

You've Got Hypertext

m.c. schraefel, Leslie Carr, David De Roure, Wendy Hall

Abstract

In this paper we consider possible “future everyday hypertext systems.” To ground our discussion, we look first at the functional and conceptual definitions of hypertext that have evolved in the hypertext research community. We then consider these definitions against the Web, the best known current everyday hypertext, but one that the hypertext community has regarded as only partially a hypertext system at best. We propose, however, that a full, rich hypertext is alive and well and living in an equally successful everyday system, and that that system is email. We look at how email meets the criteria, both functionally and conceptually for rich hypertext. We then use email-as-hypertext as our touchstone for assessing future hypertext systems. In particular, we consider the newest system on the Web event horizon, the Semantic Web, and show how the potential hypertextness of the Semantic Web has been anticipated by pre- and co-Web hypertext research systems. We consider how, if informed by the attributes of our email model, the Semantic Web may be able to break away from the limited hypertext model of the Web to become a rich, everyday hypertext system like email. We present three current hypertext research efforts that use the Semantic Web platform to show how these may be seen to embody such email-like hypertext qualities.

1. Introduction

Hypertext researchers over the past decade have debated the degree to which the Web is or is not a hypertext system. The hypertext folklore of the rejection of Web founder Tim Berners-Lee’s paper first proposing the Web to the 3rd ACM Hypertext Conference in 1991 has gone down as the first salvo of the debate. There is some agreement that, at best, the Web is a highly successful, limited version of a hypertext system: it supports non-linear navigation between predefined nodes and links. A more complete hypertext system along the lines of Halaz’s Reflections on NoteCards paper [Halasz97] would provide a more complete definition of the attributes of a fully implemented hypertext system. Attributes like annotation, user-authored links, typed links, and calculation would support, in functional terms, more what hypertext is said to be about. In conceptual terms, these functions would better support the more visionary notion of hypertext as implied by Vannevar Bush’s “As We May Think” [Bush45], where a researcher can construct paths of *associations* through various documents that have been captured both by the researcher and by other authors within and beyond the researcher’s own domain. The concept of human-determined association-building as part of a knowledge-building activity is presented as one of Bush’s core rationales for collecting and storing information. The Web’s emphasis on pre-authored, embedded links has been one of the greatest constraints for enabling this conceptual aspect of hypertext. So when Tim Berners-Lee suggests that the Semantic Web, with its emphasis on machine-based reasoning for better information retrieval on the one side and service discovery for more powerful application delivery on the other, is the Web’s next evolutionary step [Berners-Lee01], hypertext researchers may wonder if this is one more step away from the conceptual and functional model of hypertext [Marshall03].

In this paper, we propose that the Semantic Web has the potential to become a rich hypertext system. Indeed, the Semantic Web’s service-oriented platform lends itself more effectively to hypertext development than does the Web. Additionally, the emphasis on semantics allows a far greater range of dynamic linking possibilities than the Web. One of the goals of the Semantic Web is to provide semantic structures for content to be represented. Such structures are simple but powerful relations. Semantic information on a page might indicate that this paper is authored by that person. An ontology that defines

these spaces might say that an author is a person and that people have home pages. A separate piece of semantic information might say that this person's home page is web page X. Given that information, the computer can reason that the first page should be linked to page X. This link is new information, created dynamically by machine-based reasoning over the relations expressed in the semantic representation of the pages. Semantic Web research is informing a growing number of research initiatives including the Global Grid Forum and eScience [De Roure03]. Its reach into the business community is also beginning [Berners-Lee03]. The Semantic Web is fundamentally designed to be delivered using Web protocols and so far, via Web browsers. While it is a powerful concept, the Semantic Web is still in flux. The emphasis in research is very much on low-level protocols for connecting semantic services with data and the relations expressed among the data. Most emphasis, in other words is on the machine side of the Semantic Web: the logics and mechanisms to bring together *related* information and manipulators of that information. Little has yet been explored about models of use and consumption on the human side. How, for instance, might we use these new inferential powers to support real scenarios of knowledge building for human users across disciplines and of varying degrees of expertise? In terms of hypertext, how might we use the semantic affordances of this new technology to support user-determined or user-directed association-building, or paths in the Vannevar Bush sense of the term?

Because the Semantic Web is still in its early days (it made it's public debut in Scientific America in 2001 [Berners-Lee01]), we have the opportunity to engage with the research in this space to bring to it these kinds of hypertext qualities we have not perhaps seen in the Web, outside research systems presented at conferences. There is a real possibility here to help chart the course of the Semantic Web's hypertext path. If we are concerned, however, that the Web is a poor model to inform the Semantic Web in terms of a fuller range of hypertext possibilities, we need to point to models which can help us make apparent both to ourselves and to the Semantic Web research community what we mean by "real hypertext."

In the next section, we look at exactly what we mean by hypertext, beyond the commonly understood hyperlink. We also review some early efforts to incorporate hypertext with the Semantic Web, and lessons learned from these systems for moving forward our consideration of what hypertext means in the Semantic Web. By way of counterpoint to this discussion, we present an alternative approach to understand the possible relation of hypertext and the Semantic Web: we present email. The value of considering email-as-hypertext is that it is a system equally as popular and as successful as the Web. It also scales. One might argue it is as elegant in its simplicity as the Web. And yet, it has, as we will show, hypertext qualities on both an explicit and implicit level that more closely models the functional and conceptual attributes of hypertext than the Web. . If something as simple and direct as email can capture rich hypertext, we argue, surely the Semantic Web can, too.

In the sections following our discussion of email, we describe several research efforts currently underway where we have been working to bring the kinds of hypertext attributes we find in email to the Semantic Web. We present

- a formal hypertext interaction model called mSpace which we have incorporated into a Semantic Web application called CS AKTive Space
- an interactive hypertext writing application called WiCK which pulls in appropriate documents for reference while writing
- eScience, a large suite of projects in the Grid and Semantic Web space to help scientists communicate their work-in-progress to their communities for parallel use.

2. Understanding Hypertext: What is It?

It is a salutary experience to hear many Semantic Web researchers talk about hypertext research, if they talk about it at all, as something in the past tense. In invited talks, Jim Hendler (a leading Semantic Web researcher) [Hendler02] speaks about the success of the Web being based on not making the mistake of the hypertext community in its perceived obsession with not having dangling links. The Web allowed for broken links and so thrived, despite certain rough edges, whereas hypertext, with its insistence on better link management or nothing, withered. This is his warning to the Semantic Web community not to get bogged down in seeking perfection, but to succeed through the “scruffy works” approach advocated by Marvin Minsky for AI [Minsky91].

While Hendler’s point may be correct for the successful deployment of large scale distributed link-based systems¹, it misses some of the more profound aspects of hypertext described over the past sixty years by what has come to be known as hypertext research. In this section we first review those other aspects of hypermedia which inform our approach to the Semantic Web. This is largely a focus on hypertext-as-interaction with information to build associations, and through associations to build knowledge.

2.1 Hypertext as Association

We begin with Vannevar Bush, whose Memex has usually been referenced as the archetypal hypertext system. It was postulated mainly as a tool for researchers. The Memex itself was proposed not as a repository of linkable sources, but as a space for creating associative paths through information, and through this association, developing knowledge and understanding. The Memex stands on its own as a non-networked Desk, where the Researcher may reflect upon resources contained in the machine, make associations, reference new connections that would otherwise not be available without the support of the Memex. The researcher engages in these practices not only for “his” own benefit, but also to enrich what he can bring to his collaborative work with others. This desire to enhance knowledge is later specifically captured in Engelbart’s Augment system [Engelbart68].

In Literary Machines [Nelson87], which captures much of Nelson’s founding work on Hypertext since the 1960’s when he coined the term “hypertext,” Nelson indirectly extends the concept of the Memex by adding the notion of a network to an associative file system. The network extends the scope of the system to allow multiple authors to forge their own paths among a set of documents, as well as to add their own documents and annotations among the annotation space. These comments and contributions can be constrained in terms of who may see a given contribution, but everything – all writing – is held in this space, this “perma scroll.”

2.2 Hypertext in Contrast to the Web

It is this intertextuality [Kristeva86] of hypertext, captured in the early (pre-Web) systems like StorySpace that fire the imaginations, and critical writings, of the nascent hypertext research community. To use literary theory parlance, hypertext is imagined both as readerly and writerly. To use strikingly similar Unix parlance, hypertext is a read-write system, whereas the Web is experienced mainly as a read-only system; only the author of the document could author links through it. Transclusion, stretch-text, [Nelson99] annotation were lost in what became a broadcast medium that

¹ It may be interesting to note that in commercial publishing enterprises dangling links have been treated as mission critical. Significant effort has been put into making unbreakable references, resulting in the DOI (Digital Object Identifier) system. Here, “scruffy” does not work: reliable access is required as it affects the bottom line.

afforded fantastic delivery of pre-authored, but effectively untouchable content. The discursiveness imagined in Literary Machines and embodied in a system like Microcosm's user-determined links and annotations [Davis92] is gone. True, the Web provides "universal" access to information, but even the primitive idea of the Memex was predicated on this as a starting point for discourse, not as an end in itself.

It may be easy to see where Web and Semantic Web researchers have come to conflate hypertexts with links rather than associations as hypertext research in the moments around the start of the Web were increasingly focused on the aspirations of developing an Open Hypermedia System standard [Carr99].

One of the critical parts of this endeavour was the definition of data models to represent links. The seeds of this perceived need for a strong understanding of link properties can be seen as early as 1985 when system definitions begin to merge with interaction descriptions. In *Reading and Writing the Electronic Book*, Yankelovich *et al* state:

"The first essential capability of a good electronic document system is to provide a means for promoting the connection of ideas and the communication between individual scholars...the basic capabilities implied by the terms hypertext and hypermedia include linking together discrete blocks ... to form webs of information, following different paths through information webs, and attaching annotations to any block of information."

[Yankelovich85]

Two years later, Conklin's influential tutorial article in 1987 defines hypertext both as windows on a screen reflecting linked objects in a database, *and*, quoting Nelson's atypically systems-type definition, as: "a combination of natural language text with the computer's capacity for interactive branching, or dynamic display...of a nonlinear text...which cannot be printed conveniently on a conventional page" [Conklin87].

The following year, John Smith and Stephen Weiss as the editors of the CACM Special Issue on Hypertext [Smith88] describe hypertext next as "an approach to information management in which data is stored in a network of nodes connected by links." In the same issue is Halasz's Reflections on NoteCards: Seven issues for hypertext. The text is a split between a traditional hypertext-as-associations approach and what the system requirements of the system may be. Halasz lists linking, annotation and computation each as possible parts of a rich hypertext system that enables people to use the systems to share, publish, subscribe to and annotate ideas. Halasz's definition from this system gives precedence to hypermedia as linked documents, rather than *linking* documents "Hypermedia is a style of building systems for information representation and management around a network of multimedia nodes connected together by typed links." From such a definition, one would say that the only attribute of hypermedia the Web dropped was the typed aspect of the link. The retrospective on Intermedia/IRIS Hypermedia Services [Haan92] refers to hypermedia as "the capability to create and browse through complex networks of linked documents." It can certainly be argued that the Web embodies such a model of hypertext. Indeed, their concluding definition of hypermedia goes beyond the notion of linked documents to a richer hypertext, which anticipates the Semantic Web: "our vision of hypermedia is based on a belief that the computing environment should be able to reflect the ever-changing connections inherent in all information."

2.3 Hypertext in the Era of the Web

This review of the evolution of perceptions of hypermedia shifting to more link-based than association-based models is not to suggest that hypertext research has exclusively focused on hypertext as nodes and

links. Indeed, in the period before the Web, applications like Intermedia [Meyrowitz86] and StorySpace [Bernstein91] enabled experimental authoring in hypertexts as nouveau forms of writing and reading, informed by postmodern theory's indeterminacy [Yellowlesse Douglas94] and intertextuality [Landow92]. Applications like Bill Atkinson's HyperCard [Goodman87] and Southampton's Microcosm brought interlinking between applications and documents to the desktop. In the past ten years of the Web Era, however, the Web has left behind many of these interaction-rich forms of hypertext. It's perhaps therefore not a surprise that hypertext conference publications throughout this period have been largely engaged in software efforts to engineer the Web into something more hypertext-like. Work in link annotation [Bailey01], and adaptive content [De Bra99] have been focused on the web-as-platform and interaction constraint. With the exception of Eastgate System's well-received stand-alone application Tinderbox², even research in spatial hypertext to support associative knowledge building has largely been involved in how to port this model to the Web [Shipman01]. Beyond the realm of hypertext research, however, the Web has remained largely resistant to these efforts.

3. eMail

One could see the lack of take up of these research-prototype hypermedia additions to the Web as an implicit criticism in the hypertext concepts on which these extensions are founded: they are simply not as plastic as the very successful Web, which started life similarly as a research prototype at CERN.

We can however point to a system that, while it did not come out of hypertext research, meets the definition not of the reduced Web set of hypermedia traits, but of a Nelsonian hypertext system, technically and conceptually. We refer to email. Email, like hypertext, overrides conventional expectations of sequence and linearity and blurs the distinctions between authorship and readership. As such it is, more so than the Web or its precursors, perhaps the best example of a large scale, functional, open platform hypertext system. In the following sections, we represent email as hypertext in order to investigate the particular hypertext affordances of email as something something how those affordances may be used to inform the road ahead for future hypertext systems.

3.1 You've Got hypertext

In the following sections we make our case for email as hypertext by analogy. We describe aspects of a typical email exchange. The following took place between various members of the authors' research group. The complete exchange is published in a HyperMail archive at [Harnad02].

² Eastgate Systems Inc., www.eastgate.com

Figure 1 shows one bar for each email, with the y-axis recording the number of lines in each message. The bars are coloured according to the origin of each line in the message – black indicates a line which is native to that message, whereas white indicates a line which is quoted from a previous message. This figure clearly demonstrates the hypertextual intertextuality [Genette97] exhibited by the so-called quote/commentary [Harnad03] in an email exchange.

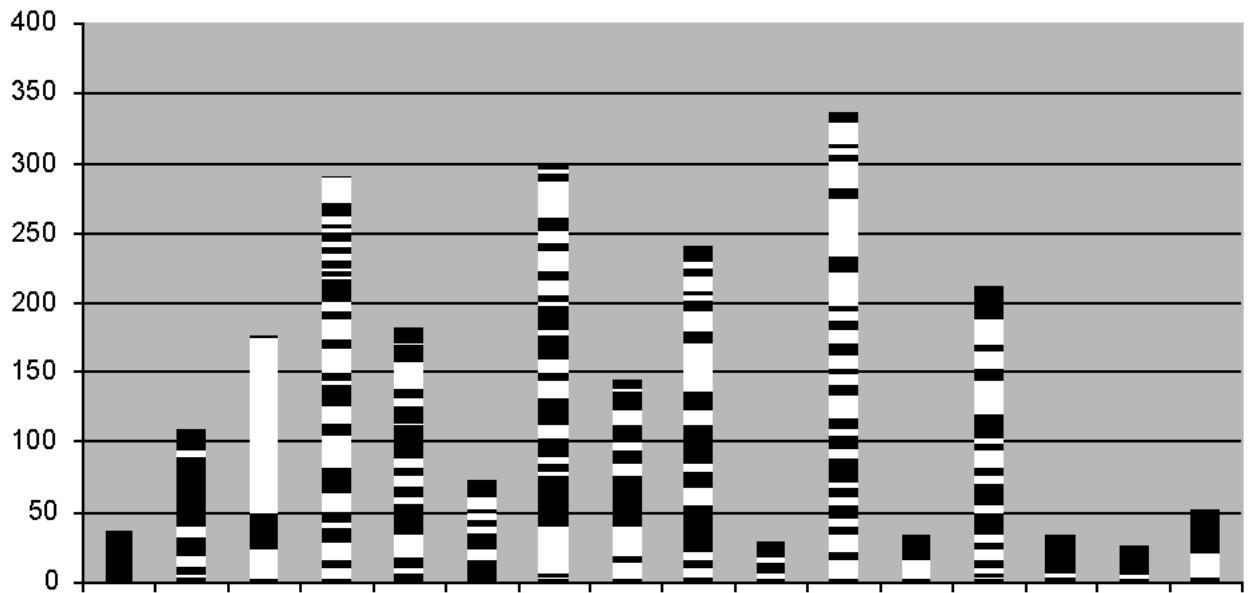


Figure 1: Original vs Quoted Text in Email Exchange

The first email is the seminar announcement and is entirely original. The second email picks out several quotes which are used as the basis of a dispute. (The small quote at the beginning of the message is a single line containing the title of the seminar, used to establish the context.) The third email contains a solicited contribution from the seminar speaker (lower white block) followed by a single uninterrupted block of response, before finishing up with the entire previous email being repeated at the end of the message (an artifact of the mail client). After this point, the quotations and comments in subsequent emails demonstrate a significantly increased granularity. Individual paragraphs or sentences may be the target of subsequent responses, with up to 18 quote/comment pairs per message.

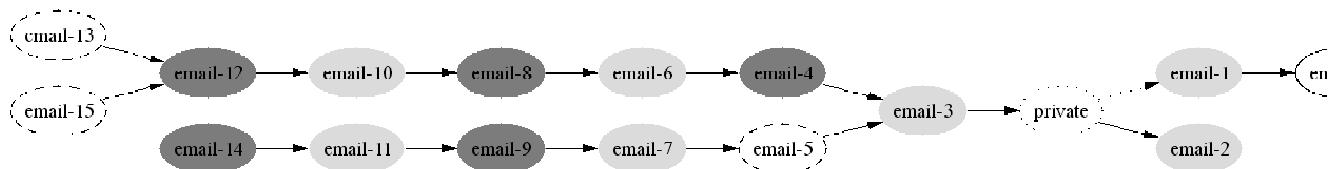


Figure 2: Dependencies between Messages

The figure serves to highlight the extent to which such a common form of writing demonstrates significant hypertext features – even though there is only minimal computer support for it and it has no explicit, encoded links; these remain implicit, indicated by quotation marks and date referencing of quoted parts within the messages. The diagram fails to record the *depth* of intertextuality, *i.e.* the extent to which one quoted fragment contains other quotes from previous messages.

The fundamentals of email reflect another core attribute of hypertext: authors and readers have similar status. Each email can be quoted and annotated upon and redistributed and requoted and redistributed ad

infinitum. No formal storage mechanism is defined, so no recipient or author has a privileged position of being allowed to make the “official” record. The flexibility of addressing allows a completely controllable distribution, much more flexibly than can be obtained by storage in a Web or FTP server.

Figure 2 shows the relationship between the emails. The original email is to the far right. The initial definition of the problem soon leads to a split in the discussion. The emails which match a particular correspondent/author are given a unique style – the discussion(s) can be seen to be driven by the original disputant (light grey) interleaved by two other participants (dark grey and dashed-white).

Figure 3 demonstrates the transclusive ordering of fragments in a particular email. The lines connect the quotes in the later email with the sections from which they were taken. Mostly, the quotes are taken in scroll (reading) order down the document. However, an early (out-of-turn) response is given to a piece of material which occurs near the end of the original email. Also, several apparently individual quotes are joined from separated blocks of material in the original email.

3.2 Faux or Fay hypertext?

Taken together, these diagrams offer evidence for collaboratively authored commentary/annotation, quotation, ordering and multilinearity in email. A reason why email might not be considered as “true” hypertext is that these hypertexts are not “stored” and the links/quotations are not explicitly coded. In fact, there is little help from the email environment in the creation of all these features – the details of storage and intertextual manipulation have to be controlled by the user.

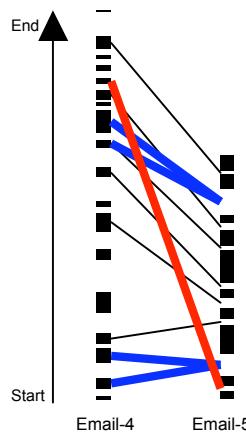


Figure 3: Ordering of Text in Response

If stored as a single mailbox, a discussion such as the above can be processed to provide explicit hypertext features, just as HyperMail does for a Web browser. However, this form merely mimics the functionality of an email client (with the advantage of open, distributed viewing) and notably does not reconstitute the intertextuality of the discussed material. Instead it concentrates on recovering the threading of the original messages, providing tables of contents ordered by date, subject and author.

The disadvantage of this state of affairs (level of computer-support) is that the user (correspondent) is forced into sculpting the form of their message – deleting, inserting, cutting and pasting, discarding and repositioning. Having gone to this effort, the ‘link’ with the original words in the original position in the original email is lost and has to be reconstructed by hand (or by search engine – which HyperMail does not provide).

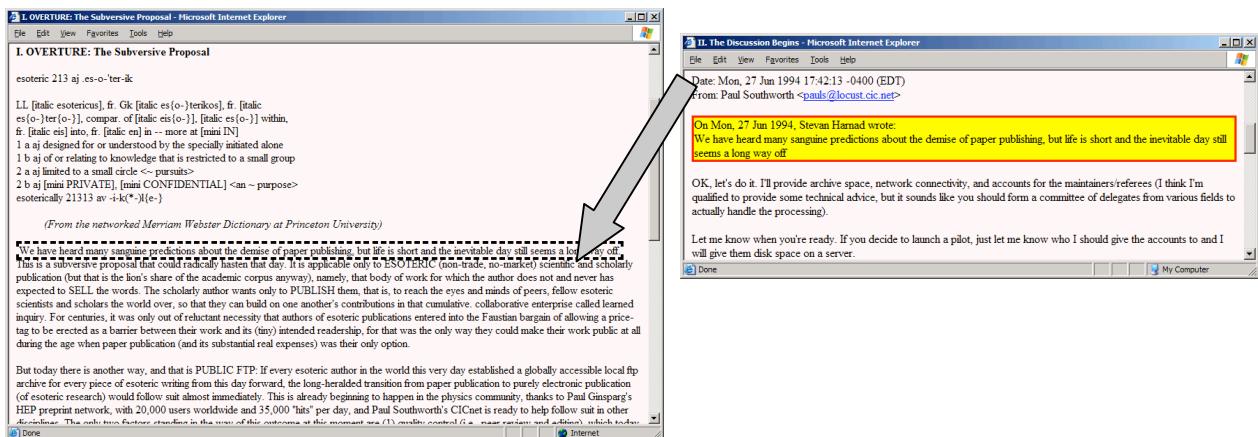


Figure 4: Email quotation rendered with transclusion and link

Figure 4 shows a simple experiment³ in which the quoted text of an email is rendered by a transclusion – the quoted text is fetched at display time from the body of the original email. Clicking on the quote itself triggers a link which brings up the quoted email highlighting the context of the quotation. The example is from a different email discussion, but illustrates the advantage of actualising the intertextual relationship – the ability to follow easily through a chain of argument.

3.3 No, it's hypertext

The seeming disadvantage listed above of the necessity of manually sculpting a response using (effectively) a text editor can be seen in quite a different light. By automatically copying the text of a message for which a response is to be made, the system *allows* the user to work in a *natural fashion* with the material with which they wish to engage. The ability to edit text, although not innate by any means, is at least reflective of a fundamental skill in a computing environment (even the most ardent hypertext practitioner would be hard pressed to argue the same about *making links*).

Quote/commentary [Harnad 2003] is a particular email rhetoric which has arisen naturally out of this simple facility of the mail client's copy on response. Although not present in the earliest UNIX email clients (the mail client in the 1979 Version 7 UNIX did not even have a 'r'eply command), the ability to isolate specific points in a message and respond to them individually provides a genuine enhancement to normal patterns of communication. In the face-to-face synchronous medium of speech the flow of dialogue does not allow for such constant interruption whereas at the other end of the spectrum the time-delay in a written or printed exchange requires the kind of batch rhetoric seen in scientific articles.

The remaining characteristic of email which is of particular relevance to the Semantic Web is the *default* tagging of each message with identifiable semantic attributes in the header: the date, author, subject and details of distribution and delivery. These not only facilitate the normal functions of an email environment (sorting by sender, threading on subject *etc.*) but also support spam blocking, message filtering and issue tracking.

Email has of course been adapted to the web environment. As well as being considered a hypertext in its own right, each message can carry URL references to Web pages that modern mail clients will turn into clickable hypertext links. Such chimeric hypertexts take the form of a dynamic email conversation about

³ The transclusion support used here is provided by the *VTLit* module of the EPrints.org archiving software.

a static web resource. Visually, the message may even take the appearance of a Web page - using the same HTML rendering - but the effect is the same. Even further behind the user experience, at the protocol level, the MIME standard provides a (rarely used) way of inserting a URL to specify the content instead of the content itself – so that rather than having huge bodies in an email message, the content can be pointed to instead of transferring them, thus offering an instantiation of transclusions directly.

In sum, email-as-hypertext provides us with a successful model of an *everyday* hypertext system which embodies the core properties of hypertext such as sharable, linked information, annotation, some control of dissemination from one to the group to the world. The ebb and flow of an email discussion has the potential to afford the kind of knowledge building through association that is the founding rationale for Bush's Memex.

4. Hypertext meets the Semantic Web

In the last section, we showed how email represents a possible embodiment of the strain of hypermedia which is focused on what might be called the human side of hypertext: the ability to engage with documents by annotation and linking those documents, and by being able to share those constructions with a participating community. This practice of engaging with text in a user-determined way is part of creating new meanings from exploring these associations and annotations. In the following section, we look at hypertext's first efforts to engage with the Semantic Web in order to provide some of this kind of human-centered overlay on top of the machine-affordances of ontologically driven, Web-delivered information. The result, as we will see, is not exactly the fluidity of email-as-hypertext. Looking at these examples, however, helps us tease out where the role of hypertext is readily seen to be of value in enriching the Semantic Web endeavour with meaning-making. In the section following this, we will see where these lessons have taken us in terms of the future hypertexts we describe.

The Semantic Web is driven by the belief in the potential of a more machine-processable Web, where data can be represented and understood against suites of ontologies that describe concepts and their relations to other concepts within an information space. Early examples of the use of the relations between concepts as part of a hypertext system include contextually aware, ontology based linking:

There are many other possibilities, a synonym generator applied to selections could expand the number of links found in a sparsely linked system, whilst an intelligent filter could build a context model of the user's requirements and restrict inappropriate links in a heavily linked system to help reduce 'link overload'. [Hill92]

This evolved into the notion of the "synonym filter" in Microcosm and in Mavis, became the multimedia thesaurus. The similarities to this approach and ontologies are clear in the following citation:

One way to approach a solution to this problem for text is to introduce a digital thesaurus. This should contain all the vocabulary (words or terms) appropriate to the application domain, arranged as a network with at least the following relations: broader term, narrower term, equivalent term and related term. The thesaurus could be invoked either at the link authoring or the link following stage or both. In the first case, if the thesaurus is invoked when a generic link is authored, links could be automatically generated and stored from all terms equivalent to the selected source anchor term. This would increase the linkbase size but have the advantage of increasing the speed of link following. Alternatively, the authoring stage could remain unchanged, ie a single link is generated from the author's selection, but at the link

following stage the user's selection is expanded via the thesaurus to a set of equivalent terms and each is compared with the linkbase source anchors to find any available links. This will be less expensive in linkbase storage but will slow down the linkbase search. [Lewis 96]

Subsequently other hypermedia environments have staked out territory within the Semantic Web specifically. Indeed, following from MAVIS, the COHSE project [Carr01] uses semantic models (ontologies) of the subject domain of a page (rather than thesauri) to add links to pages in principled fashion. By using the semantic model to understand the terms in the document it becomes possible to make sensible decisions about linking – for example, should links to be added to a concept in a page which is in any case a resource devoted to that concept.

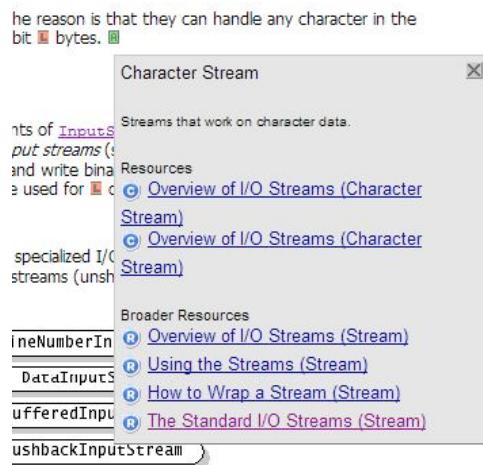


Figure 5: COHSE Linking in the Java Tutorial

The benefits of the ontology come from the existence of a conclusive set of natural language terms plus a description of the relationships (especially the subsumption relationships defining super- and sub-concepts) between those terms. It is the terms which can be used to suggest linking opportunities; it is the relationships that provide the linking system sufficient information to be able to start to judge the applicability of the links.

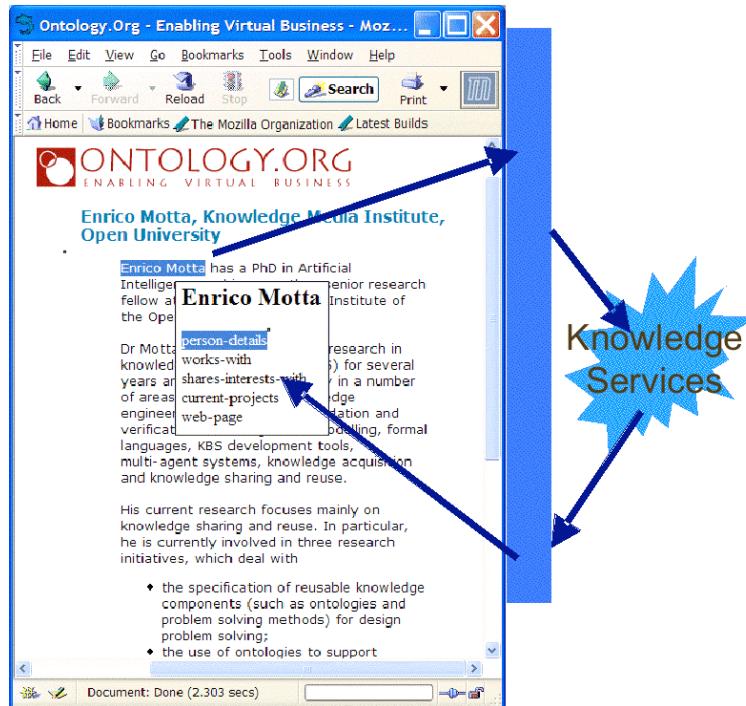


Figure 6: AKT-2 Semantically-Enhanced Portal

AKT-2 (a simple web portal interface developed using the knowledge management tools of the AKT project [AKT01]) demonstrates a similar approach. Like COHSE, it adds links to pages based on the appearance of semantically recognisable terms on the page. Unlike COHSE, the terms are named entities (a particular person, place or paper) rather than more abstract “concepts” (“lecturer”, “university” or “journal article”). Figure 7 shows a particular name being recognised by a “knowledge service” (a set of queries operating on a large knowledge-base), and responding with a menu of categories appropriate to that entity. In this case, a researcher’s name has been identified, and so “personal details,” “works with whom” and “current projects” are some of the choices available.

The weakness of these systems (and similar ones such as CREAM [Staab02] and Magpie [Domingue03]) is that they focus on ‘knowledge fragments’, recognising concepts or instances of concepts buried inside the text. Linking these occurrences certainly can be useful for reference purposes, just as dictionaries and thesauri are useful, but perhaps readers need more and other support than simple reference [Sellen02]; we already know that there is more to hypertext than putting links in nodes. The disadvantage of these systems is that they tend to simply regurgitate facts in response to mouse clicks. The interaction model for both systems is “click and tell”, triggered by isolated phrases, words or entities rather than (as one would hope in a semantic web) deeper issues of understanding of the text.

4.1 Adamantine Links

One way to model what we mean by providing users with better support for their knowledge building tasks can be shown with what Harold Bloom referred to as a “Taxonomy of Educational Outcomes” (Table 1). The taxonomy lists the ways in which students should, and our systems might, engage with knowledge and put it to use. The systems above function at “level 1” – responding to questions by simply recalling facts. Higher levels of ability require that the knowledge that is stored in a semantic web, be deployed in a more sophisticated fashion.

Table 1: Bloom's Taxonomy of Educational Outcomes
[Bloom 1965]

6. Evaluation	Ability to judge the worth of something
5. Synthesis	Ability to combine separate elements into a whole
4. Analysis	Ability to break a problem into its constituent parts and establish the relationships between each one
3. Application	Ability to apply rephrased knowledge in novel situation
2. Manipulation	Ability to rephrase or paraphrase knowledge
1. Knowledge	That which can be recalled

Another example of this behaviour is a Digital Library which provides reference linking. Although not necessarily a part of the Semantic Web (no use of RDF and related standards), this is nevertheless an example of information previously buried in the text of a document being made explicit. However, it is still stuck at the base level of Bloom's taxonomy – the recall of explicitly stated relationships. By contrast, ESKIMO [Kampa02], is an example of a system which attempts to build on these basic bibliographic relationships, augmented by other community relationships (membership of institutions, contribution to projects *etc.*), to offer links from an article to the *seminal papers* or to the *experts* in the field. The dimension in which linking is being accomplished is inference and abduction in the knowledge domain; the domain of the link is not a micro-fact, but the paper as a whole. Such associations are certainly moving towards Bloom's higher levels of analysis, synthesis and evaluation⁴.

The systems mentioned so far in the Semantic Web area rely on the very Web-like interaction of clicks-on-links (what we call 'clinks'). They are constrained by these clinks for the types of services they can deliver. For instance, it is particularly obvious in ESKIMO's user interface that a link anchor has to be generated – an artificial place for the user to click, in fact a separate knowledge navigation pane embodying a choice of semantic queries which reduce even the most complex evaluation to the appearance of a knowledge recall exercise: who is (an expert)? which is (the seminal paper in this field)?

This clink-only interaction begs the question: can web services only mean "that which can be explicitly linked"? If so, will that mean that the Semantic Web, with its familiar Web interface, will likewise be constrained to exist at the lower levels of Bloom's taxonomy? We would suggest that in order for the next generation web services to support more than information retrieval or knowledge recall we need to move from supporting only the clink of the Web to supporting the explicit and implicit associativeness of email model.

⁴ No *intelligence* is being claimed here; once a formal model of the bibliographic and community relationships is obtained, and a satisfactory set of rules is produced which matches human intuition about the evidence which supports 'expertness' then everything reduces to 'data manipulation'. This is the core of the Semantic Web – appropriate modelling and appropriate inference leading to semantically grounded inference.

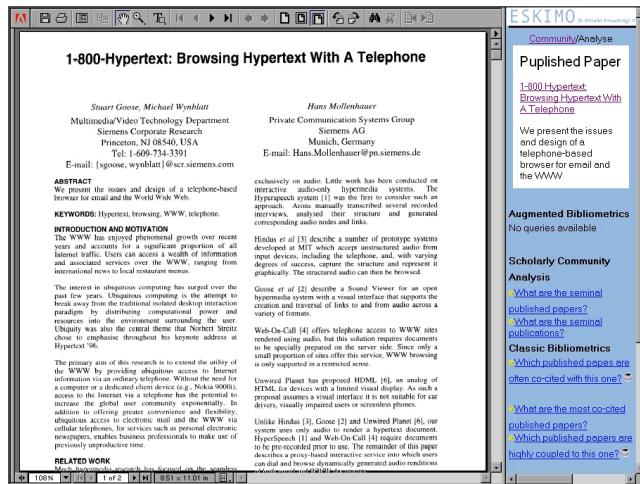


Figure 7: ESKIMO Assisting the User towards Evaluation

4.2 Break the Chain

Email, as we have seen, affords a space for reflection, annotation, evaluation, development of ideas, collaboration, scheduling, and more. Its main enabling features, in a modern mail system, are a quote commentary reply function, clickable URLs, rudimentary text editing functions, and search. As we have stated, it is the high degree of manual-ness enforced by the tool that affords the creation of the kind of exchange modelled in section 3. This direct manipulation is not a bug; it's a feature. The tools allow users to do what they wish to do in the email space – annotate, comment, link, collaborate, *i.e. make hypertexts*. The makers of the Semantic Web may wish to consider that it is this conceptual level of hypertext, enabled by email-as-model, rather than only the explicit, link level of technical hypertext where the higher levels of Bloom's knowledge building taxonomy will be enabled. In other words, it will be important for the Semantic Web to afford user-determined *implicit* link making as well as its current system-determined explicit link offering.

In the next part of the paper we look at some examples of future everyday hypertext work that are driven by this model. We consider three systems in progress: mSpace, WiCK, and eScience:

- mSpace is an approach to support the exploration of domains from user-determined perspectives.
- WiCK is a writing augmentation environment to support a person's writing task when that task means having to pull together document information from multiple sources. WiCK's hypertext model supports making appropriate documents and information available to the user while working so that they do not have to shift focus to look for data while they're thinking thoughts about that data.
- eScience hypertext proposes a kind of pervasive hypertext where information is made available on demand and in context.

5. Future Hypertexts in the Semantic Web

As described above, Web-based hypertext foregrounds the clickable link paradigm of hypertext as non-linear navigation. The success of search engines and commercial services has further constrained the link as navigation to the link mainly as page clicks in digital catalogues. Search engines return lists of links. If we click on more than one it is only in order to find a page that better matches our search

interest. This is only the most limited sense of navigation – it is a kind of hub and spoke use of links: we click out to a result; we click the back button to return to the index of search results. The items in the search list are not associated in any other way than in being part of this transitory search list based on keywords and keyword frequency algorithms. One does not get a sense from the results that they may be representative of the domain of which they are a part. Tools like Vivisimo [Caroll03] have developed methods for light categorization of search results so that for instance, search results for the keyword search “dolphins” will organize results for the football team and the mammal in distinct parts of the results page. But these sub categories themselves still do little to give the searcher a sense of where a given instance/document fits into the domain of which it is a part. In other words, we cannot tell how this document is associated with any others in that space. We get no context.

Context has always been a critical part of hypertext research. User-determined associations provide context. The quote-commentary of email makes context explicit. In the Web era, spatial hypertext has been developed as a hypertext model that privileges the construction and maintenance of user-determined contexts for collected information. Tools like VIKI [Marshall94] and the commercial Tinderbox let users create domains in which associated elements of information can be captured, annotated and organized visually in associated piles, and piles in associated blobs. The piles represent ‘implicitly’ linked groups of information. Tinderbox adds agents as a mechanism to sift through these collections to propose additional possible associations that may be of interest to the user. Tools like Hunter Gatherer [schraefel02] also support user-determined contexts for a more constrained class of information: selections of content from *within* Web pages. Users indicate components *within* Web pages which they wish to include in a new Web page. The new Web page acts as a container for these components. The links to the source documents are maintained with the selections from the various pages, thus enabling transclusions [Nelson99], the referencing back by link to the source of the component. These tools have proven useful for supporting the collection and management of personal, associated information spaces as rich hypertexts: in each case, users can annotate their collections, rearrange them, edit them, and so communicate new information as much by how they are organized as by the annotations within them.

Both of the above approaches rely on repurposing content from one source into another source – and in that other source the user creates the associations, arrangements and annotations over the selections. While powerful for individual or potentially collaborative work, these tools focus on user-determined reconstructions of information and creations of context. There are times, however, where we may wish to leverage the context of the source itself in order to build up new knowledge about the domain of that source. Similarly, we may wish to use that information to help us compare or contrast instances within a shared context. This comparison/analysis is fundamentally the nature of scholarship: comparing work within related or among associated domains. Prior to the Web, this approach was also how we frequently initiated ourselves into a new domain. Going to a library meant we could quickly get a sense of the scope of a topic by seeing how many books it occupied on a library shelf relative to others. We could suss out the relations in the topic by looking through the “card” catalogue. The library itself supported multiple mechanisms for reaching beyond its own walls to supplement its physical assets: indexes in the reference area organized additional information into associated categories; interlibrary loan arrangements meant that books could be brought in on request. Most of these observations are all at the meta-level of the topic, revealing its context, and yet this meta-knowledge can be fundamental to an understanding of the topic. Hypermedia concepts like multi-pointing links, linkbases, and dynamic links [Herzog95] have sought to support these kinds of associations for topics such that the word “camera” might spawn a categorized list of associated links from History Of, to Brands, to Optics involved.

Implementations of such link bases have largely been limited, however, to locally controlled sources of data.

The unstructured or semi-structured approach of the Web does not provide for the automated construction of associated contexts of information. The closest Web analogue to this kind of contextual information is the Web store where its limited catalogue shows how many models of a given object it carries, along with pricing and product information. In this case, the store, as in past link base systems, knows only about the content over which it has control. Other Web applications do crawl the Web to carry out comparisons on attributes like prices on the same objects at different stores. While useful, these kinds of services do not provide an integrated view for considering a range of metadata information. For instance, if we were considering the purchase of a digital camera, we must go to a variety of sites to do product comparisons among brands, then go to a number of sites to look for reviews, then a number of sites for pricing. The collective experience of these comparisons is certainly assistive but it is not efficient; similarly, the range of questions we can ask are constrained to those informing a purchase, rather than perhaps supporting an understanding of optics, which may also be of associated interest.

Interestingly, the Semantic Web's foundational use of ontologies makes this kind of high level representation of associated contextual information across domains and across heterogeneous sources, possible. Indeed, in the new eScience work, enabled by the Semantic Web, tracking just such associations, the *provenance* of the data, is essential for determining the source and possible reliability of the data. Much of the Semantic Web research in this space has been involved with the protocols for expressing ontologies, encoding metadata and defining services. By reaching out to the Web as the data source, the Semantic Web effort is reaching well beyond the constraints of local link base sources. In a sense, then this World Wide access goes beyond what the hypertext community has attempted, or thought of as tractable with pre-Web hypertext systems. The Semantic Web promises a huge, large scale opportunity to explore hypertext concepts that have only been explored at the theoretical or local level. This is a great opportunity: we as hypertext researchers already have understandings of how metadata can be used for exploring information spaces. It is just at this application modelling level where we have the opportunity to have a significant impact on the evolution of the Semantic Web, to take it beyond the *clink* of Web-based hypertext. In many respects it is an ideal platform to enable and explore the conceptual value of associative hypertext to the community.

5.1 mSpace

We have suggested that email facilitates both more of hypertext's human-centered association building and thus potentially more of bloom's taxonomy for building knowledge than does the Web. We suggest that this is so because email goes beyond the linked *clink* of the Web user interface. In this section, we describe a hypertext interaction model, mSpace, designed to support hypertexts that are more email-like in association building, than Web-like clinking alone.

mSpace is an interaction model for hypertext designed to facilitate user-determined exploration of a domain [schraefel03]. By exploration, we mean something other than the Web sense of browsing or surfing via *clinks*. Surfing or browsing Web pages suggests moving among discrete Web pages that potentially have only the very loose associations between them as described above. By exploration, we mean that the user is making information selections within a structured information domain, based on known or inferred associations among its parts. We can build this kind of interaction on a local database (indeed, that's what we did for our first prototype [schraefel03]), but with the Semantic Web's use of semantics and ontologies, we have the potential to establish these kinds of associations across globally

distributed and heterogeneous sources. The interaction design for exploration provides mechanisms for the user to take advantage of these associations in exploring the domain. In exploration, context and the availability of contextual information to support users' exploration is a critical component of the interaction design. An mSpace, designed to support exploration, therefore privileges associations and contexts in the domain interaction and representation. It also privileges user-determined orientation in a domain. For instance, an mSpace about classical music may have a default hierarchical representation of the space which shows Period | Composer | Genre | Instrument | Piece. Selecting an attribute such as Romantic from the Period area populates the next category, Composer, with the composers whose metadata maps them to the Romantic period. Selecting a composer, such as Beethoven, then populates the next category and so on. Composer and Genre may not mean much to some users, but perhaps Instrument does. An mSpace would allow users to reorganize the categories to Instrument | Composer | Genre | Piece. This action has supported both the rearrangement of the dimensions and the contraction of the space by removing Period from consideration. Other manipulations supported are swapping and Expansion of categories. For instance, Period might have been swapped instead for Recording so that the selection of a piece might produce a list of possible recordings. This might cue up an interest in the recording artists involved and so adding a Performers category could list the people involved in a particular recording.

Beyond the arrangement of a domain, mSpaces associate domain information with selected instances. We call this First Level Zoom. Each selected attribute calls up associated information so that the selection of a given Period provides information about that period: what are the date Ranges; what are the significant features of the period and representative works/composers? The selection of a composer provides an overview of the artist as well as a dynamically generated list of available associated topics. One might wish to explore the historical period of Beethoven or his political relationships or significant themes in his music. Selecting one of these paths in the First level zone or Overview lets a user focus in on that information. In each case, the context of the associated information is maintained.

In brief, the model lets users arrange an n-dimensional information space such that they can determine both a slice through the space, and then the scope, orientation, and arrangement of the attributes in that slice. A slice is determined first by the selection of categories (in the Semantic Web these are translated as class expressions within an ontology). This selection acts like a projection through an n-dimensional space, which is then flattened. The result is a hierarchical representation of the dependencies of attributes in that hierarchy, based on the ordering of the selections. The first attribute in the order represents a query for all the instances matching that attribute/expression. Selection of instances within that listing then acts as a constraint in populating the instances of the next attribute in the hierarchy and so on. The model therefore supports two levels of user interaction: manipulation of the ontology representation itself and selection of the instances of the data associated with those configurations. The logic of the model also provides for automatic reasoning across the domain to ensure that only meaningful attribute orderings/ selections can occur. Which of these affordances of the interaction model a system designer wishes to implement are up to the designer. The formal mSpace model is more fully defined in [Gibbins03].

The goal of an mSpace is to support user-determined access to a domain. By letting people orient the domain in ways that support their interests and their knowledge of a domain, we provide a way for users to manipulate information resources for moving up Bloom's taxonomy of learning. We go beyond the clink of Web hypertext to facilitate a what-if type exploration – to see immediately how one part of a domain is associated with another; to explore relations among contexts; to build expertise about the domain itself through such exploration. By way of demonstration, in the following section, we briefly

describe CS AKTive Space, our first implementation of an mSpace as a Semantic Web application. The example demonstrates how, even within a Web-delivered application, we can go beyond the clink of the Web to more enriched hypertext. Such enriched, post-clink interactions enabled by Semantic Web technologies, support the kind of hypertext association building we've only seen at this scale in email.

5.1.1 CS AKTive Space



Figure 8. CS AKTive Space

CS AKTive Space [Shadbolt03], which recently won Best Semantic Web Application [Visser03], is a version of an mSpace designed to represent the Computer Science domain in the UK. Here, the main slice through the domain is a hierarchy of Research Area, Region of the Country and Researcher. In this case, what the anticipated users know about the domain may be their own area. They can select that area and then select on a map, what area of the country is of interest to them to see who might be working in their research area. This provides them with a list of groups and their contact information working in the area. If they select the person (the view shown above), they see contact details about the person, the list of publications, and the Community of Practice, the list of people that this researcher collaborates with, with links to those other researchers. By lightweight interactions on the user's part, like brushing or clicking, users can explore an array of information within one window. At any point they may wish to

look at an element in detail, they can do that, to see the information sources going into the representation of that part of the representation.

This emphasis on user-determined, domain exploration offers as well a complement to adaptive hypermedia. Adaptive hypermedia attempts to anticipate the most appropriate path through or versions of a set of instances/Web pages that various users may encounter, depending either on their previous interactions with the system or the system's understanding of their experience. mSpace takes the approach that if given an appropriate representation of that domain space, users can drive themselves through the space. This is not to say that adaptive methods are not appropriate: a person who goes from composer to composer, and only collects piano pieces, might be assisted if the system said it could arrange the space to put instrument first, to facilitate the collection of piano pieces. What the model does suggest is that future hypertext can better facilitate users exploring a domain by taking the ontological power of the a domain representation, and letting the user ride that in order to create their own path through the information, to create their own associations for their own knowledge building. mSpace proposes another interaction model for hypertext: domain exploration. It will be increasingly possible in the future to support this interaction model *because* of the Semantic Web. It is an interaction model that suggests an alternate to keyword search as User Interface, and a list of links as the way to look at information as discrete entities rather than as part of a context.

There are two key advantages to this model for application developers. First, mSpace provides system designers with a way to support fast visual data inspection in the domain from numerous perspectives: views quickly let a designer see where there are holes in the available domain data. Second, the model also gives designers an automatic way to leverage ontologies to present the domain to a user such that users can readily perceive the scope and relations within the domain from the available attributes, and then can explore the domain from an orientation of the information that suits their interests.

In the CS AKTive case, as well, we are now integrating Semantic Web services, like chat components [Eisenstadt02], and task/to-do list components [Tate02] that will facilitate integrating one person's discovery in a space with the sharing of this information in other potentially more writerly or collaborative contexts. Something discovered in the manipulation of the domain can quickly and automatically be populated into an actionable item to be shared with a collaborative group. The process of discovery can be played back with these services in a way reminiscent of quote commentary; alternative paths can be explored by different participants. In other words, while the mSpace model starts with an exploration of associated contexts for knowledge access and knowledge building, the Semantic Web as a platform provides for opportunities to expand these discoveries in new, associated ways. It is worth noting that one of our motivations in developing the mSpace model was to help domain-naïve users gain access to domains where they may want information, but where they do not have the lexical expertise to formulate the query. The kind of strongly contextually associated, hypertextual domain representation afforded by mSpaces makes domain access via exploration and user-determined orientation far more possible. The Semantic Web makes such hypertexts feasible.

5.2 WiCK

As mSpace deals with exploring and discovering new information, WiCK (Writing in the Context of Knowledge) deals with the problem of synthesising and creating new documents. At the moment, computers support *word* processing, assembling words into appropriate graphical presentations according to stylistic rules concerning fonts and layout. Although these presentations may even be destined for the Web, the effort to write in a Web context, *ie* to link relevant texts together requires significant manual interventions. In the future, we can look forward to software that will support

meaning processing, assembling pages into appropriate logical presentations according to rules of reasoning and argument. This is hypertext as discursive, intertextual knowledge building as elaborated in the opening section.

When creating a new web page it is worth bearing in mind that there are something like a billion other public web pages forming a significant context for the new material. It is quite likely that there already exists a large body of similar or pertinent information online that could be revealed by a Web search engine. The difficulties and expense of searching for links (information overload, disambiguation of similarly-named entities, multiple textual expressions of the same idea) means that a new page is seldom threaded into this context.

However, in a future containing the Semantic Web, a new form of context will exist. Alongside the prose output of millions of authors, there will be a database of formally expressed and computationally useful knowledge. This knowledge will be harvested from the metadata statements published with the prose: *bibliographic information* relating to the publication and rights management of each Web page, *domain information* relating to the subjects, activities, people and organisations under discussion by the page, *argumentation information* expressing how the statements in the page support or contradict ideas in other pages.

The objective of the Semantic Web is to bypass the complexity of natural language understanding that makes web pages so difficult to use for any purpose other than reading. Commercial applications have already exploited this for easing business transactions, scientific and engineering processing have similarly improved data exchange, but the production of documents can be enhanced by giving authors access to a large pool of source information.

Web pages are harvested by search engines that decompose them into their constituent words against which user searches are compared. Metadata which is similarly harvested can be decomposed into component facts (or ‘triples’ in current semantic web parlance) and by correlation against its various structured and overlapping ontologies, be used to build up an understanding of a field. In an academic sphere for example, project metadata from a funding organisation can be integrated with publication metadata from a digital library and career information from a professional society to produce a useful picture of the significant members of a particular research community [Shadbolt03b]. This picture can be provided simply by listing all the known facts, or by applying visualisations, summaries or some kind of reasoned analysis.

There are many routine background-reading activities that an author would have to undertake to take advantage of the ‘readable Web’ context (searching, identifying relevant documents, extracting pertinent information, collating multiple sources); in the semantic web much of this has been taken care of by machine-assisted metadata harvesting and analysis. The semantic web therefore could provide a very different and more useful context for creating new material – a context of directly useful knowledge, rather than readable documents. To take another academic example, when writing the Background section a project funding proposal, typing the name of one of the investigators could bring up a list of their recent funded projects (ordered by monetary value) for direct inclusion in the text, together with a set of hypertext links to the original Web page which listed this information. For this information to be suggested a formal description of a funding proposal would need to be created – specifying that a ‘Background’ section refers to people and their significant research achievements. The same information could of course be manually derived from a simple Google search, noting the figures from several separate project pages. Figure 9 shows this activity being supported by the WiCK editor – writing the name of one of the project investigators in the document in the left-hand pane causes appropriate

semantic listings to be generated in the right hand column. In this case the reference number of a project to be cited in the supporting material (as evidence of previous funding gained by that investigator) has been dragged from the knowledge pane into the document. The outcome is that the proposal document is intricately and explicitly linked into the knowledge base and into the documents which originally provided the information to which it refers.

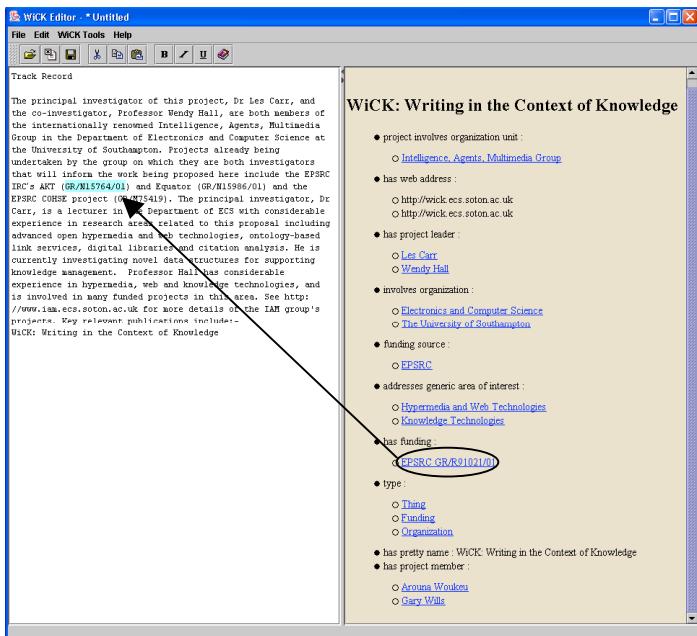


Figure 9: WiCK Editor

5.3 Escience

Scientific research is looking towards a future which involves a massive increase in scale on several fronts: vast amounts of data generated by new experimental techniques, vast amounts of computation to process it, very large scale collaborations between ‘virtual organisations’ of scientists, and with this an acceleration of the scientific and scholarly process enabled by working increasingly in the digital domain. In this section we describe the nature of the emerging ‘eScience hypertext’ and again draw upon the notions of email as hypertext and Semantic Web as an enabling technology. We choose eScience as a case study but anticipate similar trends in the e-research environment across a very broad set of disciplines.

The UK e-Science programme defines e-Science as being “about global collaboration in key areas of science, and the next generation of infrastructure that will enable it” [Hey02]. eScience brings a rich array of interlinked digital artefacts, including of course the huge data stores but also the description of experiments, data, analysis and results. There is a flow of data, annotation of that data, and derivation of results which then feed into subsequent work. These interlinked artefacts constitute a hypertext, and the causal flow associated with these processes brings backward-linking in a similar way to email, leading to dependency diagrams like Figure 2. Unlike the largely textual territory of email, in eScience we are working with scientific data: there is a very clear separation of data and annotation, and we are more likely to point at the data we are describing than to embed it – this is a domain in which transclusion becomes a necessity.

The traditional scientific and scholarly publishing processes are increasingly augmented with the less formal discourse supported by mailing lists, web forums and Internet chat, which provide further annotation of the artefacts; i.e. email can be regarded as part of the eScience hypertext. The result is three layers of digital information flow: the flow of data, the flow of academic papers, and the scientific

discussion associated with both. Each layer can be regarded as a hypertext, with similar characteristics due to the processes of derivation and annotation: like email, each is structured around a temporal axis, with backward links which express provenance. Each layer is deeply intertwined with the others, though currently these inter-layer links may not be represented explicitly – perhaps because, historically, the three layers were not all in the digital domain.

As the eScience infrastructure becomes more widely deployed, we envisage the merger of the three hypertexts – data, scholarly communication, informal discourse – into one eScience hypertext. The vision is that a scientist looking at a ‘paper’ can navigate a rich hypertext which will provide access to data, experimental procedure and discussion – though they need not do this explicitly but rather it will be done by their software. This is the notion of “publication at source”, giving access to original data and making explicit the provenance of derived results, which is very important in the scientific context. Furthermore the e-scientists can build on this material.

To achieve this requires explicit machine representation of the relationships between entities, both within and between the layers. This rich and open hypertext is an exemplary application of Semantic Web technologies, representing the relationships as the predicates of triples that describe the links. This approach is being adopted in several eScience projects such as [Frey02]. While current Semantic Web deployments typically work with persistent stores with a low rate of change, we envisage rapidly changing relationships in future systems - the processes of eScience generate data in real-time, to be processed in a timely fashion. Indeed, these dynamic, notification-oriented aspects of the eScience content again mirror a characteristic of email, i.e. live flow of messages with notification.

With increasing scale of data processing, and increasing scale of collaboration, the three hypertexts increase in scale too. When increasing the number of users of a product increases the value of the product to the individual user, we have a *network effect* [Whitehead99]; e.g. an email facility increases in value to the individual as more people use it. Similarly, the denser the linking in the hypertexts, the more valuable they may be, provided we can manage the scale. To realise these effects, our hypertext nodes must be shared widely – in Semantic Web terms, we need shared, unique identifiers for all entities, so that all the metadata that is generated which refers to those nodes is effectively linked together via them. If we can achieve this, e-Science will build huge hypertexts which gather considerable added value with increasing scale; for example, multiple independent discussions about the same chemical compound will be connected up by the reference to that compound. This added value through shared annotation will in turn facilitate the generation of new scientific results.

We can bring the more synchronous scientific discourse of meetings into the eScience hypertext too, interlinking temporal media artefacts from physical meetings as well as computer mediated events such as videoconferences. In the CoAKTinG project [Shum02], the discourse and argumentation structure of discussions is represented explicitly in a graphical notation and this graph provides a hypertext perspective which is directly linked to the appropriate artefacts. It also supports instant messaging and visualisation of presence to enhance group peripheral awareness, planning and task support, and meeting capture and replay. These tools interwork through the use of ontologies.

When discussing email as hypertext, we noted that email allows the user to work in a natural fashion. The natural fashion of work for the scientist is their interaction with the laboratory and their laboratory notebook, and this is therefore part of the interface to our eScience hypertext. Indeed, the promise of eScience hypertext is the promise of hypertext becoming an everyday part of the actual, in-lab experience. Thus we are exploring models of use on the human side of the Semantic Web. The user interface which supports this ongoing process of annotation and derivation might not look like a Web

browser, nor email, but rather it may be the devices in the chemist's laboratory [schraefel04]. Annotation may occur through a human process but there are two other important sources of metadata: the metadata captured automatically at the time of data capture, relating to the circumstances of that capture, and annotation by recording automatically the usage of data and information; e.g. who used it and for what. It is important to note that this automatically captured information is brought in and out of context at the behest of the scientists— not of the system.

Hence the eScience hypertext pushes us up Bloom's taxonomy. All the recorded information, such as the ambient temperature of the lab, the people in the lab at the time of a given experiment, the images of a filtration process, are all available resources for the scientist to pull in on demand as appropriate for analysis or for support of claims. This takes us back to the WiCK model of making appropriate resources available in the context of the task performed. Similarly, this association on demands affords an on-the-spot investigation of provenance: like chasing back to the original email in an exchange of information, hyperlinked information resources in eScience means that data trust can be made available with no effort by the scientist and scrutinized just as readily by the readers in the community. It is possible to imagine in this ready exchange of resources a new kind of scholarly publication where WiCK'd annotations on existing work become themselves a founding reference for a new document, a new analysis.

6. CONCLUSIONS

"Next, let's ask what is different about hypertext? Is there really anything new to it? ... hypertext is basically clay and we have to mold it; that is what this workshop is all about: starting to mold that clay

Quote from Andries van Dam's keynote speech at the initial hypertext workshop in 1987

In this paper, we have looked at what future hypertext systems may be, by looking at what our understandings of hypertext have been, and how these definitions have gone beyond the limited click and link of the Web. We have also shown, however, that we encounter and create rich hypertexts daily not just in the Web, but in email as well, an application we use perhaps as much if not more than the Web. By reflecting on email-as-hypertext, we have also shown that the cost for implementing hypertext properties like transclusion and annotation can be remarkably low. One of our main points in establishing this relation between the Web and email as complementary hypertext models is to stake out a research direction for future everyday hypertext systems based on the Semantic Web platform. We have shown that the Semantic Web is a powerful platform for enabling globally the kinds of context support, association building and dynamic linking embodied locally in pre- and co-Web hypertext systems like Intermedia and Microcosm. The Semantic Web offers an opportunity to return to the research agenda initiated in these systems, but at a Web scale. We have also been keen to show however, that Web scale should not mean, constrained to the Web interaction paradigm of clinks. We have therefore also pointed to more current hypertext/Semantic Web research in information exploration, writing, and eScience where the interaction breaks away from the Web clink and is more informed by the reflective interchange of email-as-hypertext for knowledge building. We have proposed Bloom's taxonomy as a criteria for evaluation of the success of such applications: that we should challenge our future everyday hypertexts by the question do they allow us to move from simple regurgitation of facts to support synthesis and analysis of that information? If they do, we have suggested, then you've got hypertext.

7. ACKNOWLEDGMENTS

The projects mentioned in section 5 have been supported by the following grants:

CombeChem: Structure-Property Mapping: Combinatorial Chemistry & The Grid (Combechem) UK Engineering and Physical Sciences Research Council (EPSRC) GR/R67729/01

CoAKTinG: Collaborative Advanced Knowledge Technologies in the Grid. EPSRC GR/R85143/01

AKT: Interdisciplinary Research Collaboration in Advanced Knowledge Technologies (AKT), EPSRC GR/N15764/01.

WiCK: Part of the Knowledge Writing in Context (KWIC), EPSRC GR/R91021/01.

We also wish to thank both the editor of this issue for her patience and support, and the many readers whom we have asked for comments on this paper. The work is better for your thoughts.

8. REFERENCES

- [AKT01] AKT Advanced Knowledge Technologies, 2001: <http://www.aktors.org/>
- [Bailey01] Bailey, C., Samhaa R. El-Beltagy and Wendy Hall Link Augmentation: A Context-Based Approach to Support Adaptive Hypermedia. In *Proceedings of Third Workshop on Adaptive hypertext and Hypermedia*. (2001): 55-63.
- [Berners-Lee01] Berners-Lee, T, Hendler, J., Lasilia, O. The Semantic Web, Scientific American, May, 2001. <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>.
- [Berners-Lee03] Berners-Lee, T. Semantic Web: Where to Direct Our Energy. Invited Talk, 2nd International Semantic Web Conference, 2003. <http://www.w3.org/2003/Talks/1023-iswc-tbl/>
- [Bernstein91] Bernstein, M. Storyspace: hypertext and the process of writing. Hypertext/Hypermedia handbook, McGraw-Hill, Hightstown, NJ., 1991.
- [Bernstein03] Bernstein, M. Collage, composites, construction. Proc. of Hypertext 2003: 122-123.
- [BuckinghamShum02] Buckingham Shum, S., De Roure, D., Eisenstadt, M., Shadbolt, N. and Tate, A. CoAKTinG: Collaborative Advanced Knowledge Technologies in the Grid. 2nd Workshop on Advanced Collaborative Environments, 11th IEEE Int. Symposium on High Performance Distributed Computing (HPDC-11), July 24-26, 2002, Edinburgh, Scotland.
- [De Bra99] De Bra, P, Brusilovsky, P. Houben G-P. Adaptive Hypermedia: from System to Framework. ACM Computing Surveys 31.4 (1999): 12pp.
- [Bloom65] Bloom. B.S. A Taxonomy of Educational Objectives Handbook 1: Cognitive Domain New York: McKay 2nd Edition
- [Brown87] Brown, P. J. Turning Ideas into Products: The Guide System, HT87 (November 1987): 33-40.
- [Bush45] Bush, V. (1945) As We May Think, *Atlantic Monthly*, (July 1945): 101-108,
- [Caroll03] Caroll, Sean. Site of the Week: Vivisimo. PC Magazine (September 19, 2003): <http://www.pcmag.com/article2/0,4149,1274924,00.asp>.
- [Carr95] Carr, L., De Roure, D., Hall, W. and Hill, G. The Distributed Link Service: A Tool for Publishers, Authors and Readers. World Wide Web Journal 1.1(1995): 647-656.
- [Carr01] Carr, L., S. Bechhofer, C. A. Goble, , and W. Hall. Conceptual Linking: Ontology-based Open Hypermedia. *WWW10*, May 2001.
- [Carr99] Carr, L., Hall, W. and De Roure, D. Link Service Evolution. ACM Computing Surveys. Symposium on Hypertext and Hypermedia. 31.4es (December 1999): Article 9.
- [Conklin87] Conklin, EJ. Hypertext: An introduction and survey. IEEE Computer, 20.9 (1987): 17-41.
- [Davis92] Davis, H., Wendy Hall, Ian Heath, Gary Hill, Robert Wilkins. Towards an Integrated Information Environment with Open Hypermedia Systems. In ECHT '92: ACM HT 1992: 181-190
- [De Roure03] De Roure, D., Jennings, N., Shadbolt, N. Research Agenda for the Semantic Grid: A Future eScience Infrastructure, in Grid Computing: Making the global infrastructure a reality, Berman, F., Fox, G., and Hey, T. (eds), Wiley Europe, 2003: 437-470.
- [Dean03] Dean, M., Dan Connolly, Frank van Harmelen, James Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein eds. (2003) OWL Web Ontology Language 1.0 Reference. W3C Working Draft <http://www.w3.org/TR/owl-ref/>
- [Domingue03] Domingue, J., Dzbor, M., and Motta, E., Semantic Layering with Magpie. Handbook on Ontologies, Staab, S. and Studer, R. (eds.), Springer Series on Handbooks in Information Systems XVI, Springer Verlag, 2003.
- [Eisenstadt02] Eisenstadt, M. and Dzbor, M. "BuddySpace: Enhanced Presence Management for Collaborative Learning, Working, Gaming and Beyond", JabberConf Europe, Munich Germany, June 2002. <http://www.aktors.org/technologies/buddyspace/>.

[Engelbart68] Engelbart, D., and W. English, "A Research Centre for Augmenting Human Intellect", Fall Joint Computer Conference 1968, pp. 395-410.

[Frey02] Frey, J., De Roure D. and Carr L. Publishing at Source: Scientific Publication from a Web to a Data Grid, EuroWeb 2002 Conference, Oxford, Dec 2002. <http://www.ecs.soton.ac.uk/~lac/publicationAtSource>.

[Frey03] Frey, Jeremy G., David De Roure, m.c. schraefel, Hugo Mills, Hongchen Fu, Sam Peppe, Gareth Hughes, Graham Smith, Terry R. Payne. Context Slicing the Chemical Aether. First Workshop on Hypermedia and the Semantic Web, in conjuncntions with HT 2003.

[Genette97] Genette, G. *Palimpsests* (trans. Channa Newman & Claude Doubinsky). Lincoln, NB: University of Nebraska Press, 1997.

[Gibbins03] Gibbins, N., Harris, S. schraefel, m.c., Applying mSpace Interfaces to the Semantic Web, 2003. Preprint <http://eprints.ecs.soton.ac.uk/archive/00008639/>.

[Goodman87] Goodman, D. The Complete HyperCard Handbook. US: Bantam Books, 1987. Excerpt interview with Atkinson: <http://www.globalideasbank.org/BOV/BV-462.HTML>.

[Gronbaek00] Gronbaek, K., Lennert Sloth, and Niels Olof Bouvin. Open Hypermedia as User Controlled Meta Data for the Web. WWW, (2000): 553--566,

[Handschrift02] Handschuh, S., Staab, S. *Authoring and annotation of web pages in CREAM*, WWW02, New York, NY, ACM Press, 2002: 462 – 473.

[Hann92] Haan, B. J., Paul Kahn , Victor A. Riley , James H. Coombs , Norman K. Meyrowitz (1992) IRIS Hpermedia Services. C ACM, 35.1(1992): 36-51.

[Halasz87] Halasz, F., Moran, T. P. and Trigg R. H. Notecards in a Nutshell, In *CHI87*, (1987): 45-52

[Hendler02] Hendler, J. Introducing the Semantic Web. <http://www.cs.umd.edu/~hendler/presentations/SWEO-talk.ppt>

[Harnad02] Harnad, S. Hypermail exchange regarding Jim Hendler's talk at Southampton, Introducing the Semantic Web. <http://www.ecs.soton.ac.uk/~harnad/Hypermail/Syntactic.Web/>.

[Harnad03] Harnad, S. Back to the Oral Tradition Through Skywriting at the Speed of Thought. Online conference "The Future of Web Publishing: Hyper-Reading, CyberTexts and Meta-Publishing". Centre Jacques Cartier. <http://www.interdisciplines.org/defispublishingweb/papers/6>.

[Herzog95] Herzog, M. The Dexter Hypertext Reference Model, 1995: <http://www.dbai.tuwien.ac.at/staff/herzog/thesis/node25.html>.

[Hey02] Hey T. and A. E. Trefethen, The UK e-Science Core Programme and the Grid, Future Generation Computing Systems. 18(2002): 1017.

[Hill92] Hill, G. J., Wilkins, R. J. and Hall, W. *Open and Reconfigurable Hypermedia Systems: A Filter-Based Model*. Technical Report CSTR 92-12, Electronics and Computer Science, University of Southampton, UK.

[Kampa02] Kampa, S. Who are the experts? e-Scholars in the Semantic Web (PhD Thesis) University of Southampton <http://eprints.ecs.soton.ac.uk/archive/00007222/>.

[Kristeva86] Kristeva, J. Word, dialogue, and the novel. In T. Moi (Ed.), *The Kristeva Reader*. New York: Columbia University Press. 1986: 35-61

[Landow92] Landow, G.P. Hypertext and Intertextuality. Hypertext: the Convergence of Contemporary Critical Theory & Technology, John Hopkins University, 1992: 10-11. <http://www.cyberartsweb.org/cpace/ht/jhup/intertext.html>

[Lassila99] Lassila, O. and Swick, R. (eds) Resource Description Framework (RDF): Model and Syntax Specification. W3C Proposed Recommendation (January 1999). <http://www.w3.org/TR/PR-rdf-syntax/>.

[Lewis96] Lewis, P.H., Davis, H. C., Dobie, M. R. Hall, W. Towards multimedia thesaurus support for media-based navigation. First International Workshop on Image Databases and Multi-Media Search, Amsterdam University Press, August 1996: 83-90.

[Marshall91] Marshall, C.C., Halasz, F., Rogers, R., and Janssen, W. 1991. Aquanet: a hypertext tool to hold your knowledge in place. In *ACM HT91* (1991): 261-275.

[Marshall94] Marshall, C.C., Shipman, F., and Coombs, J. 1994. VIKI: Spatial hypertext Supporting Emergent Structure. *ECHT '94* 1994: 13-23.

[Marshall03] Marshall, C.C., Shipman, F.M.. Which Semantic Web? ACM HT 2003: 57-66.

[McCracken84] McCracken D., Akscyn, R. Experiences with the ZOG HCI System, International Journey of Man-Machine Studies, 121(1984): 293-310.

[Meyrowitz86] Meyrowitz, N. "Intermedia: The Architecture and Construction of an Object-Oriented Hypermedia System and Applications Framework", ACMSIGPLAN OOPSLA '86, 21.11(Nov 1986): 186-201.

[Meyrowitz89] Meyrowitz N. The Missing Link: Why We're All Doing hypertext Wrong. E. Barrett, editor, *The society of text: Hypertext, hypermedia and the social construction of information*. MIT Press, Cambridge, USA, 1989: 107--114.

[Minsky91] Minsky, M. Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy", AI Magazine 12.21991: 34-51.

[Nelson81] Nelson, T. Literary Machines. Self-published. Swarthmore, PA, 1981.

[Nelson87] Nelson, T. Computer Lib, 2nd Edition, Microsoft Press, 1987.

[Nelson99] Nelson, T. H. "A Literary Structure with Two Fundamentally Different Means of Connection." Xanalogical Structure, Needed Now More than Ever: Parallel Documents, Deep Links to Content, Deep Versioning and Deep Re-Use." ACM Computing Surveys 31.4es (1999).

[Price98] Price, M., Golovchinsky, G., and Schilit, B. "Linking By Inking: Trailblazing in a Paper-like hypertext," ACM HT98, 1998:30-39.

[schraefel02] schraefel, m.c., Zhu, Y., Modjeska, D., Wigdor, D. and Zhao, S., Hunter Gatherer: Interaction Support for the Creation and Management of within-Web-Page Collections. In Proceedings of the eleventh international conference on World Wide Web, (Hawaii, 2002): 172-181.

[schraefel03] schraefel, m.c. , Maria Karam, Shengdong Zhao. mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. Proceedings of the Workshop on Adaptive Hypermedia, 2003, in conjunctions with HT 2003.

[schraefel04] schraefel, m.c., Hughes, G., Mills, H. Smith, G., Payne T., Frey, J. (2004) Breaking the Book: Translating the Chemistry Lab Book into a Pervasive Computing Lab Environment. CHI 04. Preprint <http://eprints.ecs.soton.ac.uk/archive/00008641/>.

[Shadbolt03] Shadbolt, N. R., schraefel, m. m. c., Gibbins, N., Glaser, H. and Harris, S. CS AKTive Space: Representing Computer Science in the Semantic Web. Preprint <http://eprints.ecs.soton.ac.uk/archive/00008638/>.

[Shadbolt03b] Shadbolt, N. R., schraefel, m. c., Gibbins, N., Glaser, H. and Harris, S. (2003b) CS AKTive Space: A demonstration of an Integrated Semantic Web Application. Journal of Web Semantics 1.2(2003) Preprint <http://eprints.ecs.soton.ac.uk/archive/00008652/>.

[Shipman01] Shipman, F.M. ,Hsieh, H., Maloor , P., Moore, M. (2001) The visual knowledge builder: a second generation spatial hypertext. HT01 2001: 113—122.

[Shneiderman91] Shneiderman, B., Plaisant, C., Botafogo, R., Hopkins, D., and Weiland, W., Designing to facilitate browsing: A look back at the Hyperties workstation browser, Hypermedia 3.2(1991): 101-117.

[Smith88] Smith, J., Weiss, S. Special Issue on Hypertext. CACM 31.7 (November 1988).

[Staab00] Staab, S. Angele, J., Decker, S., Erdmann, M., Hotho, A., Maedche, A., Schnurr, H.-P Studer, R. Sure, Y. *Semantic community web portals*. Computer Networks, Special Issue: WWW9 – Proc. of the 9th International World Wide Web Conference, Amsterdam, The Netherlands, May, 15-19, 2000.

[Tate02] Tate, A., Dalton, J. and Stader, J. "I-P2 - Intelligent Process Panels to Support Coalition Operations", Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations (KSCO-2002), Toulouse, France, 23-24 April 2002. <http://i-x.info/documents/2002/2002-ksco-ip2.pdf>

[Uren03] Uren, V., Buckingham Shum, S. Li, G., Domingue, J., Motta, E. "Scholarly Publishing and Argument in Hyperspace", WWW03, 2003: 244 – 250.

[Viser03] Viser, U. The Semantic Web Challenge. 2003: <http://challenge.semanticweb.org/>.

[Whitehead99] Whitehead, E. James, Jr. Control Choices and Network Effects in Hypertext Systems. HT99, (1999): 75-82.

[Yankelovich85] Yankelovich, N., Meyrowitz N. and van Dam, A. (1985) Reading and Writing the Electronic Book, IEEE Computer 18.10 (1985): 18-30.

[Yankelovich88] Yankelovich, N., Haan, B.J., Meyrowitz, N.K., Drucker S.M. Intermedia: The Concept and the Construction of a Seamless Information Environment. *IEEE Computer* (1988): 81-83, 90-96

[Yellowless94] Yellowlees Douglas. J.. "How Do I Stop This Thing?": Closure and Indeterminacy in Interactive Narratives." Hyper/Text/Theory. Ed. George P. Landow. Baltimore: Johns Hopkins UP, 1994: 159-188.