Web Science

Kieron O'Hara

Wendy Hall

Published in *ALT-N*, 12 (May 2008)

Abstract
Our understanding of the Web has not kept pace with its development. It is engineered using formally specified languages and protocols, but has large scale effects on society. Certain human activities – including education – have been altered irretrievably. This article argues for the development of the discipline of Web Science, to understand the reciprocal relationship between the Web and society at a number of scales, from technical protocols to emergent social behaviour, to ensure that the Web's growth will continue, and will benefit society. The need for both analysis and engineering demands an inherently interdisciplinary approach. With this in mind, a new Web Science Research Initiative is briefly described.

The need for Web Science
Since its inception the World Wide Web has radically changed communication, collaboration and learning. However, our understanding of the Web, and the way it operates in its myriad contexts, has not kept pace. It is a piece of technology, engineered using formally specified languages and protocols, but viewed macroscopically, patterns emerge in the Web as a result of human interactions, which in turn are influenced by social conventions and laws. Furthermore, the existence of the Web as a space has its own effects on convention, while the emergent behaviour often prompts alterations in or additions to the underlying protocols. And certain human activities – not least learning, research and education – have been altered irretrievably by the technology.

In short, the Web influences the world, and the world influences the Web. So complex is the Web, with tens of billions of pages on the "surface Web" and hundreds of billions of documents in the "deep Web", and so interrelated with society (especially in the rich democracies) has it become, that its health is a matter of real importance. Understanding the Web in its context, and at a number of scales from the micro-level of individual protocols like HTTP or HTML, to the macro-level of emergent behaviour such as blogging, spamming or e-commerce, is essential for two related reasons. First, the continued growth and development of the Web depend on ensuring that it is a beneficial technology for society. And second, understanding the potential for social effects to follow from technological change would help support the basic social values of trustworthiness, privacy, and respect for social boundaries. It is important that the Web should grow, and that it should benefit society. The variance in scale and intellectual breadth, including the need for both analysis (study) and synthesis (engineering), demands an inherently interdisciplinary approach.

Unfortunately, such a research area does not yet exist in a formalised way. Within Computer Science, Web-related research has largely focused on information retrieval algorithms and the algorithms for the routing of information through the underlying Internet. Google's PageRank link analysis algorithm, for instance, is a brilliant piece of work – but its significance to the Web depends not only on its computational properties, but also on the context of its use. Nothing about the algorithm *per se* explains how the eigenvectors that it computes map onto the conversations that Web users have, nor about how it can be adapted constantly to avoid spoofing. Outside of computing, research grows ever more dependent on the Web, but no agenda has yet cohered for exploring its emerging trends nor are researchers fully engaged with the Web research community to focus more specifically on providing for the needs of scientists and educators (Berners-Lee et al 2006b).

What is Web Science?

So, what disciplines will be required, given the emerging trends on the Web as new media types, data sources and knowledge bases appear online, access becomes increasing mobile and ubiquitous, and guarantees of privacy and control ever more important? The need for better mathematical modelling of the Web is clear. We need to understand the structure and topology of the web and the laws of connectivity and scaling to which it appears to conform. Such analyses also showed the Web to have scale-free and small-world networking structures, areas that have largely been studied by physicists and mathematicians using the tools of complex dynamical system analysis (Berners-Lee et al 2006a).

Sensemaking, reuse and retrieval is vital, particularly when the Web is seen as a learning technology. The Web is currently largely made up of linked documents, often text documents, so Natural Language Processing techniques add value by extracting some form of meaning from the human-readable text of the pages, based on heuristics or statistics. But an increasingly important extension, the *Semantic Web*, envisages linking *data* resources enriched by reference to ontologies which give interpretations of terms used; in such a Semantic Web of relational data and logical assertions, computer logic is in its element, and can do much more. This development is exciting yet challenging. How can we allow independent consistent data systems to be connected locally without requiring global consistency? How do we effectively query an unbounded Web of linked information repositories? How should we align different data models, and visualise and navigate the huge connected graph of information that results? Who should control access to data resources shared on the Web (Shadbolt et al 2006)?

The field of learning technology is one example of an area which can enhance, and be enhanced by, Web Science. Understanding the processes of learning and instruction, with an eye to the creation of and support for facilitating technologies is an example of the modelling-engineering combination that Web Science advocates (cf. e.g. O'Neil & Perez 2006). Learning technologies encompass a range of artefacts, from hardware to methods of organising information. A key aim for the future Web is to remain the world's pre-eminent information store, and the timely retrieval and reuse of information placed on it is obviously a central objective for Web developers. The coherence between the missions of the Association for Learning Technologies and Web Science is plain: both wish to maintain the Web as a valuable educational tool.

The Web Science Research Initiative (WSRI)

To promote Web Science and explore its emerging agenda, a joint endeavour between the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology, and the School of Electronics and Computer Science at the University of Southampton, was set up in 2006, called the *Web Science Research Initiative* (webscience.org/). WSRI's mission is to foster the fundamental advances required for the Web's continued growth. In particular, WSRI is focusing on steering the development of the Web Science discipline, running a series of workshops and looking at the lines of an academic curriculum for teaching Web Science. There will be an International Web Science Conference held in Athens, Greece, in 2009 – hopefully the first of many – as well as a new journal *Foundations and Trends in Web Science*.

WSRI's founding directors are Tim Berners-Lee, Wendy Hall, Nigel Shadbolt and Daniel Weitzner, and James Hendler of Rensselaer Polytechnic Institute is an Associate Director. The directors are supported by an international Advisory Board, and a Scientific Council whose members have researched and engineered the Web with great distinction.

Conclusion

Web science is not just modelling the current Web. It is about engineering new infrastructure protocols by using scientific and technological tools from many disciplines to understand the human society that uses them, to create beneficial new systems – which may involve extremely radical thinking about both technology and society (Shneiderman 2007). Such new engineering must respect the invariants of the Web experience: decentralisation to avoid bottlenecks and allow increases of scale; serendipitous reuse of information; fairness, openness and trust. In this way the Web will remain a technology that enhances human society, and supports human aspiration.

References

Berners-Lee T, Hall W, Hendler J A, O'Hara K, Shadbolt N and Weitzner D A (2006a) A Framework for Web Science. *Foundations and Trends in Web Science* **1, 1**: 1-130

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

O'Neil HF and Perez R S (eds) (2006) *Web-Based Learning: Theory, Research and Practice*. Lawrence Erlbaum Associates.

Shadbolt N, Hall W and Berners-Lee T (2006) The Semantic Web Revisited. *IEEE Intelligent Systems* **21, 3**: 96-101

Shneiderman B (2007) Web Science: A Provocative Invitation To Computer Science. *Communications of the ACM* **50, 6**: 25-27

Kieron O'Hara
Senior Research Fellow
School of Electronics and Computer Science
University of Southampton
Highfield
Southampton S017 1BJ
kmo@ecs.soton.ac.uk

Wendy Hall
Professor of Computing
School of Electronics and Computer Science
University of Southampton
Highfield
Southampton S017 1BJ
wh@ecs.soton.ac.uk