. on pp. 143, 200).

[W45] *Help: Creating a bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Creating_a_bot (visited on 30/01/2019) (cit. on p. 132).

[W46] *Help: Editing*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Editing (visited on 30/01/2019) (cit. on p. 54).

[W47] *Help: Export, 'Example' section*. Revision as of 05:31, 22 June 2015. MediaWiki.org. 2016. URL: https://www.mediawiki.org/w/index.php?title=Help:Export&oldid=1686810#Example (visited on 30/01/2019) (cit. on p. 85).

[W48] *Help: Infobox*. Wikipedia.org. 2018. URL: https://en.wikipedia.org/wiki/Help:Infobox (visited on 07/09/2018) (cit. on pp. 138, 157).

[W49] *Help: Interlanguage links*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Interlanguage_links (visited on 30/01/2019) (cit. on p. 149).

[W50] *Help: Job queue*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Job_queue (visited on 30/01/2019) (cit. on p. 170).

[W51] *Help: Labeled section transclusion*. Edit as at 12:42, 20 May 2013. Wikipedia.org. 2013. URL: https://en.wikipedia.org/w/index.php?title=Help:%20Labeled_section_transclusion&oldid=555938981 (visited on 07/09/2016) (cit. on p. 82).

[W52] *Help: Labeled section transclusion*. Edit as at 04:03, 18 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Help:Labeled_section_transclusion&oldid=691180442 (visited on 07/09/2016) (cit. on pp. 81, 109, 113).

[W53] *Help: Labeled section transclusion, section 'Function names'*. Edit as at 04:03, 18 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Help:Labeled_section_transclusion&oldid=691180442#Function_names (visited on 07/09/2016) (cit. on pp. 82, 91).

[W54] *Help: Section, section 'Sections vs. separate pages vs. transclusion'*. Edit as at 23:12, 9 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Help:Section&oldid=699048851#Sections_vs._separate_pages_vs._transclusion (visited on 07/09/2016) (cit. on p. 113).

[W55] *Help: Template*. Edit as at 07:24, 29 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Help:Template&oldid=697252556 (visited on 07/09/2016) (cit. on pp. 67, 113).

[W56] *Help: Template, section 'Noinclude, includeonly, and onlyinclude'*. Edit as at 07:24, 29 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Help:Template&oldid=697252556#Noinclude.2C_includeonly.2C_and_onlyinclude (visited on 07/09/2016) (cit. on p. 113).

[W57] *Help: Transclusion/How Transclusion Works*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Transclusion/How_Transclusion_Works (visited on 30/01/2019) (cit. on pp. 77, 257).

[W58] *Help: Watchlist, section 'Options'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Help:Watchlist#Options (visited on 30/01/2019) (cit. on p. 191).

[W59] *History of The Simpsons*. Edit as at 09:50, 11 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=History_of_The_Simpsons&oldid=694756533 (visited on 07/09/2016) (cit. on p. 105).

[W60] *HitRecord*. Edit as at 18:51, 6 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=HitRecord&oldid=698534324 (visited on 10/01/2016) (cit. on p. 109).

[W61] *Interwiki links*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Interwiki_links (visited on 30/01/2019) (cit. on pp. 114, 126, 140, 149, 178–179).

[W62] *Interwiki links, talk page, section 'Wikidata and Interwiki links'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Talk:Interwiki_links#Wikidata_and_Interwiki_links (visited on 30/01/2019) (cit. on p. 153).

[W63] *Italian wiki: 96 a.C.* Edit as at 23:14, 4 February 2014. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=96_a.C.&oldid=63976187 (visited on 07/09/2016) (cit. on p. 99).

[W64] *Italian wiki: Boston Bruins*. Edit as at 22:42, 3 January 2016. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=77591494 (visited on 07/09/2016) (cit. on p. 100).

[W65] *Italian wiki: Manchester United Football Club*. Edit as at 17:22, 7 January 2016. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=Manchester_United_Football_Club&oldid=77831412 (visited on 07/09/2016) (cit. on p. 100).

[W66] *Italian wiki: Morti nel 96 a.C.* Edit as at 13:56, 7 January 2016. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=Morti_nel_96_a.C.&oldid=77815887 (visited on 07/09/2016) (cit. on p. 99).

[W67] *Italian wiki: Nati nel 96 a.C.* Edit as at 13:56, 7 January 2016. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=Nati_nel_96_a.C.&oldid=77815886 (visited on 07/09/2016) (cit. on p. 99).

[W68] *Italian wiki: Speciale: Contributi/GacbotUser (i.e. user Gacbot's, edits)*. Wikipedia.org. 2016. URL: https://it.wikipedia.org/wiki/Speciale:Contributi/Gacbot (visited on 07/09/2016) (cit. on p. 99).

[W69] *Italian wiki: Videogiochi per Commodore 64*. Edit as at 15:41, 7 January 2016. Wikipedia.org. 2016. URL: https://it.wikipedia.org/w/index.php?title=Videogiochi_per_Commodore_64&oldid=77824354 (visited on 07/09/2016) (cit. on p. 99).

[W70] *iZombie (TV series)*. Edit as at 06:39, 31 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=IZombie_(TV_series)&oldid=697553452 (visited on 07/09/2016) (cit. on p. 105).

[W71] *Japanese wiki:* ボストン・ブルーインズ. Edit as at 03:02, 21 February 2015. Wikipedia.org. 2016. URL: https://ja.wikipedia.org/w/index.php?title=%5Cbegin%7BCJK%7D%7BUTF8%7D%7Bmin%7D%E3%83%9C%E3%82%B9%E3%83%88%E3%83%B3%E3%83%BB%E3%83%96%E3%83%AB%E3%83%BC%E3%82%A4%E3%83%B3%E3%82%BA%5Cend%7BCJK%7D&oldid=54490920 (visited on 07/09/2016) (cit. on p. 100).

[W72] *Japanese wiki:* マンチェスター・ユナイテッド*FC*. Edit as at 14:12, 9 January 2016. Wikipedia.org. 2016. URL: https://ja.wikipedia.org/w/index.php?title=%5Cbegin%7BCJK%7D%7BUTF8%7D%7Bmin%7D%E3%83%9E%E3%83%B3%E3%83%81%E3%82%A7%E3%82%B9%E3%82%BF%E3%83%BC%E3%83%BB%E3%83%A6%E3%83%8A%E3%82%A4%E3%83%86%E3%83%83%E3%83%89FC%5Cend%7BCJK%7D&oldid=58186511 (visited on 07/09/2016) (cit. on p. 100).

[W73] *Japanese wiki:* 年度別映画興行成績. Creation edit, as at 05:17, 26 September 2004. Wikipedia.org. 2004. URL: https://ja.wikipedia.org/w/index.php?title=%5Cbegin%7BCJK%7D%7BUTF8%7D%7Bmin%7D%E5%B9%B4%E5%BA%A6%E5%88%A5%E6%98%A0%E7%94%BB%E8%88%88%E8%A1%8C%E6%88%90%E7%B8%BE%5Cend%7BCJK%7D&oldid=820052 (visited on 07/09/2016) (cit. on p. 103).

[W74] *Japanese wiki:* 年度別映画興行成績. Edit as at 04:40, 24 November 2015. Wikipedia.org. 2015. URL: https://ja.wikipedia.org/w/index.php?title=%5Cbegin%7BCJK%7D%7BUTF8%7D%7Bmin%7D%E5%B9%B4%E5%BA%A6%E5%88%A5%E6%98%A0%E7%94%BB%E8%88%88%E8%A1%8C%E6%88%90%E7%B8%BE%5Cend%7BCJK%7D&oldid=57650778 (visited on 07/09/2016) (cit. on p. 103).

[W75] *Japanese wiki:* 年度別映画興行成績. Edit as at 09:07, 30 December 2015. Wikipedia.org. 2016. URL: https://ja.wikipedia.org/w/index.php?title=%5Cbegin%7BCJK%7D%7BUTF8%7D%7Bmin%7D%E5%B9%B4%E5%BA%A6%E5%88%A5%E6%98%A0%E7%94%BB%E8%88%88%E8%A1%8C%E6%88%90%E7%B8%BE%5Cend%7BCJK%7D&oldid=58063614 (visited on 07/09/2016) (cit. on p. 103).

[W76] *Joseph Gordon-Levitt, section: HitRecord*. Edit as at 07:48, 10 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Joseph_Gordon-Levitt&oldid=699104576#HitRecord (visited on 10/01/2016) (cit. on p. 109).

[W77] *List of American television series*. Edit as at 12:50, 10 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_American_television_series&oldid=699130409 (visited on 07/09/2016) (cit. on p. 96).

[W78] *List of British television programmes*. Edit as at 02:06, 28 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_British_television_programmes&oldid=697072088 (visited on 07/09/2016) (cit. on p. 96).

[W79] *List of Criminal Minds characters*. Edit as at 14:33, 28 April 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Criminal_Minds_characters&oldid=717575300 (visited on 07/09/2016) (cit. on p. 103).

[W80] *List of Criminal Minds episodes*. Edit as at 23:35, 28 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Criminal_Minds_episodes&oldid=697202954 (visited on 07/09/2016) (cit. on p. 103).

[W81] *List of devices using Mediatek tablet processors*. The Template's page has since been deleted and can no longer be accessed. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_devices_using_Mediatek_tablet_processors&oldid=691189367 (visited on 07/09/2016) (cit. on p. 107).

[W82] *List of heads of diplomats of the United Kingdom to Germany*. Edit as at 09:41, 7 January 2016, in page edit mode. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_diplomats_of_the_United_Kingdom_to_Germany&action=edit&oldid=698637349 (visited on 07/09/2016) (cit. on p. 102).

[W83] *List of heads of missions of the United Kingdom*. Edit as at 22:13, 11 August 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_heads_of_missions_of_the_United_Kingdom&oldid=690855670 (visited on 07/09/2016) (cit. on p. 102).

[W84] *List of heads of missions of the United Kingdom*. Edit as at 02:54, 16 November 2015, in page edit mode. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_heads_of_missions_of_the_United_Kingdom&action=edit&oldid=690855670 (visited on 07/09/2016) (cit. on p. 102).

[W85] *List of iZombie episodes*. Edit as at 05:06, 7 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_iZombie_episodes&oldid=698611749 (visited on 07/09/2016) (cit. on p. 105).

[W86] *List of longest-running U.S. broadcast network television series*. Edit as at 19:06, 25 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_longest-running_U.S._broadcast_network_television_series&oldid=696771006 (visited on 07/09/2016) (cit. on p. 105).

[W87] *List of longest-running U.S. primetime television series*. Edit as at 12:01, 11 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_longest-running_U.S._primetime_television_series&oldid=699285724 (visited on 07/09/2016) (cit. on p. 105).

[W88] *List of longest-running United States television series*. Edit as at 12:07, 11 January 201. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_longest-running_United_States_television_series&oldid=699286317 (visited on 07/09/2016) (cit. on p. 105).

[W89] *List of minor planets*. Edit as at 02:25, 31 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_minor_planets&oldid=697528627 (visited on 07/09/2016) (cit. on p. 96).

[W90] *List of minor planets: 147001–148000*. Edit as at 23:39, 11 April 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_minor_planets:_147001%E2%80%93148000&oldid=656041150 (visited on 07/09/2016) (cit. on p. 103).

[W91] *List of most popular websites*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/List_of_most_popular_websites (visited on 12/08/2019) (cit. on p. 7).

[W92] *List of Polish gminas*. Edit as at 20:08, 6 January 2013. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Polish_gminas&oldid=531667874 (visited on 07/09/2016) (cit. on pp. 94, 106).

[W93] *List of Polish gminas (A)*. Edit as at 02:37, 21 October 2012. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Polish_gminas_(A)&oldid=518960535 (visited on 07/09/2016) (cit. on p. 106).

[W94] *List of Polish gminas (R)*. Edit as at 20:51, 18 April 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Polish_gminas_(R)&oldid=604791608 (visited on 07/09/2016) (cit. on p. 107).

[W95] *List of Talkin' 'Bout Your Generation episodes*. Edit as at 15:28, 29 August 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Talkin%27_%27Bout_Your_Generation_episodes&oldid=678464189 (visited on 07/09/2016) (cit. on p. 106).

[W96] *List of Talkin' 'Bout Your Generation episodes (series 1)*. Edit as at 22:44, 29 August 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Talkin%27_%27Bout_Your_Generation_episodes_(series_1)&oldid=678524887 (visited on 07/09/2016) (cit. on p. 106).

[W97] *List of Talkin' 'Bout Your Generation episodes (series 2)*. Edit as at 14:43, 15 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Talkin%27_%27Bout_Your_Generation_episodes_(series_2)&oldid=685872015 (visited on 07/09/2016) (cit. on p. 106).

[W98] *List of Talkin' 'Bout Your Generation episodes (series 3)*. Edit as at 14:43, 15 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Talkin%27_%27Bout_Your_Generation_episodes_(series_3)&oldid=681147835 (visited on 07/09/2016) (cit. on p. 106).

[W99] *List of Talkin' 'Bout Your Generation episodes (series 4)*. Edit as at 22:43, 29 August 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_Talkin%27_%27Bout_Your_Generation_episodes_(series_4)&oldid=678524807 (visited on 07/09/2016) (cit. on p. 106).

[W100] *List of the Penguins of Madagascar episodes*. Edit as at 20:55, 23 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_The_Penguins_of_Madagascar_episodes&oldid=696535021 (visited on 07/09/2016) (cit. on p. 104).

[W101] *List of The Simpsons episodes*. Edit as at 22:44, 11 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_The_Simpsons_episodes&oldid=699371338 (visited on 07/09/2016) (cit. on p. 105).

[W102] *List of The Simpsons guest stars*. Edit as at 09:36, 8 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=List_of_The_Simpsons_guest_stars&oldid=698801871 (visited on 07/09/2016) (cit. on p. 105).

[W103] *List of Wikipedias*. Non-logged report, archived by the Internet Archive as at 10 January 2016, 15:08:49. Wikipedia.org. 2019. URL: https://web.archive.org/web/20160110150849/https://en.wikipedia.org/wiki/List_of_Wikipedias (visited on 30/01/2019) (cit. on p. 69).

[W104] *List of Wikipedias, section 'Detailed list'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/List_of_Wikipedias#Detailed_list (visited on 07/09/2016) (cit. on p. 85).

[W105] *List of Wikipedias, section 'Detailed list'*. Non-logged report, archived by the Internet Archive as at 10 January 2016, 15:08:49. Wikipedia.org. 2019. URL: https://web.archive.org/web/20160110150849/https://en.wikipedia.org/wiki/List_of_Wikipedias#Detailed_list (visited on 30/01/2019) (cit. on pp. 70, 85, 179).

[W106] *List of Wikipedias, section 'List'*. Non-logged report, archived by the Internet Archive as at 10 January 2016, 15:08:49. Wikipedia.org. 2019. URL: https://web.archive.org/web/20160110150849/https://en.wikipedia.org/wiki/List_of_Wikipedias#List (visited on 30/01/2019) (cit. on p. 69).

[W107] *Lsjbot - article about user 'Lsjbot'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Lsjbot (visited on 07/09/2016) (cit. on p. 70).

[W108] *Lua*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Lua_(programming_language) (visited on 30/01/2019) (cit. on p. 67).

[W109] *Manchester United F.C.* Edit as at 19:06, 9 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Manchester_United_F.C.&oldid=699017841 (visited on 07/09/2016) (cit. on p. 100).

[W110] *Matt Groening*. Edit as at 05:27, 9 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Matt_Groening&oldid=698938040 (visited on 07/09/2016) (cit. on p. 105).

[W111] *MediaWiki Wikimedia Developer Support forum thread: 'Rate limit clarifications'*. Wikimedia Labs. 2018. URL: https://discourse-mediawiki.wmflabs.org/t/rate-limit-clarifications/714/6 (visited on 09/07/2019) (cit. on p. 64).

[W112] *Mediawiki Wikimedia Developer Support: API Etiquette and Request Limits*. Wikipedia.org. 2018. URL: https://www.mediawiki.org/wiki/API:Etiquette#Request_limit (visited on 09/07/2019) (cit. on p. 64).

[W113] *MediaWiki: Manual: User Rights*. MediaWiki.org. 2019. URL: https://www.mediawiki.org/wiki/Manual:User_rights#List_of_permissions (visited on 30/01/2019) (cit. on p. 126).

[W114] *MediaWiki: Special: r9367 - Code Review*. MediaWiki.org. 2019. URL: https://www.mediawiki.org/wiki/Special:Code/MediaWiki/9367 (visited on 30/01/2019) (cit. on p. 126).

[W115] *MediaWiki's Extension: Labeled Section Transclusion*. Edit as at 17:57, 3 November 2006. MediaWiki.org. 2006. URL: https://www.mediawiki.org/w/index.php?title=Extension:Labeled_Section_Transclusion&oldid=58952 (visited on 07/09/2016) (cit. on p. 82).

[W116] *MediaWiki's Extension: Labeled Section Transclusion*. Edit as at 09:50, 10 September 2015. MediaWiki.org. 2016. URL: https://www.mediawiki.org/w/index.php?title=Extension:Labeled_Section_Transclusion&oldid=1878674 (visited on 07/09/2016) (cit. on p. 113).

[W117] *MediaWiki's Extension: Labeled section transclusion, section 'Localisation'*. Edit as at 09:50, 10 September 2015. MediaWiki.org. 2016. URL: https://www.mediawiki.org/w/index.php?title=Extension:Labeled_Section_Transclusion&oldid=1878674#Localisation (visited on 07/09/2016) (cit. on p. 81).

[W118] *MediaWiki's Help: Namespaces, section 'Renaming namespaces'*. Edit as at 01:34, 23 December 2015. MediaWiki.org. 2016. URL: https://www.mediawiki.org/w/index.php?title=Help:Namespaces&oldid=1973970#Renaming_namespaces (visited on 07/09/2016) (cit. on p. 85).

[W119] *Metawiki: Bot*. Wikipedia.org. 2019. URL: https://meta.wikimedia.org/wiki/Bot (visited on 30/01/2019) (cit. on p. 122).

[W120] *Metawiki: Data dumps/FAQ*. Wikimedia Foundation. 2018. URL: https://meta.wikimedia.org/wiki/Data_dumps/FAQ (visited on 07/09/2018) (cit. on p. 69).

[W121] *Metawiki: Research FAQ*. Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/Research:FAQ (visited on 30/01/2019) (cit. on p. 63).

[W122] *Metawiki: Research FAQ, 'Does my project need approval?'* Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/Research:FAQ#Does_my_project_need_approval? (visited on 30/01/2019) (cit. on p. 63).

[W123] *Metawiki: Research: Resources*. Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/Research:Resources (visited on 30/01/2019) (cit. on p. 63).

[W124] *Metawiki: Steward requests/Bot status/2006-03*. Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/Steward_requests/Bot_status/2006-03#en:User:Pegasusbot (visited on 30/01/2019) (cit. on p. 127).

[W125] *Metawiki: Steward requests/Permissions/2004*. Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/Steward_requests/Permissions/2004#February (visited on 30/01/2019) (cit. on p. 126).

[W126] *Metawiki: Toolhub*. Edit as at 07:46, 13 October 2018. Wikipedia.org. 2019. URL: https://meta.wikimedia.org/w/index.php?title=Toolhub&oldid=18466433 (visited on 30/01/2019) (cit. on p. 170).

[W127] *Metawiki: What's available for download*. Wikimedia Foundation. 2018. URL: https://meta.wikimedia.org/wiki/Data_dumps/What%27s_available_for_download (visited on 07/09/2018) (cit. on pp. 65, 69).

[W128] *Metawiki: Wikipedia article depth*. Wikimedia Foundation. 2016. URL: https://meta.wikimedia.org/wiki/Wikipedia_article_depth (visited on 08/03/2016) (cit. on p. 85).

[W129] *Metawiki: WikiXRay*. Wikimedia Foundation. 2019. URL: https://meta.wikimedia.org/wiki/WikiXRay (visited on 30/01/2019) (cit. on p. 63).

[W130] *Module: Episode list*. Edit as at 14:19, 21 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Module:Episode_list&oldid=696188778 (visited on 07/09/2016) (cit. on p. 104).

[W131] *Pardon of Nixon*. Edit at 06:12, 1 February 2016. Originally sampled 27 July 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Pardon_of_Richard_Nixon&oldid=702709814 (visited on 07/09/2016) (cit. on p. 94).

[W132] *Pardon of Nixon*. Sampled 27 July 2016, when most recent edit was as at 16:58, 23 July 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Pardon_of_Richard_Nixon&oldid=728785009 (visited on 07/09/2016) (cit. on p. 94).

[W133] *Pareto principle*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Pareto_principle (visited on 30/01/2019) (cit. on p. 178).

[W134] *Pathology*. Edit as at 23 April 2008. Wikipedia.org. 2008. URL: https://en.wikipedia.org/w/index.php?title=Pathology&oldid=207669487 (visited on 07/09/2016) (cit. on p. 110).

[W135] *Pathology*. Edit as at 23 April 2008, edit view. Wikipedia.org. 2008. URL: https://en.wikipedia.org/w/index.php?title=Pathology&action=edit&oldid=207669487 (visited on 07/09/2016) (cit. on p. 110).

[W136] *Phabricator blog: Labs and Tool Labs being renamed*. Wikipedia.org. 2019. URL: https://phabricator.wikimedia.org/phame/post/view/59/labs_and_tool_labs_being_renamed/ (visited on 30/01/2019) (cit. on p. 170).

[W137] *Research*. Wikimedia Foundation. 2019. URL: https://wikimediafoundation.org/research/ (visited on 30/01/2019) (cit. on p. 63).

[W138] *Russian wiki:* Бостон Брюинз. Edit as at 13:43, 20 December 2015. Wikipedia.org. 2016. URL: https://ru.wikipedia.org/w/index.php?title=%5Cbegin%7Botherlanguage*%7D%7Brussian%7D%D0%91%D0%BE%D1%81%D1%82%D0%BE%D0%BD_%D0%91%D1%80%D1%8E%D0%B8%D0%BD%D0%B7%5Cend%7Botherlanguage*%7D&oldid=75211807 (visited on 07/09/2016) (cit. on p. 100).

[W139] *Russian wiki:* Манчестер Юнаитед. Edit as at 12:05, 9 January 2016. Wikipedia.org. 2016. URL: https://ru.wikipedia.org/w/index.php?title=%5Cbegin%7Botherlanguage*%7D%7Brussian%7D%D0%9C%D0%B0%D0%BD%D1%87%D0%B5%D1%81%D0%B8%D1%82%D0%B5%D1%80%20%D0%AE%D0%BD%D0%B0%D1%82%D0%B5%D0%B4%5Cend%7Botherlanguage*%7D&oldid=75647293 (visited on 07/09/2016) (cit. on p. 100).

[W140] *Sockpuppet (Internet)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Sockpuppet_(Internet) (visited on 30/01/2019) (cit. on p. 399).

[W141] *Songs of Anarchy: Music from Sons of Anarchy Seasons 1–4*. Edit as at 17:03, 24 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Songs_of_Anarchy:_Music_from_Sons_of_Anarchy_Seasons_1%E2%80%934&oldid=692274939 (visited on 07/09/2016) (cit. on p. 78).

[W142] *Spanish wiki: Boston Bruins*. Edit as at 07:32 16 Jun 2015. Wikipedia.org. 2016. URL: https://es.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=83218095 (visited on 07/09/2016) (cit. on p. 100).

[W143] *Spanish wiki: Manchester United Football Club*. Edit as at 16:25, 11 January 2016. Wikipedia.org. 2016. URL: https://es.wikipedia.org/w/index.php?title=Manchester_United_Football_Club&oldid=88372991 (visited on 07/09/2016) (cit. on p. 100).

[W144] *Special: (User) contributions*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Special:Contributions (visited on 07/09/2016) (cit. on p. 264).

[W145] *Special: Contributions/KonstableBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/KonstableBot (visited on 30/01/2019) (cit. on p. 138).

[W146] *Special: Global accounts list: Global bots*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:GlobalUsers?username=&group=global-bot&limit=500 (visited on 30/01/2019) (cit. on pp. 127, 132).

[W147] *Special: ListUsers/bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:ListUsers/bot (visited on 30/01/2019) (cit. on pp. 121, 129, 132, 178).

[W148] *Special: ListUsers/bot (500 items per page)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:ListUsers/bot&offset=&limit=500&group=bot (visited on 30/01/2019) (cit. on pp. 130, 132).

[W149] *Special: Log entries for user Zbot370*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:Log&type=&user=&page=User:Zbot370 (visited on 30/01/2019) (cit. on p. 127).

[W150] *Special: Most Transcluded Pages*. Data is as from 03:38, 28 April 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Special:MostTranscludedPages (visited on 30/04/2016) (cit. on p. 94).

[W151] *Special: Pages with badges*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Special:PagesWithBadges (visited on 07/09/2016) (cit. on p. 264).

[W152] *Special: Public Logs - filtered for User:RambotRambot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:Log&type=&user=&page=User:Rambot (visited on 30/01/2019) (cit. on p. 126).

[W153] *Special: Statistics*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Special:Statistics (visited on 07/09/2016) (cit. on pp. 262–263).

[W154] *Special: User logs for User:Δbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:%CE%94bot&redirect=no (visited on 30/01/2019) (cit. on p. 141).

[W155] *Special: User logs for User:DarafshBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Log?type=&user=&page=User%3ADarafshBot&wpdate=&tagfilter= (visited on 30/01/2019) (cit. on p. 141).

[W156] *Special: User logs for User:Lancbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Log?type=&user=&page=User%3ALancbot&wpdate=&tagfilter= (visited on 30/01/2019) (cit. on p. 140).

[W157] *Special: User logs for User:Lt-wiki-bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:Log&type=&user=&page=User:Lt-wiki-bot&year=2017&month=6 (visited on 30/01/2019) (cit. on pp. 140–141).

[W158] *Special: User logs for User:Shadowbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Shadowbot~enwiki&action=history (visited on 30/01/2019) (cit. on p. 136).

[W159] *Special: User Rambot - User Contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:Contributions/Rambot&target=Rambot (visited on 30/01/2019) (cit. on p. 125).

[W160] *Special: Users filtered for 'Importers'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Special%3AListUsers&username=&group=import&limit=500 (visited on 07/09/2016) (cit. on p. 262).

[W161] *Special: What Links Here*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:WhatLinksHere (visited on 30/01/2019) (cit. on p. 155).

[W162] *Special: What Links Here/The Simpsons (only showing transcludes)*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Special:WhatLinksHere/The_Simpsons&hidelinks=1&hideredirs=1 (visited on 20/07/2016) (cit. on p. 108).

[W163] *Special: What Links Here/The Simpsons';* Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Special:WhatLinksHere/The_Simpsons (visited on 20/07/2016) (cit. on p. 108).

[W164] *Special: WhatLinksHere/Trace amine*. 'Trace Amine' is transcluded by 'monoamine neurotransmitter' and 'Biogenic amine'. Note: special pages are not cached or saved in dumps. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Special:WhatLinksHere/Trace_amine&hidelinks=1&hideredirs=1 (visited on 18/07/2016) (cit. on p. 113).

[W165] *Swedish wiki: Boston Bruins*. Edit as at 00:12, 6 January 2016. Wikipedia.org. 2016. URL: https://sv.wikipedia.org/w/index.php?title=Boston_Bruins&oldid=32261476 (visited on 07/09/2016) (cit. on p. 100).

[W166] *Swedish wiki: Manchester United FC*. Edit as at 03:12, 29 December 2015. Wikipedia.org. 2016. URL: https://sv.wikipedia.org/w/index.php?title=Manchester_United_FC&oldid=31973956 (visited on 07/09/2016) (cit. on p. 100).

[W167] *Talk: List of heads of missions of the United Kingdom*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Talk:List_of_heads_of_missions_of_the_United_Kingdom (visited on 07/09/2016) (cit. on p. 103).

[W168] *Talk: List of minor planets*. Edit as at 20:53, 21 September 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Talk:List_of_minor_planets&oldid=682147806 (visited on 07/09/2016) (cit. on p. 103).

[W169] *Talk: Pathology*. Article's talk page section 'Experimental transclusion technique', comment dated 20 July 2009. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Talk:Pathology#Experimental_transclusion_technique (visited on 07/09/2016) (cit. on pp. 94, 111).

[W170] *Talkin' 'Bout Your Generation*. Edit as at 10:50, 24 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Talkin%27_%27Bout_Your_Generation&oldid=687260359 (visited on 07/09/2016) (cit. on p. 106).

[W171] *Ted Nelson*. Edit as at 07:27, 31 January 2016, in edit mode. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Ted_Nelson&action=edit&oldid=702551651&editintro=Template:BLP_editintro (visited on 07/09/2016) (cit. on p. 108).

[W172] *Template: BAG Tools*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:BAG_Tools (visited on 30/01/2019) (cit. on p. 132).

[W173] *Template: Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Bot (visited on 30/01/2019) (cit. on p. 139).

[W174] *Template: Bots*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Bots (visited on 30/01/2019) (cit. on p. 191).

[W175] *Template: Country data United States*. Edit as at 05:29, 4 September 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Country_data_United_States&oldid=679376046 (visited on 07/09/2016) (cit. on p. 94).

[W176] *Template: Episode list*. Edit as at 02:06, 8 July 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Episode_list&oldid=616030121 (visited on 07/09/2016) (cit. on p. 104).

[W177] *Template: Episode list/sublist*. Edit as at 02:15, 8 July 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Episode_list/sublist&oldid=616030747 (visited on 07/09/2016) (cit. on p. 104).

[W178] *Template: FA-star discussion*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Templates_for_discussion/Log/2010_May_9#Template:FA-star (visited on 30/01/2019) (cit. on p. 164).

[W179] *Template: FIFA Club World Cup winners*. Edit as at 07:44, 21 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:FIFA_Club_World_Cup_winners&oldid=696155763 (visited on 07/09/2016) (cit. on p. 100).

[W180] *Template: Football in Greater Manchester*. Edit as at 03:19, 26 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Football_in_Greater_Manchester&oldid=687525537 (visited on 07/09/2016) (cit. on p. 101).

[W181] *Template: Infobox Wikipedia bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Infobox_Wikipedia_bot (visited on 30/01/2019) (cit. on pp. 135, 138).

[W182] *Template: List of Polish gminas*. The Template's page has since been deleted and can no longer be accessed. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:List_of_Polish_gminas&oldid=621930896 (visited on 07/09/2016) (cit. on p. 107).

[W183] *Template: Manchester United F.C.* Edit as at 16:30, 3 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Manchester_United_F.C.&oldid=688880163 (visited on 07/09/2016) (cit. on p. 101).

[W184] *Template: Notability*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Notability (visited on 30/01/2019) (cit. on p. 168).

[W185] *Template: Onlyinclude*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Template:Onlyinclude (visited on 07/09/2016) (cit. on p. 103).

[W186] *Template: Premier League*. Edit as at 17:04, 6 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Premier_League&oldid=689360575 (visited on 07/09/2016) (cit. on p. 101).

[W187] *Template: Retired*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Retired (visited on 30/01/2019) (cit. on pp. 136, 140).

[W188] *Template: Start date*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Template:Start_date (visited on 30/01/2019) (cit. on p. 166).

[W189] *Template: Transcluded*. Edit as at 17:13, 27 January 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Transcluded&oldid=644427855 (visited on 07/09/2016) (cit. on p. 93).

[W190] *Template: Transcluded section*. Edit as at 18:41, 3 November 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:Transcluded_section&oldid=632312487 (visited on 07/09/2016) (cit. on p. 93).

[W191] *Template: Transcluded section*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Transcluded_section (visited on 30/01/2019) (cit. on p. 83).

[W192] *Template: Transcluding article*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Template:Transcluding_article (visited on 30/01/2019) (cit. on p. 83).

[W193] *Template: UEFA Champions League winners*. Edit as at 21:21, 31 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:UEFA_Champions_League_winners&oldid=697643077 (visited on 07/09/2016) (cit. on p. 101).

[W194] *Template: UEFA Cup Winners' Cup winners*. Edit as at 17:50, 3 November 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:UEFA_Cup_Winners%27_Cup_winners&oldid=632306239 (visited on 07/09/2016) (cit. on p. 101).

[W195] *Template: UEFA Super Cup winners*. Edit as at 22:13, 11 August 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Template:UEFA_Super_Cup_winners&oldid=675660514 (visited on 07/09/2016) (cit. on p. 101).

[W196] *The Penguins of Madagascar (season 1)*. Edit as at 15:10, 24 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=The_Penguins_of_Madagascar_(season_1)&oldid=692248708 (visited on 07/09/2016) (cit. on p. 104).

[W197] *The Penguins of Madagascar (season 2)*. Edit as at 09:12, 1 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=The_Penguins_of_Madagascar_(season_2)&oldid=697706866 (visited on 07/09/2016) (cit. on p. 104).

[W198] *The Penguins of Madagascar (season 3)*. Edit as at 08:08, 3 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=The_Penguins_of_Madagascar_(season_3)&oldid=697989906 (visited on 07/09/2016) (cit. on p. 104).

[W199] *The Simpsons*. Edit as at 01:27, 11 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=The_Simpsons&oldid=699225461 (visited on 07/09/2016) (cit. on p. 105).

[W200] *The Simpsons' October 2015*. Edit as at 05:25, 24 October 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=The_Simpsons&oldid=687233629 (visited on 07/09/2016) (cit. on p. 105).

[W201] *The Turk*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/The_Turk (visited on 30/01/2019) (cit. on pp. 181, 186).

[W202] *Transclusion*. Edit as at 15:25, 14 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Transclusion&oldid=695215728 (visited on 07/09/2016) (cit. on p. 109).

[W203] *User talk: Anomie*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Anomie (visited on 30/01/2019) (cit. on p. 194).

[W204] *User talk: CanisRufus#Bot name*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:CanisRufus#Bot_name (visited on 30/01/2019) (cit. on p. 133).

[W205] *User talk: Kleinzach (Archive page 34): section 'Diplomat notability'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Kleinzach/Archive_34#Diplomat_notability (visited on 30/01/2019) (cit. on p. 168).

[W206] *User talk: Kleinzach (Archive page 35): section 'RileyBot'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Kleinzach/Archive_35#RileyBot (visited on 30/01/2019) (cit. on p. 168).

[W207] *User talk: Konstable#KonstableBot-deflagged*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Konstable#KonstableBot_deflagged (visited on 30/01/2019) (cit. on p. 138).

[W208] *User talk: Mobius Clock topic 'Bot amok'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Mobius_Clock#Bot_amok (visited on 30/01/2019) (cit. on p. 138).

[W209] *User talk: riley (Archive page 4)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:~riley/Archive_4 (visited on 30/01/2019) (cit. on p. 168).

[W210] *User talk: TedderBot*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/User_talk:TedderBot (visited on 30/01/2019) (cit. on p. 156).

[W211] *User talk: Tim1357 (Archive page 3): section 'Taking over the bot request where Betacommand left off?*' Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Tim1357/Archive_3#Taking_over_the_bot_request_where_Betacommand_left_off%3F (visited on 30/01/2019) (cit. on p. 168).

[W212] *User talk: Xenobot Mk*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/User_talk:Xenobot_Mk_V/requests#WP:USFILMS (visited on 30/01/2019) (cit. on p. 165).

[W213] *User: Δ*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:$%5CDelta$ (visited on 30/01/2019) (cit. on pp. 160, 177).

[W214] *User: 718 Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:718%20Bot (visited on 30/01/2019) (cit. on p. 173).

[W215] *User: A-lú-mih-bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:A-l%C3%BA-mih-bot (visited on 30/01/2019) (cit. on p. 135).

[W216] *User: Addbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Addbot (visited on 30/01/2019) (cit. on p. 153).

[W217] *User: AFD Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AFD_Bot (visited on 30/01/2019) (cit. on p. 150).

[W218] *User: AfDStatBot (edit current during 2014)*. Edit as at 14:04, 26 October 2013. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:AfDStatBot&oldid=578824278 (visited on 30/01/2019) (cit. on p. 125).

[W219] *User: AlexNewArtBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AlexNewArtBot (visited on 30/01/2019) (cit. on p. 142).

[W220] *User: AllyUnion*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AllyUnion (visited on 30/01/2019) (cit. on pp. 158–159).

[W221] *User: Anchor Link Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Anchor_Link_Bot (visited on 30/01/2019) (cit. on p. 167).

[W222] *User: Android Mouse Bot 2*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Android_Mouse_Bot_2 (visited on 30/01/2019) (cit. on p. 167).

[W223] *User: Android Mouse Bot 3*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Android_Mouse_Bot_3 (visited on 30/01/2019) (cit. on p. 167).

[W224] *User: Android Mouse Bot 4*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Android_Mouse_Bot_4 (visited on 30/01/2019) (cit. on p. 167).

[W225] *User: AnomiBOT*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AnomieBOT (visited on 30/01/2019) (cit. on pp. 67, 127, 152, 176).

[W226] *User: Anomie*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Anomie (visited on 30/01/2019) (cit. on p. 176).

[W227] *User: AnomieBOT II*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AnomieBOT_II (visited on 30/01/2019) (cit. on p. 176).

[W228] *User: AnomieBOT III*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AnomieBOT_III (visited on 30/01/2019) (cit. on p. 176).

[W229] *User: AntiSpamBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AntiSpamBot (visited on 30/01/2019) (cit. on pp. 136, 150).

[W230] *User: AntiSpamBot edit history*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:AntiSpamBot&action=history (visited on 30/01/2019) (cit. on p. 136).

[W231] *User: AntiVandalBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:AntiVandalBot (visited on 30/01/2019) (cit. on pp. 141, 161).

[W232] *User: Anybot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Anybot (visited on 30/01/2019) (cit. on p. 137).

[W233] *User: Avicennasis*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Avicennasis (visited on 30/01/2019) (cit. on p. 163).

[W234] *User: Betacommand*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Betacommand (visited on 30/01/2019) (cit. on p. 160).

[W235] *User: BetacommandBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:BetacommandBot (visited on 30/01/2019) (cit. on pp. 151, 173).

[W236] *User: Bgbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Bgbot (visited on 30/01/2019) (cit. on p. 126).

[W237] *User: Carnildo*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Carnildo (visited on 30/01/2019) (cit. on p. 135).

[W238] *User: ClueBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ClueBot (visited on 30/01/2019) (cit. on p. 152).

[W239] *User: ClueBot II*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ClueBot_II (visited on 30/01/2019) (cit. on p. 173).

[W240] *User: ClueBot NG*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ClueBot_NG (visited on 30/01/2019) (cit. on pp. 123, 141, 152, 179, 196, 199).

[W241] *User: Cobi*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Cobi (visited on 30/01/2019) (cit. on pp. 131, 141).

[W242] *User: Crispy1989*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Crispy1989 (visited on 30/01/2019) (cit. on pp. 141, 179).

[W243] *User: Dake enwiki*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Dake~enwiki (visited on 30/01/2019) (cit. on p. 140).

[W244] *User: DarafshBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DarafshBot (visited on 30/01/2019) (cit. on p. 141).

[W245] *User: DASHBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DASHBot (visited on 30/01/2019) (cit. on p. 155).

[W246] *User: DavidRF*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DavidRF (visited on 30/01/2019) (cit. on p. 162).

[W247] *User: DeltaBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DeltaBot (visited on 30/01/2019) (cit. on p. 141).

[W248] *User: DinoSizedBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DinoSizedBot (visited on 30/01/2019) (cit. on p. 167).

[W249] *User: Dispenser*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Dispenser (visited on 30/01/2019) (cit. on pp. 155, 164).

[W250] *User: DumZiBoT*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:DumZiBoT (visited on 30/01/2019) (cit. on p. 160).

[W251] *User: Dycedarg*. Wikipedia.org. 2019. URL: User:%20https://en.wikipedia.org/wiki/User:Dycedarg (visited on 30/01/2019) (cit. on p. 177).

[W252] *User: EarwigBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:EarwigBot (visited on 30/01/2019) (cit. on p. 142).

[W253] *User: ENewsBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ENewsBot (visited on 30/01/2019) (cit. on p. 136).

[W254] *User: FastilyBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:FastilyBot (visited on 30/01/2019) (cit. on p. 142).

[W255] *User: Ganeshbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Ganeshbot (visited on 30/01/2019) (cit. on p. 142).

[W256] *User: GoingBatty*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:GoingBatty (visited on 30/01/2019) (cit. on p. 176).

[W257] *User: GoneAwayNowAndRetired*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:GoneAwayNowAndRetired (visited on 30/01/2019) (cit. on pp. 160, 192).

[W258] *User: GoneAwayNowAndRetired, article 'Wikipedia is broken and failing'*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/User:GoneAwayNowAndRetired/Wikipedia_is_broken_and_failing (visited on 30/01/2019) (cit. on pp. 160, 192).

[W259] *User: Grutness*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Grutness (visited on 30/01/2019) (cit. on p. 159).

[W260] *User: HagermanBot*. Wikipedia.org. 2018. URL: https://en.wikipedia.org/wiki/User:HagermanBot (visited on 07/09/2018) (cit. on p. 273).

[W261] *User: Hasteur*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Hasteur (visited on 30/01/2019) (cit. on p. 164).

[W262] *User: HBC AIV Helperbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:HBC_AIV_Helperbot (visited on 30/01/2019) (cit. on p. 123).

[W263] *User: HBC AIV helperbot5*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:HBC_AIV_helperbot5 (visited on 30/01/2019) (cit. on p. 153).

[W264] *User: Helpful Pixie Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Helpful_Pixie_Bot (visited on 30/01/2019) (cit. on pp. 123, 152).

[W265] *User: HerculeBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:HerculeBot (visited on 30/01/2019) (cit. on p. 136).

[W266] *User: HerculeBot*. Edit as at 22:06, 25 July 2015. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:HerculeBot&oldid=673072955 (visited on 30/01/2019) (cit. on p. 136).

[W267] *User: HersfoldArbClerkBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:HersfoldArbClerkBot (visited on 30/01/2019) (cit. on p. 139).

[W268] *User: HiW-Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:HiW-Bot (visited on 30/01/2019) (cit. on p. 136).

[W269] *User: Jayden54Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Jayden54Bot (visited on 30/01/2019) (cit. on p. 140).

[W270] *User: JerryBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:JerryBot (visited on 30/01/2019) (cit. on p. 140).

[W271] *User: JJMC89 bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:JJMC89_bot (visited on 30/01/2019) (cit. on p. 142).

[W272] *User: JLaTondre*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:JLaTondre (visited on 30/01/2019) (cit. on p. 162).

[W273] *User: Joe's Null Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Joe's_Null_Bot (visited on 30/01/2019) (cit. on p. 147).

[W274] *User: John Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:John_Bot (visited on 30/01/2019) (cit. on p. 138).

[W275] *User: John Bot*. Edit as at 14:07, 13 January 2009. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:John_Bot&oldid=263805147 (visited on 30/01/2019) (cit. on p. 138).

[W276] *User: John Bot II*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:John_Bot_II (visited on 30/01/2019) (cit. on p. 138).

[W277] *User: John Bot III*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:John_Bot_III (visited on 30/01/2019) (cit. on p. 138).

[W278] *User: John Bot IV*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:John_Bot_IV (visited on 30/01/2019) (cit. on p. 138).

[W279] *User: John Bot, edit history*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:John_Bot&action=history (visited on 30/01/2019) (cit. on p. 138).

[W280] *User: John Bot, user contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/John_Bot (visited on 30/01/2019) (cit. on p. 138).

[W281] *User: Josh Parris*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Josh_Parris (visited on 30/01/2019) (cit. on p. 172).

[W282] *User: Josvebot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Josvebot (visited on 30/01/2019) (cit. on p. 139).

[W283] *User: Jumbuck*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Jumbuck (visited on 30/01/2019) (cit. on p. 140).

[W284] *User: JYBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:JYBot (visited on 30/01/2019) (cit. on p. 139).

[W285] *User: KamikazeBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:KamikazeBot (visited on 30/01/2019) (cit. on p. 139).

[W286] *User: KarlsenBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:KarlsenBot (visited on 30/01/2019) (cit. on p. 139).

[W287] *User: Kleinzach*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Kleinzach (visited on 30/01/2019) (cit. on p. 168).

[W288] *User: Konstable*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User_talk:Konstable (visited on 30/01/2019) (cit. on p. 138).

[W289] *User: KonstableBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:KonstableBot (visited on 30/01/2019) (cit. on p. 138).

[W290] *User: Kumi-Taskbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 137).

[W291] *User: Kumioko*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Kumioko (visited on 30/01/2019) (cit. on p. 160).

[W292] *User: Lancbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Lancbot (visited on 30/01/2019) (cit. on p. 140).

[W293] *User: LaraBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:LaraBot (visited on 30/01/2019) (cit. on p. 139).

[W294] *User: Le Pied-bot enwiki*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Le_Pied-bot~enwiki (visited on 30/01/2019) (cit. on p. 140).

[W295] *User: Legobot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Legobot (visited on 30/01/2019) (cit. on pp. 136, 142).

[W296] *User: Lightbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Lightbot (visited on 30/01/2019) (cit. on p. 140).

[W297] *User: Lightmouse*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Lightmouse (visited on 30/01/2019) (cit. on p. 156).

[W298] *User: ListeriaBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ListeriaBot (visited on 30/01/2019) (cit. on p. 139).

[W299] *User: Lt-wiki-bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Lt-wiki-bot (visited on 30/01/2019) (cit. on p. 140).

[W300] *User: Lucasbfrbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Lucasbfrbot (visited on 30/01/2019) (cit. on p. 140).

[W301] *User: Luuvabot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Luuvabot&redirect=no (visited on 30/01/2019) (cit. on p. 135).

[W302] *User: Magioladitis*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Magioladitis (visited on 30/01/2019) (cit. on pp. 152, 158, 162).

[W303] *User: MandelBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MandelBot (visited on 30/01/2019) (cit. on p. 164).

[W304] *User: MarshBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MarshBot (visited on 30/01/2019) (cit. on p. 140).

[W305] *User: MarshBot*. Edit as at 01:39, 1 February 2008. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:MarshBot&oldid=188288598 (visited on 30/01/2019) (cit. on p. 140).

[W306] *User: MartinBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MartinBot (visited on 30/01/2019) (cit. on p. 141).

[W307] *User: MaxSem*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MaxSem (visited on 30/01/2019) (cit. on p. 157).

[W308] *User: MedcabBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MedcabBot (visited on 30/01/2019) (cit. on p. 142).

[W309] *User: Melsaran*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Melsaran (visited on 30/01/2019) (cit. on p. 138).

[W310] *User: MelsaranAWB*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MelsaranAWB (visited on 30/01/2019) (cit. on p. 138).

[W311] *User: MelsaranAWB, user contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/MelsaranAWB (visited on 30/01/2019) (cit. on p. 138).

[W312] *User: MenoBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MenoBot (visited on 30/01/2019) (cit. on p. 139).

[W313] *User: MessageDeliveryBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MessageDeliveryBot (visited on 30/01/2019) (cit. on pp. 136, 140).

[W314] *User: MessedRobot II*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MessedRobot_II (visited on 30/01/2019) (cit. on p. 140).

[W315] *User: MetsBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:MetsBot (visited on 30/01/2019) (cit. on p. 142).

[W316] *User: Mobius Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Mobius_Bot (visited on 30/01/2019) (cit. on p. 138).

[W317] *User: Mobius Bot*. Edit as at 09:30, 19 May 2010. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Mobius_Bot&oldid=362977380 (visited on 30/01/2019) (cit. on p. 138).

[W318] *User: Mobius Bot, diff for destructive edit #396499589 as at 12:49, 13 November 2010*. Edit difference: edit as at 16:38, 11 November 2010 vs. edit as at 12:49, 13 November 2010. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Agriculture&diff=prev&oldid=396499589&diffmode=source (visited on 30/01/2019) (cit. on p. 138).

[W319] *User: Mobius Clock'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Mobius_Clock (visited on 30/01/2019) (cit. on p. 138).

[W320] *User: NicDumZ*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:NicDumZ (visited on 30/01/2019) (cit. on p. 160).

[W321] *User: Nilfanion*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Nilfanion (visited on 30/01/2019) (cit. on p. 163).

[W322] *User: NVS(bot)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:NVS(bot) (visited on 30/01/2019) (cit. on p. 136).

[W323] *User: NVS(bot)*. Edit difference: edit as at 18:45, 20 December 2008 vs. edit as at 04:25, 21 December 2008. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:NVS(bot)&diff=259288250&oldid=259203572 (visited on 30/01/2019) (cit. on p. 136).

[W324] *User: O*. Wikipedia.org. 2019. URL: User:O (visited on 30/01/2019) (cit. on p. 134).

[W325] *User: OldMedCabBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:OldMedCabBot (visited on 30/01/2019) (cit. on p. 142).

[W326] *User: One bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:One%20bot (visited on 30/01/2019) (cit. on p. 136).

[W327] *User: Pigsonthewing*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Pigsonthewing (visited on 30/01/2019) (cit. on p. 177).

[W328] *User: Polbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Polbot (visited on 30/01/2019) (cit. on p. 138).

[W329] *User: Ptbotgourou*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Ptbotgourou (visited on 30/01/2019) (cit. on p. 140).

[W330] *User: R*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:R (visited on 30/01/2019) (cit. on p. 134).

[W331] *User: Radiant!/Classification of admins*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Radiant!/Classification_of_admins (visited on 30/01/2019) (cit. on p. 154).

[W332] *User: Ram-man*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Ram-Man (visited on 30/01/2019) (cit. on pp. 125, 131, 273).

[W333] *User: Rambot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Rambot (visited on 30/01/2019) (cit. on pp. 123, 125, 143–144).

[W334] *User: Rambot*. Edit as at 03:28, 9 December 2002. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Rambot&oldid=471880 (visited on 30/01/2019) (cit. on p. 126).

[W335] *User: Redrocketbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Redrocketbot (visited on 30/01/2019) (cit. on p. 137).

[W336] *User: Redrocketbot, User contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/Redrocketbot (visited on 30/01/2019) (cit. on p. 137).

[W337] *User: RedWolf*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:RedWolf (visited on 30/01/2019) (cit. on pp. 133–134).

[W338] *User: Rich Farmbrough*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Rich_Farmbrough (visited on 30/01/2019) (cit. on pp. 152, 171).

[W339] *User: Rich Farmbrough, Talk archive page*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Rich_Farmbrough/Talk_archive_index (visited on 30/01/2019) (cit. on p. 194).

[W340] *User: Rich Smith*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Rich_Smith (visited on 30/01/2019) (cit. on pp. 141, 150).

[W341] *User: Riley Huntley*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:~riley (visited on 30/01/2019) (cit. on p. 167).

[W342] *User: Riley Huntley*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:RileyBot (visited on 30/01/2019) (cit. on pp. 167–168).

[W343] *User: RileyBot, task log for bot task 10*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:RileyBot/Logs/10 (visited on 30/01/2019) (cit. on p. 168).

[W344] *User: Rubinbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Rubinbot (visited on 30/01/2019) (cit. on p. 140).

[W345] *User: RussBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:RussBot (visited on 30/01/2019) (cit. on p. 144).

[W346] *User: SassoBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SassoBot (visited on 30/01/2019) (cit. on p. 140).

[W347] *User: SatyrBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SatyrBot (visited on 30/01/2019) (cit. on p. 156).

[W348] *User: Seedbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Seedbot (visited on 30/01/2019) (cit. on p. 140).

[W349] *User: Shadowbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Shadowbot (visited on 30/01/2019) (cit. on p. 135).

[W350] *User: SineBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SineBot (visited on 30/01/2019) (cit. on p. 123).

[W351] *User: Skier Dude*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Skier_Dude (visited on 30/01/2019) (cit. on p. 166).

[W352] *User: Skynet*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Skynet&action=edit&redlink=1 (visited on 30/01/2019) (cit. on p. 135).

[W353] *User: Skynet enwiki*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Skynet~enwiki (visited on 30/01/2019) (cit. on p. 135).

[W354] *User: SkynetAI*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SkynetAI (visited on 30/01/2019) (cit. on p. 135).

[W355] *User: SkynetAI SUL Global account information*. Wikipedia.org. 2019. URL: https://meta.wikimedia.org/wiki/Special:CentralAuth?target=SkynetAI (visited on 30/01/2019) (cit. on p. 135).

[W356] *User: SkynetAI, log entries*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Log?type=&user=&page=User%3ASkynet%7Enwiki&wpdate=&tagfilter= (visited on 30/01/2019) (cit. on p. 135).

[W357] *User: SkynetAI, User contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/SkynetAI (visited on 30/01/2019) (cit. on p. 135).

[W358] *User: SmackBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:SmackBot&redirect=no (visited on 30/01/2019) (cit. on pp. 152, 156).

[W359] *User: Soulbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Soulbot (visited on 30/01/2019) (cit. on p. 141).

[W360] *User: SpebiBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SpebiBot (visited on 30/01/2019) (cit. on p. 136).

[W361] *User: SpebiBot*. Edit as at 04:52, 22 July 2007. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:SpebiBot&oldid=146256345 (visited on 30/01/2019) (cit. on p. 136).

[W362] *User: SpellCheckerBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SpellCheckerBot (visited on 30/01/2019) (cit. on p. 123).

[W363] *User: SquelchBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SquelchBot (visited on 30/01/2019) (cit. on p. 136).

[W364] *User: STBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:STBot (visited on 30/01/2019) (cit. on p. 136).

[W365] *User: STBotD*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:STBotD (visited on 30/01/2019) (cit. on p. 136).

[W366] *User: STBotD*. Edit as at 15:52, 16 August 2007. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:STBotD&oldid=151623648 (visited on 30/01/2019) (cit. on p. 136).

[W367] *User: SunCreatorBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:SunCreatorBot (visited on 30/01/2019) (cit. on p. 139).

[W368] *User: T13bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:T13bot (visited on 30/01/2019) (cit. on p. 137).

[W369] *User: T13bot: Revision history*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:T13bot&action=history (visited on 30/01/2019) (cit. on p. 137).

[W370] *User: T13bot: User contributions*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Special:Contributions/T13bot (visited on 30/01/2019) (cit. on p. 137).

[W371] *User: T13bot/tasks*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:T13bot/tasks (visited on 30/01/2019) (cit. on p. 137).

[W372] *User: TARBOT*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:TARBOT (visited on 30/01/2019) (cit. on p. 140).

[W373] *User: Tawkerbot2*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Tawkerbot2 (visited on 30/01/2019) (cit. on pp. 141, 198).

[W374] *User: The Anome*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:The_Anome (visited on 30/01/2019) (cit. on p. 125).

[W375] *User: Theo's Little Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Theo%27s_Little_Bot (visited on 30/01/2019) (cit. on p. 169).

[W376] *User: Theopolisme*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Theopolisme (visited on 30/01/2019) (cit. on p. 168).

[W377] *User: UgenBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:UgenBot (visited on 30/01/2019) (cit. on p. 141).

[W378] *User: VIAFbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:VIAFbot (visited on 30/01/2019) (cit. on p. 139).

[W379] *User: WdefconBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:WdefconBot (visited on 30/01/2019) (cit. on p. 142).

[W380] *User: Werdnabot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Werdnabot (visited on 30/01/2019) (cit. on pp. 140, 161).

[W381] *User: Xenobot Mk V*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Xenobot_Mk_V (visited on 30/01/2019) (cit. on p. 139).

[W382] *User: Xenophon (bot)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Xenophon_(bot) (visited on 30/01/2019) (cit. on p. 141).

[W383] *User: XLinkBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:XLinkBot (visited on 30/01/2019) (cit. on pp. 123, 136, 150).

[W384] *User: Yobot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Yobot (visited on 30/01/2019) (cit. on pp. 151–152, 158).

[W385] *User: Yuzhong*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=User:Yuzhong (visited on 30/01/2019) (cit. on p. 155).

[W386] *User: Zbot370*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Zbot370 (visited on 30/01/2019) (cit. on p. 127).

[W387] *User: ZscoutBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:ZscoutBot (visited on 30/01/2019) (cit. on p. 138).

[W388] *User: Zzyzx11*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/User:Zzyzx11 (visited on 30/01/2019) (cit. on p. 170).

[W389] *Vicipaedia: Pagina prima (Latin Wikipedia)*. Wikipedia.org. 2019. URL: https://la.wikipedia.org/wiki/Vicipaedia:Pagina_prima (visited on 30/01/2019) (cit. on p. 273).

[W390] *Wikidata*. Wikidata.org. 2019. URL: https://www.wikidata.org/wiki/Wikidata:Main_Page (visited on 30/01/2019) (cit. on pp. 48, 83, 179).

[W391] *Wikidata all around the world*. Wikimedia.de. 2019. URL: https://blog.wikimedia.de/2013/04/24/wikidata-all-around-the-world/ (visited on 30/01/2019) (cit. on p. 83).

[W392] *Wikimedia Commons*. Wikipedia.org. 2019. URL: https://commons.wikimedia.org/wiki/Main_Page (visited on 30/01/2019) (cit. on p. 172).

[W393] *Wikimedia Commons: Bots/Requests/AvicBot*. Wikipedia.org. 2019. URL: https://commons.wikimedia.org/wiki/Commons:Bots/Requests/AvicBot (visited on 30/01/2019) (cit. on p. 172).

[W394] *Wikimedia lecture: The Future of Libraries and Wikipedia*. Wikimedia.org. 2019. URL: https://upload.wikimedia.org/wikipedia/commons/0/07/The_Future_of_Libraries_and_Wikipedia.pdf (visited on 30/01/2019) (cit. on p. 9).

[W395] *Wikipapers: List of conference papers*. Wikipapers. 2019. URL: http://wikipapers.referata.com/wiki/List_of_conference_papers (visited on 30/01/2019) (cit. on p. 62).

[W396] *Wikipapers: List of journal articles*. Wikipapers. 2019. URL: http://wikipapers.referata.com/wiki/List_of_journal_articles (visited on 30/01/2019) (cit. on p. 62).

[W397] *Wikipapers: List of Tools*. Wikipapers. 2019. URL: http://wikipapers.referata.com/wiki/List_of_tools (visited on 30/01/2019) (cit. on p. 62).

[W398] *Wikipedia home page, 'Wikipedia, the free encyclopedia' (English language wiki)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Main_Page (visited on 30/01/2019) (cit. on p. 69).

[W399] *Wikipedia public logs for User: Bgbot (to 30 June 2017)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Special:Log&page=User:Bgbot&year=2017&month=6 (visited on 30/01/2019) (cit. on p. 126).

[W400] *Wikipedia talk (Archive page 25): Activity requirements*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_policy/Archive_25#Activity_requirements (visited on 30/01/2019) (cit. on p. 131).

[W401] *Wikipedia talk: AutoWikiBrowser, section 'Convert .28 and .29 to parentheses'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:AutoWikiBrowser#Convert_.28_and_.29_to_parentheses (visited on 30/01/2019) (cit. on p. 157).

[W402] *Wikipedia talk: AutoWikiBrowser/Feature requests, (Archive page 2), section 'Convert .28 and .29 to parentheses'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:AutoWikiBrowser/Feature_requests/Archive_2#Convert_.28_and_.29_to_parentheses (visited on 30/01/2019) (cit. on p. 157).

[W403] *Wikipedia talk: AutoWikiBrowser/Feature requests, section 'Convert .28 and .29 to parentheses'*. Wikipedia.org. 2019. URL: Wikipedia_talk:AutoWikiBrowser/Feature_requests#Convert_.28_and_.29_to_parentheses (visited on 30/01/2019) (cit. on p. 156).

[W404] *Wikipedia talk: Bot Approvals Group*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_Approvals_Group (visited on 30/01/2019) (cit. on p. 131).

[W405] *Wikipedia talk: Bot Approvals Group (Archive page 8) Kumi-Taskbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_Approvals_Group/Archive_8#Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 137).

[W406] *Wikipedia talk: Bot Approvals Group Kumi-Taskbot*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia_talk:Bot_Approvals_Group&oldid=477526099#Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 165).

[W407] *Wikipedia talk: Bot Approvals Group Kumi-Taskbot*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia_talk:Bot_Approvals_Group&oldid=487234536#Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 165).

[W408] *Wikipedia talk: Bot policy*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_policy (visited on 30/01/2019) (cit. on p. 131).

[W409] *Wikipedia talk: Bot requests*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests (visited on 30/01/2019) (cit. on p. 132).

[W410] *Wikipedia talk: Bot requests*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests (visited on 30/01/2019) (cit. on pp. 164, 176).

[W411] *Wikipedia talk: Bot requests (Archive page 1): Archiving 2*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_1#Archiving_2 (visited on 30/01/2019) (cit. on p. 177).

[W412] *Wikipedia talk: Bot requests (Archive page 1): Archiving idea*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_1#Archiving_idea (visited on 30/01/2019) (cit. on p. 177).

[W413] *Wikipedia talk: Bot requests (Archive page 1): Extending Archiving to 7 days?* Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_1#Extending_Archiving_to_7_days? (visited on 30/01/2019) (cit. on p. 176).

[W414] *Wikipedia talk: Bot requests (Archive page 1): Template:BOTREQ*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_1#Template:BOTREQ (visited on 30/01/2019) (cit. on p. 165).

[W415] *Wikipedia talk: Bot requests (Archive page 1): When & how to archive Wikipedia:Bot requests*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_1#When_&_how_to_archive_Wikipedia:Bot_requests (visited on 30/01/2019) (cit. on p. 177).

[W416] *Wikipedia talk: Bot requests (Archive page 2): Erm, what happens here?* Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_2#Erm,_what_happens_here? (visited on 30/01/2019) (cit. on p. 177).

[W417] *Wikipedia talk: Bot requests (Archive page 3): Archived without action*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bot_requests/Archive_2#Archived_without_action (visited on 30/01/2019) (cit. on p. 177).

[W418] *Wikipedia talk: Bots/Noticeboard*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bots/Noticeboard (visited on 30/01/2019) (cit. on p. 131).

[W419] *Wikipedia talk: Bots/Requests for approval*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bots/Requests_for_approval (visited on 30/01/2019) (cit. on p. 132).

[W420] *Wikipedia talk: Subpages section 'Re. Articles do not have sub-pages – wrong'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Subpages#Re._Articles_do_not_have_sub-pages_--_wrong (visited on 07/09/2016) (cit. on p. 110).

[W421] *Wikipedia talk: Summary style (Archive page 1): section 'Summary section vs. subarticle lead'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Summary_style/Archive_1#Summary_section_vs._subarticle_lead (visited on 07/09/2016) (cit. on p. 110).

[W422] *Wikipedia talk: Summary style (Archive page 2)*. See sections 1, 2, 4 and 15. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Summary_style/Archive_2 (visited on 07/09/2016) (cit. on p. 110).

[W423] *Wikipedia talk: Template namespace,(Archive page 29): section 'transcluding prose'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Template_namespace/Archive_1#transcluding_prose (visited on 07/09/2016) (cit. on pp. 112–113).

[W424] *Wikipedia talk: Transclusion section 'Please add clearer real example text'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Transclusion#Please_add_clearer_real_example_text (visited on 07/09/2016) (cit. on p. 112).

[W425] *Wikipedia: A navbox on every page*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:A_navbox_on_every_page (visited on 07/09/2016) (cit. on pp. 101, 107).

[W426] *Wikipedia: Academic studies of Wikipedia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia (visited on 30/01/2019) (cit. on p. 62).

[W427] *Wikipedia: Academic use*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Academic_use (visited on 30/01/2019) (cit. on p. 62).

[W428] *Wikipedia: Administrators, section 'Expectations of adminship'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Administrators#Expectations_of_adminship (visited on 30/01/2019) (cit. on pp. 188, 201).

[W429] *Wikipedia: Administrators' noticeboard*. Edit difference: edit as at 22:34, 30 March 2012 vs. edit as at 23:25, 30 March 2012. Wikipedia.org. 2012. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Administrators%27_noticeboard&diff=484771864&oldid=484771780&diffmode=source (visited on 30/01/2019) (cit. on p. 171).

[W430] *Wikipedia: Administrators/Tools*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Administrators/Tools (visited on 07/09/2016) (cit. on p. 262).

[W431] *Wikipedia: Anarchism task force (ATF)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:ATF (visited on 30/01/2019) (cit. on p. 158).

[W432] *Wikipedia: Arbitration Committee*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Arbitration_Committee (visited on 30/01/2019) (cit. on p. 188).

[W433] *Wikipedia: Arbitration: user Rich Farmbrough*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Arbitration/Requests/Case/Rich_Farmbrough (visited on 30/01/2019) (cit. on p. 168).

[W434] *Wikipedia: Arbitration: user Rich Farmbrough, section 'Rich Farmbrough prohibited from using automation'*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Arbitration/Requests/Case/Rich_Farmbrough#Rich_Farmbrough_prohibited_from_using_automation (visited on 30/01/2019) (cit. on p. 169).

[W435] *Wikipedia: Arbitration: user Rich Farmbrough, section 'Rich Farmbrough: Motions'*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/w/index.php?oldid=729042824#Rich_Farmbrough:_Motions (visited on 30/01/2019) (cit. on p. 169).

[W436] *Wikipedia: Articles for creation*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Articles_for_creation (visited on 30/01/2019) (cit. on p. 283).

[W437] *Wikipedia: Articles for deletion*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion (visited on 30/01/2019) (cit. on p. 283).

[W438] *Wikipedia: Articles for deletion: Anybot's algae articles*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion/Anybot%27s_algae_articles (visited on 30/01/2019) (cit. on p. 137).

[W439] *Wikipedia: AutoWikiBrowser*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:AutoWikiBrowser (visited on 30/01/2019) (cit. on p. 397).

[W440] *Wikipedia: Biographies of living persons*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons (visited on 30/01/2019) (cit. on pp. 159, 283, 291).

[W441] *Wikipedia: Books*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Books (visited on 07/09/2016) (cit. on p. 261).

[W442] *Wikipedia: Bot Approvals Group*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_Approvals_Group (visited on 30/01/2019) (cit. on pp. 128, 131).

[W443] *Wikipedia: Bot Approvals Group - discussion of Kumi-Taskbot*. Edit as at 12:51, 18 February 2012. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia_talk:Bot_Approvals_Group&oldid=477526099#Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 137).

[W444] *Wikipedia: Bot Approvals Group, initial page edit*. Edit as at 02:23, 26 March 2006. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bot_Approvals_Group&oldid=45503763 (visited on 30/01/2019) (cit. on p. 128).

[W445] *Wikipedia: Bot Approvals Group/Guide*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_Approvals_Group/Guide (visited on 30/01/2019) (cit. on p. 131).

[W446] *Wikipedia: Bot Approvals Group/nominations*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_Approvals_Group/nominations (visited on 30/01/2019) (cit. on p. 131).

[W447] *Wikipedia: Bot Approvals Group/nominations/Archive*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_Approvals_Group/nominations/Archive (visited on 30/01/2019) (cit. on p. 131).

[W448] *Wikipedia: Bot policy*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy (visited on 30/01/2019) (cit. on p. 131).

[W449] *Wikipedia: Bot policy, edit as at 15:13, 3 January 2017*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bot_policy&oldid=758109815 (visited on 30/01/2019) (cit. on p. 131).

[W450] *Wikipedia: Bot policy, section 'Cosmetic changes'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy#Cosmetic_changes (visited on 30/01/2019) (cit. on p. 161).

[W451] *Wikipedia: Bot policy, section 'Approval process'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy#Approval_process (visited on 30/01/2019) (cit. on pp. 127, 129).

[W452] *Wikipedia: Bot policy, section 'Interwikilinks'*. Wikimedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy#Interwiki_links (visited on 30/01/2019) (cit. on p. 149).

[W453] *Wikipedia: Bot policy, section 'Valid operations without approval'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy#Valid_operations_without_approval (visited on 30/01/2019) (cit. on p. 126).

[W454] *Wikipedia: Bot policy, section: Activity requirements*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_policy#Activity_requirements (visited on 30/01/2019) (cit. on p. 131).

[W455] *Wikipedia: Bot requests*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests (visited on 30/01/2019) (cit. on pp. 132, 151, 153, 173–174, 178, 180, 194, 289).

[W456] *Wikipedia: Bot requests (Archive page 08): Dead Playboy Playmates*. Wikipedia.org. 2006. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_8#Dead_Playboy_Playmates (visited on 30/01/2019) (cit. on p. 165).

[W457] *Wikipedia: Bot requests (Archive page 1): Italians in the wrong country*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_1#Italians_in_the_wrong_country (visited on 30/01/2019) (cit. on p. 159).

[W458] *Wikipedia: Bot requests (Archive page 11): Bad word and Rss Feed Bot*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_11#Bad_word_and_Rss_Feed_Bot (visited on 30/01/2019) (cit. on p. 161).

[W459] *Wikipedia: Bot requests (Archive page 11): Bot For Wikiproject Aircraft*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_11#Bot_For_Wikiproject_Aircraft (visited on 30/01/2019) (cit. on p. 160).

[W460] *Wikipedia: Bot requests (Archive page 11): Requests for bot approval request tracker*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_11#Requests_for_bot_approval_request_tracker (visited on 30/01/2019) (cit. on p. 169).

[W461] *Wikipedia: Bot requests (Archive page 12): A bot to update Wikicalendar articles*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_12#A_bot_to_update_Wikicalendar_articles (visited on 30/01/2019) (cit. on p. 159).

[W462] *Wikipedia: Bot requests (Archive page 12): Need a bot to set up assessment schemes*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_12#Need_a_bot_to_set_up_assessment_schemes (visited on 30/01/2019) (cit. on p. 166).

[W463] *Wikipedia: Bot requests (Archive page 12): Tchaikovsky navbox*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_12#Tchaikovsky_navbox (visited on 30/01/2019) (cit. on p. 162).

[W464] *Wikipedia: Bot requests (Archive page 13): Bot to identify recreated pages*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_13#Bot_to_identify_recreated_pages (visited on 30/01/2019) (cit. on p. 166).

[W465] *Wikipedia: Bot requests (Archive page 13): Replacement bots*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_13#Replacement_bots (visited on 30/01/2019) (cit. on p. 167).

[W466] *Wikipedia: Bot requests (Archive page 14): 'Request bot to convert %28 and %29 to parentheses'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_14#Request_bot_to_convert_%28_and_%29_to_parentheses (visited on 30/01/2019) (cit. on p. 156).

[W467] *Wikipedia: Bot requests (Archive page 14): Protection check bot*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_14#Protection_check_bot (visited on 30/01/2019) (cit. on p. 166).

[W468] *Wikipedia: Bot requests (Archive page 14): Redirect and template bot*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_14#Redirect_and_template_bot (visited on 30/01/2019) (cit. on p. 161).

[W469] *Wikipedia: Bot requests (Archive page 14): School IP tagging bot*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_14#School_IP_tagging_bot (visited on 30/01/2019) (cit. on p. 165).

[W470] *Wikipedia: Bot requests (Archive page 16): ? a bot to convert {{Somerset}} to {{WikiProject Somerset}}*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#?_a_bot_to_convert_%7B%7BSomerset%7D%7D_to_%7B%7BWikiProject_Somerset%7D%7D (visited on 30/01/2019) (cit. on p. 169).

[W471] *Wikipedia: Bot requests (Archive page 16): EditorBot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#EditorBot (visited on 30/01/2019) (cit. on p. 160).

[W472] *Wikipedia: Bot requests (Archive page 16): Fair-use in non-mainspace*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#Fair-use_in_non-mainspace (visited on 30/01/2019) (cit. on p. 159).

[W473] *Wikipedia: Bot requests (Archive page 16): For someone with way too much time on his hands*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#For_someone_with_way_too_much_time_on_his_hands (visited on 30/01/2019) (cit. on p. 154).

[W474] *Wikipedia: Bot requests (Archive page 16): New Bot Idea - Links*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#New_Bot_Idea_-_Links (visited on 30/01/2019) (cit. on p. 160).

[W475] *Wikipedia: Bot requests (Archive page 16): One Day International cricket*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#One_Day_International_cricket (visited on 30/01/2019) (cit. on p. 156).

[W476] *Wikipedia: Bot requests (Archive page 16): Polynesia tagging*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#Polynesia_tagging (visited on 30/01/2019) (cit. on p. 161).

[W477] *Wikipedia: Bot requests (Archive page 16): Wikify bot ?* Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_16#Wikify_bot_%3F (visited on 30/01/2019) (cit. on p. 160).

[W478] *Wikipedia: Bot requests (Archive page 17): EditorBot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_17#EditorBot (visited on 30/01/2019) (cit. on p. 159).

[W479] *Wikipedia: Bot requests (Archive page 18): Bot to change old signature/user name*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_18#Bot_to_change_old_signature%2Fuser_name (visited on 30/01/2019) (cit. on p. 162).

[W480] *Wikipedia: Bot requests (Archive page 18): Missing settlements*. Wikipedia.org. 2008. URL: Wikipedia:%20https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_18#Missing_settlements (visited on 30/01/2019) (cit. on p. 162).

[W481] *Wikipedia: Bot requests (Archive page 18): Tag some articles*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_18#Tag_some_articles (visited on 30/01/2019) (cit. on p. 156).

[W482] *Wikipedia: Bot requests (Archive page 19): Tag Genetics stubs with poject template*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_19#Tag_Genetics_stubs_with_poject_template (visited on 30/01/2019) (cit. on p. 164).

[W483] *Wikipedia: Bot requests (Archive page 19): Template documentation*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_19#Template_documentation (visited on 30/01/2019) (cit. on p. 157).

[W484] *Wikipedia: Bot requests (Archive page 20): DashBot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_20#DashBot (visited on 30/01/2019) (cit. on p. 161).

[W485] *Wikipedia: Bot requests (Archive page 21): Add navbox to articles in Category:Infoshops*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_21#Add_navbox_to_articles_in_Category:Infoshops (visited on 30/01/2019) (cit. on p. 166).

[W486] *Wikipedia: Bot requests (Archive page 21): Automated anniversaries for Portals*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_21#Automated_anniversaries_for_Portals (visited on 30/01/2019) (cit. on p. 158).

[W487] *Wikipedia: Bot requests (Archive page 21): List of doping cases in sport*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_21#List_of_doping_cases_in_sport (visited on 30/01/2019) (cit. on p. 170).

[W488] *Wikipedia: Bot requests (Archive page 22)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_22 (visited on 30/01/2019) (cit. on p. 132).

[W489] *Wikipedia: Bot requests (Archive page 23): Historical "What links here" function*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_23#Historical_%22What_links_here%22_function (visited on 30/01/2019) (cit. on p. 155).

[W490] *Wikipedia: Bot requests (Archive page 24): Bulk-replace URL for Handbook of Texas Online*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_24#Bulk-replace_URL_for_Handbook_of_Texas_Online (visited on 30/01/2019) (cit. on p. 156).

[W491] *Wikipedia: Bot requests (Archive page 24): Category:Blackpool F.C. seasons*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_24#Category:Blackpool_F.C._seasons (visited on 30/01/2019) (cit. on p. 157).

[W492] *Wikipedia: Bot requests (Archive page 24): JerryBot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_24#JerryBot (visited on 30/01/2019) (cit. on p. 172).

[W493] *Wikipedia: Bot requests (Archive page 24): Most Redlinked Bot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_24#Most_Redlinked_Bot (visited on 30/01/2019) (cit. on p. 170).

[W494] *Wikipedia: Bot requests (Archive page 24): Placebox template series*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_24#Placebox_template_series (visited on 30/01/2019) (cit. on p. 158).

[W495] *Wikipedia: Bot requests (Archive page 25): Replacing calender templates*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_25#Replacing_calender_templates (visited on 30/01/2019) (cit. on p. 170).

[W496] *Wikipedia: Bot requests (Archive page 26): euphemism bot*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_26#euphemism_bot (visited on 30/01/2019) (cit. on p. 161).

[W497] *Wikipedia: Bot requests (Archive page 26): Template:Canada CP 2006*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_26#Template:Canada_CP_2006 (visited on 30/01/2019) (cit. on p. 170).

[W498] *Wikipedia: Bot requests (Archive page 27): Announcing a bot run*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_27#Announcing_a_bot_run (visited on 30/01/2019) (cit. on p. 164).

[W499] *Wikipedia: Bot requests (Archive page 27): Template change in Copy to Wikimedia Commons category*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_27#Template_change_in_Copy_to_Wikimedia_Commons_category (visited on 30/01/2019) (cit. on p. 156).

[W500] *Wikipedia: Bot requests (Archive page 27): Vague idea for a bot/technical request*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_27#Vague_idea_for_a_bot%2Ftechnical_request (visited on 30/01/2019) (cit. on p. 160).

[W501] *Wikipedia: Bot requests (Archive page 29): American film taskforce tagging*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_29#American_film_taskforce_tagging (visited on 30/01/2019) (cit. on p. 165).

[W502] *Wikipedia: Bot requests (Archive page 3): Recategorization bot*. Wikipedia.org. 2005. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_3#Recategorization_bot (visited on 30/01/2019) (cit. on p. 158).

[W503] *Wikipedia: Bot requests (Archive page 32): Paraguay Templates*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_32#Paraguay_Templates (visited on 30/01/2019) (cit. on p. 164).

[W504] *Wikipedia: Bot requests (Archive page 32): Removing user namespace pages from article namespace categories*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_32#Removing_user_namespace_pages_from_article_namespace_categories (visited on 30/01/2019) (cit. on p. 157).

[W505] *Wikipedia: Bot requests (Archive page 33): Change of template name*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_33#Change_of_template_name (visited on 30/01/2019) (cit. on p. 163).

[W506] *Wikipedia: Bot requests (Archive page 33): Need a Bot for a Job*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_33#Need_a_Bot_for_a_Job (visited on 30/01/2019) (cit. on p. 161).

[W507] *Wikipedia: Bot requests (Archive page 33): Sockpuppet investigations bot*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_33#Sockpuppet_investigations_bot (visited on 30/01/2019) (cit. on p. 167).

[W508] *Wikipedia: Bot requests (Archive page 33): Tagging request*. Wikipedia.org. 2009. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_33#Tagging_request (visited on 30/01/2019) (cit. on p. 156).

[W509] *Wikipedia: Bot requests (Archive page 33): Unsourced BLPs*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_33#Unsourced_BLPs (visited on 30/01/2019) (cit. on p. 159).

[W510] *Wikipedia: Bot requests (Archive page 34): Unreferenced biography of living persons bot to get projects involved in referencing*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_34#Unreferenced_biography_of_living_persons_bot_to_get_projects_involved_in_referencing. (visited on 30/01/2019) (cit. on p. 168).

[W511] *Wikipedia: Bot requests (Archive page 35): Changing of {{otheruses4}} to {{about}}*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_35#Changing_of_%7B%7Botheruses4%7D%7D_to_%7B%7Babout%7D%7D (visited on 30/01/2019) (cit. on pp. 162–163).

[W512] *Wikipedia: Bot requests (Archive page 35): Convert dates to template*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_35#Convert_dates_to_template (visited on 30/01/2019) (cit. on p. 162).

[W513] *Wikipedia: Bot requests (Archive page 35): Converting refs to list-defined refs*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_35#Converting_refs_to_list-defined_refs (visited on 30/01/2019) (cit. on p. 158).

[W514] *Wikipedia: Bot requests (Archive page 35): Username Change*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_35#Username_Change (visited on 30/01/2019) (cit. on p. 156).

[W515] *Wikipedia: Bot requests (Archive page 36): Book for Nobel laureates*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_36#Book_for_Nobel_laureates (visited on 30/01/2019) (cit. on p. 162).

[W516] *Wikipedia: Bot requests (Archive page 36): Holding cell*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_36#Holding_cell (visited on 30/01/2019) (cit. on p. 164).

[W517] *Wikipedia: Bot requests (Archive page 36): International law wikiproject template*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_36#International_law_wikiproject_template (visited on 30/01/2019) (cit. on p. 163).

[W518] *Wikipedia: Bot requests (Archive page 36): User:ArticleAlertbot*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_36#User:ArticleAlertbot (visited on 30/01/2019) (cit. on p. 170).

[W519] *Wikipedia: Bot requests (Archive page 36): User:ArticleAlertbot has been inoperable for a month*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_36#User:ArticleAlertbot_has_been_inoperable_for_a_month (visited on 30/01/2019) (cit. on p. 170).

[W520] *Wikipedia: Bot requests (Archive page 37): Bypassing redirects with accent characters*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_37#Bypassing_redirects_with_accent_characters (visited on 30/01/2019) (cit. on p. 164).

[W521] *Wikipedia: Bot requests (Archive page 37): We need another User:WebCiteBOT*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_37#We_need_another_User:WebCiteBOT (visited on 30/01/2019) (cit. on p. 172).

[W522] *Wikipedia: Bot requests (Archive page 39): Undated Articles*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_39#Undated_Articles (visited on 30/01/2019) (cit. on p. 171).

[W523] *Wikipedia: Bot requests (Archive page 40): Need new UAA bot*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_40#Need_new_UAA_bot (visited on 30/01/2019) (cit. on p. 168).

[W524] *Wikipedia: Bot requests (Archive page 41): Category:Infobox person using deprecated parameters*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_41#Category:Infobox_person_using_deprecated_parameters (visited on 30/01/2019) (cit. on p. 163).

[W525] *Wikipedia: Bot requests (Archive page 42): Bot to convert from citation templates to non-citation templates*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_42#Bot_to_convert_from_citation_templates_to_non-citation_templates (visited on 30/01/2019) (cit. on p. 167).

[W526] *Wikipedia: Bot requests (Archive page 42): Moving reviews out of Infobox Albums*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_42#Moving_reviews_out_of_Infobox_Albums (visited on 30/01/2019) (cit. on p. 155).

[W527] *Wikipedia: Bot requests (Archive page 42): Template:NFL player Template:Infobox gridiron football person*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_42#Template:NFL_player_%E2%86%92_Template:Infobox_gridiron_football_person (visited on 30/01/2019) (cit. on p. 163).

[W528] *Wikipedia: Bot requests (Archive page 43): Bot Creation*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_43#Bot_Creation (visited on 30/01/2019) (cit. on p. 165).

[W529] *Wikipedia: Bot requests (Archive page 43): Bot to keep track of bots*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_43#Bot_to_keep_track_of_bots (visited on 30/01/2019) (cit. on p. 169).

[W530] *Wikipedia: Bot requests (Archive page 43): Links to Commons*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_43#Links_to_Commons (visited on 30/01/2019) (cit. on p. 163).

[W531] *Wikipedia: Bot requests (Archive page 43): Removing templates from disambiguation pages*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_43#Removing_templates_from_disambiguation_pages (visited on 30/01/2019) (cit. on p. 162).

[W532] *Wikipedia: Bot requests (Archive page 44): Redirects for Olympic articles*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_44#Redirects_for_Olympic_articles (visited on 30/01/2019) (cit. on p. 159).

[W533] *Wikipedia: Bot requests (Archive page 45): Quick text replacment*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_45#Quick_text_replacment (visited on 30/01/2019) (cit. on p. 155).

[W534] *Wikipedia: Bot requests (Archive page 46): Hatnote renaming*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_46#Hatnote_renaming (visited on 30/01/2019) (cit. on p. 158).

[W535] *Wikipedia: Bot requests (Archive page 46): Quis custodiet ipsos custodes?* Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_46#Quis_custodiet_ipsos_custodes%3F (visited on 30/01/2019) (cit. on p. 172).

[W536] *Wikipedia: Bot requests (Archive page 47): add ref section to language stubs*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_47#add_ref_section_to_language_stubs (visited on 30/01/2019) (cit. on p. 171).

[W537] *Wikipedia: Bot requests (Archive page 47): Asteroid Thingy*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_47#Asteroid_Thingy (visited on 30/01/2019) (cit. on p. 169).

[W538] *Wikipedia: Bot requests (Archive page 47): Project tagging bot*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_47#Project_tagging_bot (visited on 30/01/2019) (cit. on p. 165).

[W539] *Wikipedia: Bot requests (Archive page 47): Search for references/citations*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_47#Search_for_references%2Fcitations (visited on 30/01/2019) (cit. on p. 164).

[W540] *Wikipedia: Bot requests (Archive page 47): wikiproject articles by size*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_47#wikiproject_articles_by_size (visited on 30/01/2019) (cit. on p. 172).

[W541] *Wikipedia: Bot requests (Archive page 48): Bot to cleanup Template:Duplicate file info*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_48#Bot_to_cleanup_Template:Duplicate_file_info (visited on 30/01/2019) (cit. on p. 166).

[W542] *Wikipedia: Bot requests (Archive page 48): People to pick up after RF bot's task*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_48#People_to_pick_up_after_RF_bot%27s_task (visited on 30/01/2019) (cit. on p. 169).

[W543] *Wikipedia: Bot requests (Archive page 48): Removal of deprecated template*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_48#Removal_of_deprecated_template (visited on 30/01/2019) (cit. on p. 160).

[W544] *Wikipedia: Bot requests (Archive page 49): Looking for a Commons bot operator to do a simple category merger*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_49#Looking_for_a_Commons_bot_operator_to_do_a_simple_category_merger (visited on 30/01/2019) (cit. on p. 172).

[W545] *Wikipedia: Bot requests (Archive page 52): template bombing*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_52#template_bombing (visited on 30/01/2019) (cit. on p. 157).

[W546] *Wikipedia: Bot requests (Archive page 54): Archive version bot*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Archive_version_bot (visited on 30/01/2019) (cit. on p. 168).

[W547] *Wikipedia: Bot requests (Archive page 54): Creating a set of new articles*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Creating_a_set_of_new_articles (visited on 30/01/2019) (cit. on p. 162).

[W548] *Wikipedia: Bot requests (Archive page 54): Featured Portal closings*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Featured_Portal_closings (visited on 30/01/2019) (cit. on p. 168).

[W549] *Wikipedia: Bot requests (Archive page 54): Mark a lot of pages for microformatting, round 2*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Mark_a_lot_of_pages_for_microformatting%2C_round_2 (visited on 30/01/2019) (cit. on p. 166).

[W550] *Wikipedia: Bot requests (Archive page 54): Non-notable diplomat stub articles*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Non-notable_diplomat_stub_articles (visited on 30/01/2019) (cit. on p. 167).

[W551] *Wikipedia: Bot requests (Archive page 54): Redirects to sections*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Redirects_to_sections (visited on 30/01/2019) (cit. on p. 169).

[W552] *Wikipedia: Bot requests (Archive page 54): Schuylkill*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_54#Schuylkill (visited on 30/01/2019) (cit. on p. 164).

[W553] *Wikipedia: Bot requests (Archive page 56): Bot to tag articles only sourced to National Register Information System*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_56#Bot_to_tag_articles_only_sourced_to_National_Register_Information_System (visited on 30/01/2019) (cit. on p. 163).

[W554] *Wikipedia: Bot requests (Archive page 56): Category:Pages using citations with accessdate and no URL*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_56#Category:Pages_using_citations_with_accessdate_and_no_URL (visited on 30/01/2019) (cit. on p. 166).

[W555] *Wikipedia: Bot requests (Archive page 58): Template:Infobox dam*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_58#Template:Infobox_dam (visited on 30/01/2019) (cit. on p. 169).

[W556] *Wikipedia: Bot requests (Archive page 58): YoshiBot*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_58#YoshiBot (visited on 30/01/2019) (cit. on p. 160).

[W557] *Wikipedia: Bot requests (Archive page 59): Meta: unresolved requests*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_59#Meta:_unresolved_requests (visited on 30/01/2019) (cit. on p. 160).

[W558] *Wikipedia: Bot requests (Archive page 59): MTA (New York City) dead links*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_59#MTA_%28New_York_City%29_dead_links (visited on 30/01/2019) (cit. on p. 171).

[W559] *Wikipedia: Bot requests (Archive page 6): Userbox and Userboxes bot*. Wikipedia.org. 2006. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_6#Userbox_and_Userboxes_bot (visited on 30/01/2019) (cit. on p. 158).

[W560] *Wikipedia: Bot requests (Archive page 60): Remove unnecessary 'Image requested' templates*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_60#Remove_unnecessary_%27Image_requested%27_templates (visited on 30/01/2019) (cit. on p. 162).

[W561] *Wikipedia: Bot requests (Archive page 60): WikiProject talk page tagging*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_60#WikiProject_talk_page_tagging (visited on 30/01/2019) (cit. on p. 162).

[W562] *Wikipedia: Bot requests (Archive page 61): Tagging articles for a WikiProject*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_61#Tagging_articles_for_a_WikiProject (visited on 30/01/2019) (cit. on p. 165).

[W563] *Wikipedia: Bot requests (Archive page 61): Trouble communicating with bot operator, request for new bot maintainer*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_61#Trouble_communicating_with_bot_operator%2C_request_for_new_bot_maintainer (visited on 30/01/2019) (cit. on p. 169).

[W564] *Wikipedia: Bot requests (Archive page 62): Bot to create links for unlinked pages called from template subpages*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_62#Bot_to_create_links_for_unlinked_pages_called_from_template_subpages (visited on 30/01/2019) (cit. on p. 157).

[W565] *Wikipedia: Bot requests (Archive page 62): Changing a foreign word to its translation*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_62#Changing_a_foreign_word_to_its_translation (visited on 30/01/2019) (cit. on p. 158).

[W566] *Wikipedia: Bot requests (Archive page 63): Adding information templates to files that don't have one*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_63#Adding_information_templates_to_files_that_don%27t_have_one (visited on 30/01/2019) (cit. on p. 165).

[W567] *Wikipedia: Bot requests (Archive page 63): Automated identification of excessive WP:Tag-bombing*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_63#Automated_identification_of_excessive_WP:Tag-bombing (visited on 30/01/2019) (cit. on p. 157).

[W568] *Wikipedia: Bot requests (Archive page 64): SpongeBob SquarePants task*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_64#SpongeBob_SquarePants_task (visited on 30/01/2019) (cit. on p. 165).

[W569] *Wikipedia: Bot requests (Archive page 65): Convert Rs. to ₹*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_65#Convert_Rs._to_%E2%82%B9 (visited on 30/01/2019) (cit. on p. 161).

[W570] *Wikipedia: Bot requests (Archive page 65): Infobox*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_65#Infobox (visited on 30/01/2019) (cit. on p. 161).

[W571] *Wikipedia: Bot requests (Archive page 65): Jain Agamas*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_65#Jain_Agamas (visited on 30/01/2019) (cit. on p. 161).

[W572] *Wikipedia: Bot requests (Archive page 69): Bot for dashes*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_69#Bot_for_dashes (visited on 30/01/2019) (cit. on p. 158).

[W573] *Wikipedia: Bot requests (Archive page 69): Redirects to lists, from the things they are lists of*. Wikipedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_69#Redirects_to_lists%2C_from_the_things_they_are_lists_of (visited on 30/01/2019) (cit. on p. 171).

[W574] *Wikipedia: Bot requests (Archive page 73): BSicons*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_73#BSicons (visited on 30/01/2019) (cit. on p. 164).

[W575] *Wikipedia: Bot requests (Archive page 73): Create a simple TemplateData for all Infoboxes*. Wikipedia.org. 2017. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_73#Create_a_simple_TemplateData_for_all_Infoboxes (visited on 30/01/2019) (cit. on p. 165).

[W576] *Wikipedia: Bot requests (Archive page 74): WP 1.0 bot*. Wikipedia.org. 2017. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Archive_74#WP_1.0_bot (visited on 30/01/2019) (cit. on p. 169).

[W577] *Wikipedia: Bot requests/Frequently denied bots*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bot_requests/Frequently_denied_bots (visited on 30/01/2019) (cit. on pp. 132, 175).

[W578] *Wikipedia: Bot/Noticeboard: - Rich Farmbrough's bots*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard/Archive_7#Rich_Farmbrough's_bots (visited on 30/01/2019) (cit. on p. 169).

[W579] *Wikipedia: BOTACC*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:BOTACC (visited on 30/01/2019) (cit. on p. 133).

[W580] *Wikipedia: Bots*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots (visited on 07/09/2016) (cit. on pp. 122, 130–131, 263).

[W581] *Wikipedia: Bots talk*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bots (visited on 30/01/2019) (cit. on p. 131).

[W582] *Wikipedia: Bots, Requests for approval (original edit)*. Edit as at 18:27, 13 March 2006. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bots/Requests_for_approval&oldid=43605786 (visited on 30/01/2019) (cit. on p. 127).

[W583] *Wikipedia: Bots: Requests for approval (Archive page 4): Request for deflagging and blocking of Anybot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bots/Requests_for_approval/Archive_4#Request_for_deflagging_and_blocking_of_Anybot (visited on 30/01/2019) (cit. on p. 137).

[W584] *Wikipedia: Bots: Requests for approval (Archive page 7): Request for deflagging and blocking of Anybot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia_talk:Bots/Requests_for_approval/Archive_7#Request_for_deflagging_and_blocking_of_Anybot (visited on 30/01/2019) (cit. on p. 137).

[W585] *Wikipedia: Bots/ Noticeboard (Archive 08) Bot takeover*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard/Archive_8#Bot_takeover (visited on 30/01/2019) (cit. on p. 168).

[W586] *Wikipedia: Bots/Dictionary*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Dictionary (visited on 30/01/2019) (cit. on p. 131).

[W587] *Wikipedia: Bots/News*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/News (visited on 30/01/2019) (cit. on p. 132).

[W588] *Wikipedia: Bots/Noticeboard*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard (visited on 30/01/2019) (cit. on p. 131).

[W589] *Wikipedia: Bots/Noticeboard (Archive page 8): Wikidata*. Wikimedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard/Archive_8#Wikidata (visited on 30/01/2019) (cit. on p. 149).

[W590] *Wikipedia: Bots/Noticeboard, section 'Kumi-Taskbot and collapse parameters'*. Wikipedia.org. 2011. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard/Archive_7#Kumi-Taskbot_and_collapse_parameters (visited on 30/01/2019) (cit. on p. 165).

[W591] *Wikipedia: Bots/Noticeboard/InactiveBots/December2016*. Wikimedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Noticeboard/InactiveBots/December2016 (visited on 30/01/2019) (cit. on p. 149).

[W592] *Wikipedia: Bots/Requests for approval*. Edit difference: edit as at 20:34, 13 March 2006 vs. edit as at 21:35, 13 March 2006. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bots/Requests_for_approval&diff=43624849&oldid=43618299 (visited on 30/01/2019) (cit. on p. 127).

[W593] *Wikipedia: Bots/Requests for approval*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?action=edit&preload=Wikipedia%3ABots%2FRequests+for+approval%2FInputInit&editintro=Wikipedia%3ABots%2FRequests+for+approval%2FEditintro&title=Wikipedia%3ABots%2FRequests+for+approval%2FTestBot&create=Create+request+page (visited on 30/01/2019) (cit. on p. 127).

[W594] *Wikipedia: Bots/Requests for approval*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval (visited on 30/01/2019) (cit. on p. 132).

[W595] *Wikipedia: Bots/Requests for approval (edit at 00:03, 23 March 2006)*. Edit difference: edit as at 23:47, 22 March 2006 vs. edit as at 00:03, 23 March 2006. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bots/Requests_for_approval&diff=45021032&oldid=45019161&diffmode=source (visited on 30/01/2019) (cit. on p. 127).

[W596] *Wikipedia: Bots/Requests for approval (edit at 14:29, 6 September 2006)*. Edit difference: edit as at 02:02, 6 September 2006 vs. edit as at 14:29, 6 September 2006. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Bots/Requests_for_approval&diff=prev&oldid=74136983&diffmode=source (visited on 30/01/2019) (cit. on p. 127).

[W597] *Wikipedia: Bots/Requests for approval/AAlertBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/AAlertBot (visited on 30/01/2019) (cit. on p. 128).

[W598] *Wikipedia: Bots/Requests for approval/Andrea105Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Andrea105Bot (visited on 30/01/2019) (cit. on p. 128).

[W599] *Wikipedia: Bots/Requests for approval/Anybot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/anybot (visited on 30/01/2019) (cit. on p. 137).

[W600] *Wikipedia: Bots/Requests for approval/BHGbot, task #3*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/BHGbot_3 (visited on 30/01/2019) (cit. on p. 129).

[W601] *Wikipedia: Bots/Requests for approval/DarafshBot, task #3*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DarafshBot_3 (visited on 30/01/2019) (cit. on p. 141).

[W602] *Wikipedia: Bots/Requests for approval/DeadLinkBOT*. Wikipedia.org. 2008. URL: Wikipedia:%20https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DeadLinkBOT (visited on 30/01/2019) (cit. on p. 156).

[W603] *Wikipedia: Bots/Requests for approval/DeltaQuadBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DeltaQuadBot (visited on 30/01/2019) (cit. on p. 128).

[W604] *Wikipedia: Bots/Requests for approval/DyceBot, task #4*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/DyceBot_4 (visited on 30/01/2019) (cit. on p. 162).

[W605] *Wikipedia: Bots/Requests for approval/HasteurBot, task #4*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/HasteurBot_4 (visited on 30/01/2019) (cit. on p. 163).

[W606] *Wikipedia: Bots/Requests for approval/Hazard-Bot, task #11*. Wikipedia.org. 2012. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Hazard-Bot_11 (visited on 30/01/2019) (cit. on p. 169).

[W607] *Wikipedia: Bots/Requests for approval/Hazard-Bot, task #21*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Hazard-Bot_21 (visited on 30/01/2019) (cit. on p. 166).

[W608] *Wikipedia: Bots/Requests for approval/Helpful Pixie Bot, task #46*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Helpful_Pixie_Bot_46 (visited on 30/01/2019) (cit. on p. 128).

[W609] *Wikipedia: Bots/Requests for approval/Helpful Pixie Bot, task #48*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Helpful_Pixie_Bot_48 (visited on 30/01/2019) (cit. on p. 171).

[W610] *Wikipedia: Bots/Requests for approval/Helpful Pixie Bot, task #50*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Helpful_Pixie_Bot_50 (visited on 30/01/2019) (cit. on p. 169).

[W611] *Wikipedia: Bots/Requests for approval/HersfoldArbClerkBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/HersfoldArbClerkBot (visited on 30/01/2019) (cit. on p. 139).

[W612] *Wikipedia: Bots/Requests for approval/JerryBot*. Wikipedia.org. 2008. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/JerryBot (visited on 30/01/2019) (cit. on p. 172).

[W613] *Wikipedia: Bots/Requests for approval/John Bot, task #3*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/John_Bot_3 (visited on 30/01/2019) (cit. on p. 138).

[W614] *Wikipedia: Bots/Requests for approval/KarlsenBot, task #6*. Wikipedia.org. 2010. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/KarlsenBot_6 (visited on 30/01/2019) (cit. on p. 171).

[W615] *Wikipedia: Bots/Requests for approval/Kumi-Taskbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Kumi-Taskbot (visited on 30/01/2019) (cit. on p. 137).

[W616] *Wikipedia: Bots/Requests for approval/Mobius Bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Mobius_Bot (visited on 30/01/2019) (cit. on p. 138).

[W617] *Wikipedia: Bots/Requests for approval/MoohanBOT, task #8*. Wikimedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/MoohanBOT_8 (visited on 30/01/2019) (cit. on p. 171).

[W618] *Wikipedia: Bots/Requests for approval/Redirectbot*. Wikimedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Redirectbot (visited on 30/01/2019) (cit. on p. 171).

[W619] *Wikipedia: Bots/Requests for approval/Redirectbot, task #2*. Wikimedia.org. 2015. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Redirectbot_2 (visited on 30/01/2019) (cit. on p. 171).

[W620] *Wikipedia: Bots/Requests for approval/Redrocketbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Redrocketbot (visited on 30/01/2019) (cit. on p. 137).

[W621] *Wikipedia: Bots/Requests for approval/RileyBot, task #15*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/RileyBot_15 (visited on 30/01/2019) (cit. on p. 169).

[W622] *Wikipedia: Bots/Requests for approval/RileyBot, task #9*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/RileyBot_9 (visited on 30/01/2019) (cit. on p. 166).

[W623] *Wikipedia: Bots/Requests for approval/RileyBot, task 10*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/RileyBot_10 (visited on 30/01/2019) (cit. on p. 167).

[W624] *Wikipedia: Bots/Requests for approval/Seedbot, task #2*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Seedbot2 (visited on 30/01/2019) (cit. on p. 159).

[W625] *Wikipedia: Bots/Requests for approval/Skynet*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Skynet (visited on 30/01/2019) (cit. on p. 135).

[W626] *Wikipedia: Bots/Requests for approval/SoxBot, task #17*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/SoxBot_17 (visited on 30/01/2019) (cit. on p. 156).

[W627] *Wikipedia: Bots/Requests for approval/SquelchBot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/SquelchBot (visited on 30/01/2019) (cit. on p. 136).

[W628] *Wikipedia: Bots/Requests for approval/T13bot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/T13bot (visited on 30/01/2019) (cit. on p. 137).

[W629] *Wikipedia: Bots/Requests for approval/Theo's Little Bot, task #8*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Theo%27s_Little_Bot_8 (visited on 30/01/2019) (cit. on p. 162).

[W630] *Wikipedia: Bots/Requests for approval/Zscoutbot*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/ZscoutBot (visited on 30/01/2019) (cit. on p. 138).

[W631] *Wikipedia: Categorization*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Categorization (visited on 30/01/2019) (cit. on pp. 56, 157, 200).

[W632] *Wikipedia: Conflict of interest*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest (visited on 30/01/2019) (cit. on pp. 283, 292).

[W633] *Wikipedia: Consensus*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Consensus (visited on 30/01/2019) (cit. on p. 191).

[W634] *Wikipedia: Content assessment*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Content_assessment (visited on 30/01/2019) (cit. on pp. 283, 291).

[W635] *Wikipedia: Courtesy vanishing (Right to Vanish)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Courtesy_vanishing (visited on 30/01/2019) (cit. on pp. 136, 156).

[W636] *Wikipedia: Criteria for speedy deletion, section 'U1' - User Request*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Criteria_for_speedy_deletion#U1 (visited on 30/01/2019) (cit. on p. 137).

[W637] *Wikipedia: Database reports/WikiProject watchers*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Database_reports/WikiProject_watchers (visited on 07/09/2016) (cit. on p. 265).

[W638] *Wikipedia: Database reports/WikiProjects by changes*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Database_reports/WikiProjects_by_changes (visited on 07/09/2016) (cit. on p. 265).

[W639] *Wikipedia: Dispute resolution noticeboard (Archive page 29): Friends episodes*. Thread in Archive page #29. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Dispute_resolution_noticeboard/Archive_29#Friends_episodes (visited on 07/09/2016) (cit. on p. 111).

[W640] *Wikipedia: Do not use subpages*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Do_not_use_subpages&oldid=2836167 (visited on 07/09/2016) (cit. on p. 112).

[W641] *Wikipedia: Edit count*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Edit_count (visited on 30/01/2019) (cit. on pp. 54, 187, 191).

[W642] *Wikipedia: Edit warring*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Edit_warring (visited on 30/01/2019) (cit. on pp. 143, 263).

[W643] *Wikipedia: Edit warring - policy*. Edit as at 20:59, 13 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Edit_warring&oldid=699684847 (visited on 07/09/2016) (cit. on p. 111).

[W644] *Wikipedia: Education program*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Education_program (visited on 07/09/2016) (cit. on p. 261).

[W645] *Wikipedia: Expert editors*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Expert_editors (visited on 30/01/2019) (cit. on p. 187).

[W646] *Wikipedia: Featured articles*. Edit as at 00:05, 11 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_articles&oldid=699215397 (visited on 07/09/2016) (cit. on p. 111).

[W647] *Wikipedia: Featured articles*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_articles (visited on 30/01/2019) (cit. on p. 190).

[W648] *Wikipedia: Featured list candidates*. Edit as at 10:03, 22 May 2012. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Featured_list_candidates&oldid=493802226 (visited on 07/09/2016) (cit. on p. 111).

[W649] *Wikipedia: Featured list candidates*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Featured_list_candidates (visited on 10/01/2016) (cit. on p. 111).

[W650] *Wikipedia: Featured lists*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Featured_lists (visited on 10/01/2016) (cit. on p. 111).

[W651] *Wikipedia: Global rights policy*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Global_rights_policy (visited on 07/09/2016) (cit. on p. 263).

[W652] *Wikipedia: Global rights policy, section 'Stewards'*. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Global_rights_policy#Stewards (visited on 30/01/2019) (cit. on p. 126).

[W653] *Wikipedia: Good articles*. Edit as at 12:40, 13 May 2014. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Good_articles&oldid=608372225 (visited on 07/09/2016) (cit. on pp. 110–111).

[W654] *Wikipedia: Good articles*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Good_articles (visited on 30/01/2019) (cit. on p. 190).

[W655] *Wikipedia: History of Wikipedia bots*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:History_of_Wikipedia_bots (visited on 30/01/2019) (cit. on pp. 125, 132).

[W656] *Wikipedia: Ignore all rules (WP:IAR)*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Ignore_all_rules (visited on 30/01/2019) (cit. on pp. 142, 192, 400).

[W657] *Wikipedia: Lamest edit wars#(Ro)bot wars*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Lamest_edit_wars#(Ro)bot_wars (visited on 30/01/2019) (cit. on p. 143).

[W658] *Wikipedia: List of Wikipedians by number of edits, section 'List of Wikipedians by number of edits'*. Archived by the Internet Archive as at 19 July 2016, 16:24:47. Wikipedia.org. 2019. URL: https://web.archive.org/web/20170719162447/https://en.wikipedia.org/wiki/Wikipedia:List_of_Wikipedians_by_number_of_edits#List_of_Wikipedians_by_number_of_edits (visited on 30/01/2019) (cit. on p. 152).

[W659] *Wikipedia: Manual of Style*. Edit as at 18:32, 6 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style&oldid=698531326 (visited on 07/09/2016) (cit. on pp. 111, 161–162).

[W660] *Wikipedia: Manual of Style, Lists*. Edit as at 14:49, 10 November 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Manual_of_Style/Lists&oldid=689970575 (visited on 07/09/2016) (cit. on p. 111).

[W661] *Wikipedia: Naming conventions (television)*. Edit as at 23:57, 7 September 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Naming_conventions_(television)&oldid=679976299 (visited on 07/09/2016) (cit. on p. 104).

[W662] *Wikipedia: Notability*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Notability (visited on 30/01/2019) (cit. on p. 168).

[W663] *Wikipedia: Ownership of content*. Edit as at 04:31, 7 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Ownership_of_content&oldid=698607754 (visited on 07/09/2016) (cit. on p. 111).

[W664] *Wikipedia: Ownership of content*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Ownership_of_content (visited on 30/01/2019) (cit. on p. 186).

[W665] *Wikipedia: Page mover*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Page_mover (visited on 07/09/2016) (cit. on p. 262).

[W666] *Wikipedia: Redirect, section 'Do not "fix" links to redirects that are not broken'*. Wikipedia.org. 2007. URL: https://en.wikipedia.org/wiki/Wikipedia:Redirect#Do_not_%22fix%22_links_to_redirects_that_are_not_broken (visited on 30/01/2019) (cit. on p. 156).

[W667] *Wikipedia: Redirects for discussion/Log/2007 October 15#WP:RAD User:Lincalinca*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Redirects_for_discussion/Log/2007_October_15#WP:RAD_%E2%86%92_User:Lincalinca (visited on 30/01/2019) (cit. on p. 138).

[W668] *Wikipedia: Requests for adminship*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship (visited on 07/09/2016) (cit. on p. 263).

[W669] *Wikipedia: Requests for adminship, section 'Expressing opinions'*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship#Expressing_opinions (visited on 07/09/2016) (cit. on p. 263).

[W670] *Wikipedia: Requests for comment: Start date in NRHP articles*. Wikipedia.org. 2013. URL: https://en.wikipedia.org/wiki/Wikipedia:Requests_for_comment/Start_date_in_NRHP_articles (visited on 30/01/2019) (cit. on p. 166).

[W671] *Wikipedia: Research Help#How Reliable Is Wikipedia?* Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Research_help#Reliable (visited on 30/01/2019) (cit. on p. 9).

[W672] *Wikipedia: Researching Wikipedia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Researching_Wikipedia (visited on 30/01/2019) (cit. on p. 62).

[W673] *Wikipedia: Researching with Wikipedia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Researching_with_Wikipedia (visited on 30/01/2019) (cit. on p. 62).

[W674] *Wikipedia: Reusing Wikipedia content*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Reusing_Wikipedia_content (visited on 30/01/2019) (cit. on p. 62).

[W675] *Wikipedia: Size of Wikipedia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia (visited on 12/08/2019) (cit. on p. 7).

[W676] *Wikipedia: Sock puppetry*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Sock_puppetry (visited on 30/01/2019) (cit. on p. 399).

[W677] *Wikipedia: Stewards requests/permissions*. Edit as at 18:13, 22 April 2004. Wikipedia.org. 2019. URL: https://meta.wikimedia.org/w/index.php?title=Steward_requests/Permissions&oldid=32944 (visited on 30/01/2019) (cit. on p. 126).

[W678] *Wikipedia: Stewards requests/permissions*. Wikipedia.org. 2019. URL: https://meta.wikimedia.org/w/index.php?title=Steward_requests/Permissions (visited on 30/01/2019) (cit. on p. 143).

[W679] *Wikipedia: Subpages*. Edit difference: 13:47, 7 June 2006 vs. 15:32, 7 June 2006. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Subpages&diff=57366220&oldid=57353082 (visited on 07/09/2016) (cit. on p. 112).

[W680] *Wikipedia: Substitution*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#Substitution (visited on 07/09/2016) (cit. on p. 84).

[W681] *Wikipedia: Television episodes*. Edit as at 18:17, 4 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Television_episodes&oldid=693759627 (visited on 07/09/2016) (cit. on p. 104).

[W682] *Wikipedia: Templates for discussion*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Templates_for_discussion (visited on 30/01/2019) (cit. on p. 139).

[W683] *Wikipedia: Toolserver*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Toolserver (visited on 30/01/2019) (cit. on pp. 155, 170).

[W684] *Wikipedia: Transclusion*. Edit difference: edit as 15:22, 18 April 2008 vs. edit as at 20:16, 20 April 2008. Wikipedia.org. 2008. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&diff=206950560&oldid=206504873 (visited on 07/09/2016) (cit. on p. 82).

[W685] *Wikipedia: Transclusion*. Edit difference: edit as 14:05, 17 June 2011 vs. edit as at 00:39, 19 June 2011. Wikipedia.org. 2011. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&diff=440222218&oldid=434765744 (visited on 07/09/2016) (cit. on p. 82).

[W686] *Wikipedia: Transclusion*. Edit difference: edit as 06:14, 20 December 2013 vs. edit as at 13:25, 12 February 2014. Wikipedia.org. 2014. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&diff=595137158&oldid=586910323 (visited on 07/09/2016) (cit. on p. 82).

[W687] *Wikipedia: Transclusion*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756 (visited on 07/09/2016) (cit. on pp. 109–110).

[W688] *Wikipedia: Transclusion costs and benefits*. Edit as at 07:31, 28 September 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion_costs_and_benefits&oldid=683118661 (visited on 07/09/2016) (cit. on p. 110).

[W689] *Wikipedia: Transclusion costs and benefits, section 'Social'*. Edit as at 07:31, 28 September 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion_costs_and_benefits&oldid=683118661#Social (visited on 07/09/2016) (cit. on p. 111).

[W690] *Wikipedia: Transclusion, section 'Application of transclusion'*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#Applications_of_transclusion (visited on 07/09/2016) (cit. on pp. 109, 112).

[W691] *Wikipedia: Transclusion, section 'Basic syntax'*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#Basic_syntax (visited on 07/09/2016) (cit. on p. 78).

[W692] *Wikipedia: Transclusion, section 'Markup'*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#Markup (visited on 07/09/2016) (cit. on p. 81).

[W693] *Wikipedia: Transclusion, section 'Partial transclusion'*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#Partial_transclusion (visited on 07/09/2016) (cit. on pp. 79, 110).

[W694] *Wikipedia: Transclusion, section 'See also'*. Edit as at 07:44, 3 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Transclusion&oldid=693549756#See_also (visited on 07/09/2016) (cit. on p. 110).

[W695] *Wikipedia: Types of bots*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Types_of_bots (visited on 30/01/2019) (cit. on p. 122).

[W696] *Wikipedia: Types of bots*. Edit as at 23:41, 20 September 2015. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Types_of_bots&oldid=682011236 (visited on 30/01/2019) (cit. on p. 130).

[W697] *Wikipedia: Use of Wikidata in Wikipedia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Use_of_Wikidata_in_Wikipedia (visited on 30/01/2019) (cit. on p. 114).

[W698] *Wikipedia: Use of Wikidata in Wikipedia*. Original edit as of 06:31, 28 September 2017. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Use_of_Wikidata_in_Wikipedia&oldid=802749930 (visited on 30/01/2019) (cit. on p. 114).

[W699] *Wikipedia: User access levels*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:User_access_levels (visited on 07/09/2016) (cit. on pp. 212, 261–263).

[W700] *Wikipedia: User access levels*. Edit as at 22:58, 19 November 2004. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:User_access_levels&oldid=7662744 (visited on 30/01/2019) (cit. on p. 126).

[W701] *Wikipedia: User access levels, 'Table' section*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:User_access_levels#Table (visited on 07/09/2016) (cit. on p. 261).

[W702] *Wikipedia: Username policy, section 'Changing your username'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Username_policy#Changing_your_username (visited on 30/01/2019) (cit. on pp. 155–156).

[W703] *Wikipedia: Village Pump, (Archive 74): When & how to archive Wikipedia:Bot requests*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Village_pump_(policy)/Archive_74#When_&_how_to_archive_Wikipedia:Bot_requests (visited on 30/01/2019) (cit. on p. 177).

[W704] *Wikipedia: Village Pump: Unresolved bot requests*. Wikipedia.org. 2014. URL: https://en.wikipedia.org/wiki/Wikipedia:Village_pump_(technical)/Archive_124#Unresolved_bot_requests (visited on 30/01/2019) (cit. on p. 160).

[W705] *Wikipedia: What is an article?* Edit as at 18:54, 1 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:What_is_an_article%3F&oldid=697759098 (visited on 07/09/2016) (cit. on pp. 86, 111).

[W706] *Wikipedia: What Wikipedia is not*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not (visited on 30/01/2019) (cit. on p. 192).

[W707] *Wikipedia: What Wikipedia is not, section 'Wikipedia is not a newspaper'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not#Wikipedia_is_not_a_newspaper (visited on 30/01/2019) (cit. on p. 188).

[W708] *Wikipedia: Wikidata*. Revision as of 18:03, 4 November 2019. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikidata&oldid=924575631 (visited on 30/01/2019) (cit. on p. 114).

[W709] *Wikipedia: Wikidata, section 'Inserting Wikidata values into Wikipedia articles'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Wikidata#Inserting_Wikidata_values_into_Wikipedia_articles (visited on 30/01/2019) (cit. on p. 48).

[W710] *Wikipedia: Wikidata, section 'Inserting Wikidata values into Wikipedia articles'*. Revision as of 18:03, 4 November 2019. Wikipedia.org. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikidata&oldid=924575631 (visited on 30/01/2019) (cit. on p. 114).

[W711] *Wikipedia: Wikimedia Cloud Services*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Wikimedia_Cloud_Services (visited on 30/01/2019) (cit. on p. 170).

[W712] *Wikipedia: Wikimedia Cloud Services, section 'Migration from Toolserver to Wikimedia Cloud Services'*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Wikimedia_Cloud_Services#Migration_from_Toolserver_to_Wikimedia_Cloud_Services (visited on 30/01/2019) (cit. on p. 170).

[W713] *Wikipedia: Wikipedia is failing*. Wiki 'essay' dating from February 2007. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_failing (visited on 30/01/2019) (cit. on p. 192).

[W714] *Wikipedia: Wikipedians*. Edit as at 21:52, 8 January 2016. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedians&oldid=698891580 (visited on 07/09/2016) (cit. on p. 62).

[W715] *Wikipedia: WikiProject*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject (visited on 07/09/2016) (cit. on pp. 56, 150, 265).

[W716] *Wikipedia: WikiProject Cooperation/Non-controversial edits*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Cooperation/Non-controversial_edits (visited on 30/01/2019) (cit. on p. 163).

[W717] *Wikipedia: WikiProject Directory/All*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Directory/All (visited on 07/09/2016) (cit. on p. 265).

[W718] *Wikipedia: WikiProject Military history*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Military_history (visited on 08/09/2016) (cit. on p. 265).

[W719] *Wikipedia: WikiProject Television/Episode coverage task force*. Edit as at 13:39, 14 May 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Television/Episode_coverage_task_force&oldid=662301009 (visited on 07/09/2016) (cit. on p. 104).

[W720] *Wikipedia: WikiProject Wikidemia*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wikidemia (visited on 30/01/2019) (cit. on p. 62).

[W721] *Wikipedia: WikiProject Women in Red*. Wikipedia.org. 2016. URL: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red (visited on 08/09/2016) (cit. on p. 265).

[W722] *Wikipedia: Your first article*. Edit as at 23:31, 22 December 2015. Wikipedia.org. 2016. URL: https://en.wikipedia.org/w/index.php?title=Wikipedia:Your_first_article&oldid=696410329 (visited on 07/09/2016) (cit. on p. 111).

[W723] *Wikipedia:Editing policy*. Wikipedia.org. 2019. URL: https://en.wikipedia.org/wiki/Wikipedia:Editing_policy (visited on 30/01/2019) (cit. on p. 54).

[W724] *Wikisource (French): Traité de documentation (Treatise on Documentation), by Paul Otlet*. Wikipedia.org. 2019. URL: https://fr.wikisource.org/wiki/Trait%C3%A9_de_documentation (visited on 30/01/2019) (cit. on p. 311).

[W725] *Wikitech: Category:Toolforge*. Edit as at 06:01, 10 January 2019. Wikipedia.org. 2019. URL: https://wikitech.wikimedia.org/w/index.php?title=Category:Toolforge_tools&oldid=1812884 (visited on 30/01/2019) (cit. on p. 170).

[W726] *Wikitech: Portal:Cloud VPS*. Wikipedia.org. 2019. URL: https://wikitech.wikimedia.org/wiki/Portal:Cloud_VPS (visited on 30/01/2019) (cit. on p. 170).

[W727] *Wikitech: Portal:Toolforge*. Wikipedia.org. 2019. URL: https://wikitech.wikimedia.org/wiki/Portal:Toolforge (visited on 30/01/2019) (cit. on p. 170).

[W728] *xtools: Rambot - Namespace totals*. Wikimedia Foundation Labs. 2019. URL: https://xtools.wmflabs.org/ec/en.wikipedia.org/Rambot (visited on 30/01/2019) (cit. on p. 125).

List of Abbreviations

A | B | C | E | F | H | I | K | M | N | O | P | R | S | T | U | V | W

A

ACM

Association for Computing Machinery. 241, 243

AI

Artificial Intelligence. 8, 45, 185, 186, 196, 199, 209, 213, 225, 226, 243, 249

API

Application Programming Interface. 12, 64

ARPA

Advanced Research Project Agency. 395

B

BAG

Bot Approvals Group. 124, 126, 128, 131, 137, 139, 141, 144, 149, 169, 178, 195–197, 221

BAHYS

BAsic Hypertext System. 34

C

CACM

Communications of The ACM. 245

CAD

Computer-Aided Design. 240

CAI

Computer-Assisted Instruction. 29

CDN

Content Delivery Network. 4

CSCW

Computer Supported Cooperative Work. 195, 196, 217

CSS

Cascading Style Sheets. 249

CYOA

Choose Your Own Adventure. 20

E

EDS

Electronic Document System. 237

ELF

Evolutionary List File. 16

ELit

Electronic Literature. 15, 37, 233, 245, 246

F

FJCC

Fall Joint Computer Conference. 27, 235

FOSS

Free and Open-Source Software. 208

H

HAM

Hypertext Abstract Machine. 240

HES

Hypertext Editing System. 28–30, 234, 235, 241

I

IBIS

Issue Based Information Systems. 243

IBM

Internation Business Machines. 19, 21

IDE

Integrated Development Environment. 244

IGD

Interactive Graphical Documents. 237

IP

Information Protocol. 64

IR

Information Retrieval. 20, 22–25, 241

IRIS

Institute for Research in Information and Scholarship. 239

IT

Information Technology. 215, 217, 268, 269

K

KIM

Knowledge and Information Management. 269

KM

Knowledge Management. 241

M

ML

Machine Learning. 8, 185, 196, 199, 209, 213, 224, 240, 268, 269

N

NACA

National Advisory Committee for Aeronautics. 22, 25

NASA

National Aeronautical and Space Administration. 235

NDRC

U.S. National Defense Research Committee. 22, 25

NEH

National Endowment for the Humanities. 20

NIC

Network Information Center. 27

NIST

National Institute of Standards and Technology. 244

NLP

Natural Language Processing. 213, 224, 226

NLS

oNLine System. 27, 31, 234, 250

O

OCR

Optical Character Recognition. 24

ODIN

Open Data INnovation. 252

OHM

Open Hypertext Model. 245

ONS

Office of Naval Research. 25

OSRD

U.S. Office of Scientific Research and Development. 22, 23, 25

OWL

Office Workstations Limited. 237

P

PARC

(Xerox) Palo Alto Research Center. 27, 35, 37, 239, 244

PMS

Program Management System. 29

PROMIS

PRoblem-Oriented Medical Information System. 235, 236

R

RBU

Répetoire Bibliographique Universel. 20

RDF

Resource Description Framework. 206, 224

RN

Royal Navy. 9

ROI

Return On Investment. 269

RQ

Research Question. 60

RTF

Rich Text Format. 72, 397

RTV

Right To Vanish. 136, 137

S

SGML

Standard Generalised Markup Language. 245

SH

spatial hypertext. 206, 216, 224

SJCC

Spring Joint Computer Conference. 29, 234

SME

Subject Matter Experts. 210, 212

SRI

Standford Research Institute. 25–27

SUL

Single User Login. 135, 144, 146, 147, 278

SVG

Scalable Vector Graphics. 251

T

TED

Transclusions in Enterprise Documents. 44

TfD

Templates for Discussion. 139

U

UDC

Universal Decimal Classification. 20, 21

UID

Unique Identifier. 72, 322, 323

UNC

University of North Carolina. 240, 242

URI

Uniform Resource Indicator. 42

URL

Uniform Resource Locator. 7, 72, 108, 117, 323

USN

United States Navy. 25, 236

USSR

Union of Soviet Socialist Republics. 34

V

VKB

Visual Knowledge Builder. 247, 250

W

WE

Writing Experience. 242

Glossary

A | B | C | E | F | G | H | K | L | R | S | T | U | V | W | X | Z

A

Aquanet

This is a pun on the then-popular hairspray brand 'Aqua Net'. Just as 'Aqua Net' holds your hair in place, so the Aquanet system holds your information in place. Thus allowing papers titled "Aquanet: A Hypertext Tool to Hold Your Knowledge in Place" [234] and "Two Years before the Mist: Experiences with Aquanet" [231]. 247, 249

ARPANet

The **A**dvanced **R**esearch Projects **P**gency **Net**work (ARPANet) was the first wide-area packet-switching network, created by Advanced Research Project Agency (ARPA) in 1966. Its creation is detailed in Hafner & Lyon's *Where Wizards Stay Up Late* [152] and Waldrop's *The Dream Machine* [365, p.259–323]. It featured distributed control and was one of the first networks to implement the TCP/IP protocol suite. 27, 32

AutoWikiBrowser

The AutoWikiBrowser [W439] is a semi-automated mediawiki editor designed to make tedious or repetitive editing Wikipedia tasks quicker and easier. 99, 152, 157, 160, 161, 164, 169, 278

B

BRFA

Bots: Requests for Approval. 124, 126, 129, 137–140, 142, 152, 158, 169, 171, 178, 277, 287

C

CC-BY-SA

Creative Commons Attribution Share-Alike [87]. 62

CO

The Cabinet Office [75] is the UK Government's central ministry. It supports the country's Prime Minister and ensures the effective running of government across all departments. It is also the corporate headquarters for government, including general policy across the Civil Service (the country's non-partisan publicly-funded bureaucracy). The author was seconded from his university to the Cabinet Office for 3 months in 2019. 267

E

EDL

A Xanadu Edit Decision List as descibed at 'Project Xanadu' [292]. 17, 30, 257

F

FRESS

File Retrieval and Editing System (no last 'S'). 20, 235, 241

G

GFDL

GNU Free Documentation License [143]. 62

grep

A unix command-line program whose name derives from the Unix `ed` command `g/re/p` (**g**obally search a **r**egular **e**xpression and **p**rint). The function does a global search using a supplied regular expression and then printing all matching lines to `stdout`. 86

H

H-LAM/T

'Human using Language, Artefact, Methodology in which he is Trained'. This is the process expressed in Engebart's 'Bootstrap' and 'Augment' concepts: see Bardini [29]. 28

HTML

Hypertext Markup Language is the native coding method for encoding a Web 'page'. 6, 117

K

KMS

A backronym for 'Knowledge Management System' as the product was always referred to as 'KMS'. [14]. 237, 238, 398

L

linkbase

Originally, in pre-Web hypertext applications a linkbase simply implied a ‘link database’, e.g. in Microcosm (q.v. *Rethinking Hypermedia: The Microcosm Approach* [161, p.38, p.167]). The term has been defined within the XML XLink’ specification as ‘Documents containing collections of inbound and third-party links are called link databases, or linkbases.’ [389]. 206, 243

LST

‘Labeled section transclude’. N.B. Wikipedia uses the American English spelling of ‘labeled’ in documentation of LST first. 81–83, 91, 94, 97, 103, 109

R

RTFD

This is an extended version of the long-standing Rich Text Format (RTF) format. RTF supports styled test, whereas RTFD also allows instertion of images and some other formats into the text stream. The format is only fully supported on the Mac. 72

S

sockpuppet

A ‘sockpuppet’ is an online identity used for purposes of deception. See: Sockpuppet (Internet) [W140] Wikipedia: Sock puppetry [W676]. 138, 286, 294

T

technical debt

This terms was coined by Ward Cunningham, inventor of the wiki, and first defined in his article on the ‘c2’ wiki in an article reporting on OOPSLA ’92 ‘OOPSLA ’92 Experience Report’ [91]. 48, 61, 120, 126, 134, 135, 137, 157, 179–181, 190, 194, 202, 210, 212, 222, 225

U

unkown unknowns

‘There are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don’t know. But there are also unknown unknowns. There are things we do not know we don’t know.’ Donald Rumsfeld, US Secretary of State for Defence, in answer to a question at a U.S. Department of Defense (DoD) news briefing on February 12, 2002. Video of the quote [106]. The phrase is not, as often thought, a slip of the tongue but references the intelligence community’s use of the ‘Johari window’ after Luft and Ingham [220]. 64, 135

V

VIKI

'VIsual and KInethetic' analysis tool, though it is believed this expansion of the acronym was never published. The name has also been assumed to also be an homage to the character 'V.I.K.I.' (for: Virtual Interactive Kinetic Intelligence) in the Isaac Asimov novel 'I, Robot', though again source information is not clear. Information via private correspondence with Frank Shipman, co-author of the first VIKI paper [233]. 247, 250

VITE

Not an acronym, but a pun on *vite* the French word for 'quick' as the program made making structured data manipulation faster. (Information via private correspondence with Frank Shipman, one of VITE's creators.). 252

W

WP:IAR

Wikipedia essay 'Ignore all rules' [W656]. The concept that whilst rules need to exist, in considered cases they should be ignored. 142

X

XML

eXtensible Markup Language, a form of structured mark-up language.. 68, 71, 108

Z

ZOG

This is not an acronym, despite the uppercase form. It was intended to have no meaning but to be a short and memorable name. See Robertson [302]. Superseded by KMS. 236