Choosing our Science

David E. Millard (University of Southampton) and Weigang Wang (University of Manchester)

Hypertext and the Web

What is Hypertext? It is well known in our community that the word Hypertext was coined by Nelson in 1965 to describe his vision of an intertwingled world of transcluded electronic texts, but we also know that the ideas and principles of Hypertext predate electronic computers. Writers and scholars have always experimented with interlinked texts, from the Talmud and the Synoptic Gospels to Borges' Garden of Forking Paths. Well's World Brain, Otlet's Mundaneum and Bush's memex set out grand visions for global stores of human knowledge based on index cards and microfiche.

Despite these broad origins Hypertext has come to be seen almost exclusively from a digital viewpoint, perhaps because of its many synergies with key concepts from informatics and computer science, such as networks, communication theory, and knowledge modeling (Conklin, 1987). But Hypertext is more than digital. It predates computers and predates computer science. Any broad approach to hypertext must therefore be interdisciplinary (Tosca, 2001), even if keeping one foot in the digital domain.

The Web is the closest that we have come to the grand visions of the hypertext pioneers. From unpromising beginnings as a basic read-only distributed hypertext, the Web has evolved over the last twenty years into the premier distributed application platform. As a platform it supports a whole ecology of hypertext tools and forms: versioned collaborative hypertexts (Wikis), bookmarks and trails (Delicious or Digg), citizen journalism (blogs and sharing sites) and social conversation and chatter (Facebook or Twitter). As a hypertext it disappointed and confounded us, but as a platform it has excited and renewed hypertext research.

A Science of the Web

Web Science is a new discipline that is concerned with the study of the Web and our behavior on it. Web Science draws on a wide range of traditional disciplines (such as sociology, economics and law) to understand, model and predict the Web's impact on our lives and societies. Through Web Science the hope is that we can better understand the affects of different technology, and the changing attitudes and social norms that emerge from its use. The aim is to better guide new developments in both technology and policy.

Web Science was first proposed by Berners-Lee et al (2006) in an article for Science, this then led to the establishment of the Web Science Research Initiative (now the Web Science Trust[1]) an agreement between MIT in the US, and the University of Southampton in the UK, to explore and promote the study of the Web; in particular by developing curricula for Web Science and through

[1] The Web Science Trust: http://webscience.org/

establishing the Web Science conference (held for the second time this year in Raleigh, North Carolina).

As the Web has become an essential service in our society (Hendler et al, 2008), so hypertext has become an essential tool of communication and interaction. If we accept that the Web is an ecology of hypertexts and that Web Science is the interdisciplinary study of the Web then we should expect that Web Science and Hypertext are natural bedfellows. It was this thought that led us (the guest editors) to ask what is or should be the relationship between Hypertext research and Web Science.

In 2008 we ran a workshop at ACM Hypertext 2008 on 'Web Science: Collaboration and Collective Intelligence'[2]. In part we were interested in what aspects of its Hypertext work that the community also considered as Web Science. The workshop was the largest at the conference, and ten position papers were presented on the day on topics ranging from trust and media bias to narrative structure.

There was a great deal of interest from participants about the idea of Web Science, but there was also uncertainty. Computer Science is itself a hybrid discipline – engineering, mathematics, logic, human factors, semiotics and semantics, etc (Shneiderman, 2007). Where were the boundaries of this new discipline of Web Science, were new tools based on interdisciplinary theories part of it, and at what point was it appropriate to call ourselves Web Scientists?

The Special Issue

This special issue of the New Review was aimed at researchers who believe that their work is in the intersection of Hypertext and Web Science. We invited papers on a variety of technical topics with a personal, cultural or societal slant and particularly welcomed work that we felt was interdisciplinary in nature.

The accepted papers have provided a good sample of such topics, which include:

- Knowledge structuring and creation
- Computational support on knowledge and metadata
- Innovative social and knowledge interfaces
- Novel community interaction and knowledge elicitation
- Studies of on-line communities and their behaviour

A key focus for Web Science in-the-small is to explore the relationship between people and knowledge, in particular the human-knowledge interface, and people's ability to create and manipulate sophisticated knowledge structures either individually or as a community. Several of the papers in the Special Issue explore these issues.

Many current approaches require authors of metadata to master Semantic Web technologies and to explicitly deal with classes, statements, URIs, namespaces,

---

[2] Web Science Workshop at ACM Hypertext 2008: http://users.ecs.soton.ac.uk/dem/workshops/webscience08/

etc. Such expertise is beyond the capability of the majority of web users (Volkel et al. 2006). In *Ontology-driven generation of wiki content and interfaces*, Iorio et al. report their efforts to design and implement a flexible and user-friendly metadata editor for making the creation of semantic wiki metadata easier, so as to involve wider user participation in the metadata creation process. The flexibility is supported by its ability to allow users to chose and apply any metadata schemes (i.e. any domain ontology that is suitable for the purpose of a Wiki site) to generate the (topic) structure of wiki pages and its corresponding metadata editor. The user friendliness is addressed by presenting the metadata editor to users as forms in wiki pages. The forms are dynamically built by analysing the class properties of the imported ontology and by mapping each property in the proper element of the GUI interface (through a GUI ontology). The automatically-generated forms (often with default values or suggestions) are finally exploited by the wiki users to actually write the semantic data.

Compared with metadata creation governed by formal metadata schemes, tagging and simple folksonomies are less formal and easier for users to understand (Marlow, 2006). However, in terms of expressiveness, they are much less powerful than knowledge structures based on the Web Ontology Language (OWL). *Leveraging Search and Content Exploration by Exploiting Context in Folksonomy Systems* by Abel et al. addresses these weaknesses by adding context information to the tags. They use a lightweight approach to model context by means of tag clouds (list of weighted tags), which can be generated automatically. They propose a new model for embedding arbitrary context information into folksonomies. To make use of such metadata, they have also developed new ranking algorithms that can perform significantly better than existing ranking algorithms on such metadata structures. In this way, they have not only created tag-based knowledge structures, but also developed computation support upon them for improving searching and content exploitation.

Another article in this special issue presents a different approach to impose a knowledge structure upon user created tags. Hargood et al. in *Capturing the Semiotic Relationship between Term*s, propose the use of a thematic model (Tomashevsky , 1965) for term expansion based on semiotic relationships between terms. They explore a method to capture semiotic understanding of particular terms using a guide to authoring definitions in the terms of the model manually. Experimentation shows that it is possible to capture valid definitions that can be used for term expansion but that the guide itself may not be sufficient to support this on a large scale. As an ongoing work, they also propose additional approaches to address the scale issue by leveraging users, including the authors of the tags, to create instances of such knowledge structures.

Web Science in-the-large studies the use and impact of web technology on communities and society, and several of the papers in the Special Issue focus on community questions.

In *PATONGO: Patterns and Tools for Non-Profit Organisations,* Schuemmer and Haake describe a pattern-based approach for helping volunteers of Non-Profit Organisations to identify and share good practice (Edwards, 1998). They have

showed how good practice patterns and the relationship between them can be captured in a hypertext structure. They have also elaborated on the specific roles and processes required in the collaborative construction and use of such a hypertext-based pattern language representation. Their work views patterns as living structures subject to social improvement in their life cycle in an interactive Web 2.0 environment. The authors also discuss patterns in relation to Engelbart's ABC model (Engelbart, 1992), which may provide a theoretical ground to further study knowledge structure evolution in such contexts.

In *Tracking Cohesive Subgroups over Time in Inferred Social Networks*, Chin et al. present an analysis method for detecting cohesive sub-groups in social networks (Piper et al. 1983). Taking interaction data as input, the approach can find cohesive subgroups and help to better understand the evolution of the subgroups over time. As the authors point out, the kind of work as reported in this paper is a step towards more powerful community trackers, a new class of online tools that may in future index and query group cohesiveness and community based on online collaboration and communication.

In *Talking About Your Health to Strangers: Understanding the use of online social networks by patients*, Colineau et al. present a survey about the use of health related social networking sites in the era of Web 2.0. In such social networks, patients are organising themselves in groups, sharing observations and helping each other (Smith and Christakis, 2008). The aim of this work is to understand why and how people use health related sites. The authors studied these sites and identified three dimensions characterising most of them: informational/supportive; general/focused and new relationships/existing ones. The user study reveals interesting insights of such communities and their behaviour.

One of the principles of Web Science is that the technical protocols underlying a Web technology and the social protocols governing its use should co-evolve hand in hand. The success story of Wikipedia and its social practice presents us a good example of such co-evolution. Wikipedia is known as the 'free encyclopedia' that 'anyone can edit'. However, such openness and easiness has made Wikipedia the home of some bitter disputes. Such observations have motivated the emergence of a collection of social norms behind its proper use. While a number of studies have examined the policies and procedures involved, the development and use of norms in communication remains understudied. In *"Be Nice": Wikipedia Norms for Supportive Communication* Reagle discusses a wide range of norms in Wikipedia and their relation to Gibb's work "Defensive Communication" (Gibb, 2006). According to Gibb, defensive behaviour arises in the context of perceived or anticipated threat, to one's self or group, and causes a decrease in the ability to perceive accurately the motives, the values, and the emotions of the sender; while a supportive climate would help the receiver to concentrate upon structure, content, and cognitive meanings of the message. Reagle argues that Wikipeida norms as documented in its policies, guidelines, and explanatory essays can be characterized as supportive and suggests that Wikipedia is seeking to create a supportive communication environment.

Are We Web Scientists?

The papers in the Special Issue concerned with Web Science in-the-small, the personal relationships that we have with information and knowledge, are on topics that are historically familiar to Hypertext research: meta-data, knowledge interfaces and narrative. Hypertext researchers should feel very comfortable exploring these issues and perhaps the emergence of Web Science is partly a reflection of our willingness to explore new interdisciplinary areas that draw on a wider range of the humanities in order to understand the relationships between users and hypertext systems.

The huge success of the Web has meant that it is possible to do Web Science in-the-large, and we also present a number of papers that study the impact of Web systems on professional and personal communities, and analyze the social structures that result. These are new areas of possibility for Hypertext researchers and represent an opportunity for new longitudinal studies and the use of statistical research methodologies.

Hypertext research is naturally growing and evolving to reflect the pervasiveness of the Web and to take advantage of a changing interdisciplinary scientific culture. These changes take it closer to Web Science, and there are many overlapping areas where Hypertext is a valuable lens with which to examine the impact of the Web. However, Hypertext remains its own subject, with its own long intellectual and scholarly history.

In the end it is our choice as researchers as to whether we choose to name what we do as a Science of the Web, or ultimately brand ourselves as Web Scientists. In making that choice we should remember that Web Science is more than rebadging existing work. Interdisciplinary research can be difficult to communicate, and by using a strong term like Web Science we create a clear space for interdisciplinary research approaches, one that encourages us to explore in new ways and to think more holistically about our subject.

The articles in this special issue represent a timely and engaging intersection of Hypertext and Web science. They illustrate the wide range of technical topics, contexts and uses of the Web that fall under this intersection, as well as the complexities, opportunities and also sometimes problems that arise when Web technologies and practices are integrated into our daily life.

We wish to express our thanks to the Editors for proposing and supporting this special issue; to all the authors who submitted articles in response to the call; and also to the reviewers who provided thoughtful and detailed feedback on the original submissions.

References

Berners-Lee T., Hall W., Hendler J., Shadbolt N., and Weitzner D. Creating a Science of the Web. SCIENCE (2006) vol. 313 (5788) pp. 769-771

Conklin. Hypertext: an Introduction and Survey. Computer (1987) vol. 20 (9) pp. 17-41

Edwards. Organizational learning in non-governmental organizations: What have we learned?. Public Administration and Development (1998) vol. 17 pp. 235-250

Engelbart, D. C. (1992). Toward high-performance organizations: a strategic rolefor groupware. Proceedings of the GroupWare '92 Conference, San Jose, CA, Morgan Kaufmann, pp. 77-100

Gibb. J.R. Defensive Communication. Journal of Communication (1961) Volume 11 Issue 3, pp. 141 - 148

Hendler J., Shadbolt N., Hall W., Berners-Lee T., and Weitzner D. Web science: an interdisciplinary approach to understanding the web. Communications of the ACM (2008) vol. 51 (7) pp. 60-69

Marlow C., Naaman M., Boyd D., and Davis M. HT06, tagging paper, taxonomy, Flickr, academic article, to read. Proc. of ACM Hypertext 2006 (2006) (Odense, Denmark), ACM New York, NY, pp. 31-39

Piper W.E., Marrache M., Lacroix R., Richardsen A.M. and Jones B.D. Cohesion as a basic bond in groups. Human Relations (1983) vol. 36 (2) pp. 93-108

Shneiderman. Web science: a provocative invitation to computer science. Communications of the ACM (2007) vol. 50 (6) pp. 25-27

Smith and Christakis. Social networks and health. Annual Review of Sociology (2008) vol. 34 (August) pp. 405-429

Tomashevsky B.. Russian Formalist Criticism: Four Essays, chapter Thematics, pages 66–68. University of Nebraska Press, 1965.

Tosca. Hypertext Criticism: introduction to a special issue. Journal of Digital Information (2001) vol. 1 (7). Available from: http://journals.tdl.org/jodi/article/view/25/26

Völkel M., Krötzsch M., Vrandecic D., Haller H., and Studer R. Semantic Wikipedia. Proc of WWW 2006 (2006) (Edinburgh, Scotland) pp. 585-594. Available from: http://portal.acm.org/citation.cfm?id=1135863