Universitat Oberta de Catalunya

uocpapers e-Journal on the Knowledge Society

http://uocpapers.uoc.edu

article

"Web Science" dossier

Science, the Web and Web Science

Daniel Riera

Submission date: September 2008
Accepted in: September 2008
Published in: October 2008

Abstract

The evolution of the World Wide Web (or simply the web) since its birth in the 1980s has been staggering. The incorporation of new users, which has been exponential over time, and the constant contributions of new ways of organising data, communicating, sharing information, etc., have led to its gaining a practically unstoppable inertia in just two decades. The Web Science Research Institute, WSRI, was created in November of 2006 with the main aim of proposing a new discipline, Web Science, to observe the web and all that surrounding it. It defends the need to analyse what is going on inside and outside the web and, thus, be able to propose improvements and corrections. This idea requires combining disciplines that have, to date, been very disperse, such as IT, psychology, law or economics. This has led to a new professional profile, the Web Scientist, and, likewise, new academic needs. This article details the work carried out by the WSRI in its first two years of existence, with regard to the new knowledge area of Web Science.

Keywords

Web Science, historical review, education, WSRI

Resum

L'evolució del World Wide Web (o simplement web) des del seu naixement a la dècada dels vuitanta del segle xx ha estat gegant. La incorporació de nous usuaris, exponencial en el temps, així com l'aportació constant de noves maneres d'organitzar les dades, comunicar-se, compartir informació, etc., han fet que el web hagi assolit en només dues dècades una inèrcia pràcticament imparable. El novembre de 2006 es crea la Web Science Research Initiative (WSRI) amb l'objectiu principal de proposar una nova disciplina, la «ciència del web» (Web Science), que observi la xarxa i tot allò que l'envolta. Defensa la necessitat d'analitzar què passa dins i fora de la xarxa i, d'aquesta manera, poder-hi proposar millores o correccions. Aquesta idea implica combinar disciplines tan llunyanes fins ara com la informàtica, la psicologia, el dret o l'economia. Tot això implica l'aparició d'un nou perfil professional, el científic del web, i, per tant, de noves necessitats acadèmiques. Aquest article mostra la feina feta per la WSRI en els primers dos anys d'existència amb relació a la nova àrea de coneixement: la ciència del web.

Paraules clau

Web Science, ressenya històrica, educació, WSRI

The inertia of the web

According to Sir Isaac Newton's first law of motion (1643-1727), "an object at rest tends to stay at rest and an object in uniform motion tends to stay in uniform motion unless acted upon by a net external force". This first law is called the law of inertia[1] and attempts to explain the attitude of systems when faced with movement: if they are at rest, they will make an effort to stay at rest, and if they are moving, they will try to continue the movement. Therefore, depending on the object (or system), we can imagine, without going too deeply into the physics behind it, that this object will have more or less "resistance" to change. Consequently, if we have a football and a cannonball of the same size, both placed side by side, it would seem easier to make the former move than the latter. Equally, it will be easier to stop a lorry that is out of control at 1 kmh^{-1} than a motorbike, which is much smaller, at about 200 kmh^{-1}.

1. *Inertia*, in Latin, means leisure or laziness.

When Sir Timothy John Berners-Lee proposed the idea of connecting hypertext with TCP and DNS in 1989, he was making available to everyone (as the infrastructure already existed) what we now call "the web" (or the World Wide Web). In order to find an easy way of instantly sharing and updating information between scientists, a network was created that would enable contents that were far apart to be viewed as though they were on an individual's own computer. From this moment on, and with this little shove, a small snowball began to roll.

In almost no time (less than two decades), the web has gained inertia. First, the incorporation of the academic and, then, the business and private fields have expanded the size of the invention. In addition, increased features and the lowering of prices, linked to swift developments in hardware, and the incorporation of new, highly attractive ideas into the web, have enabled it to mature and become enriched very quickly.

If, initially, the web was viewed as the connection of a group of static (or almost static) contents, these contents have become increasingly dynamic. The web is now called the Web 2.0[2] and appears to want to become 3.0 (Markoff, 2006). The current second version leans towards creativity, information sharing (not only among scientists) and collaboration through social networks, wikis or blogs, for example. Many of these components had been in existence since the web began, though they were not so widely known about or used. The main proposal of the future (third) version is to transform the use of the web's information into a semantic web (Berners-Lee *et al.*, 2001), with all the data globally interconnected and accessible, and with search engines that incorporate artificial intelligence. Likewise, the content will be primarily open, as will users' identities, and must be portable to any machine connected to the web.

The entity that the web has become, from the initial shove that it received from Tim Berners-Lee around 20 years ago, is immense. Large-scale expansion began in 1995, when it already had 16 million users. By 2001, there were 400 million and 2005 saw it reach the first billion-user mark. It is expected that there will be more than two billion[3] by 2010.

These two types of exponential growth (qualitative and quantitative) make the accumulated inertia seem uncontrollable. A good test of this is that, in terms of teaching, what a few years ago could just be summarised in a couple of lessons in an IT degree subject, is now a volume that could constitute a complete degree all by itself.

Given all this, on November 2 2006, four researchers from MIT and the University of Southampton founded the Web Science Research Initiative (WSRI). In their presentation, they argued the need for a new interdisciplinary field – which would embrace mathematics, IT, psychology, econòmics, law, etc. – to analyse what the web is and what happens on it and around it. A field showing the vulnerabilities, the interactions that need improving, the implications that microscopic decisions have on global design, etc. A science that works on both the technical and social aspects. They also suggested that this new area should be led by specialists skilled in this wide range of subjects. This requires the appearance of teaching devoted specifically to Web Science.

The Web Science Research Initiative

There is a great proliferation of international meetings linked in one way or another with the web. Despite this, it was not until the creation of the WSRI that the web began to be treated in a global and multidisciplinary way. From the start, the WSRI has tried to broadcast its purpose through publications and presentations at international congresses and conferences. It has tried to bring together experts from different fields so that they can all start sharing the idea of Web Science.

To achieve this, The Web Science Workshop was held in 2005 at the London offices of the British Computer Society[4]. The Emerging Science of the Web[5] was attended by 21 participants. The central topic of the encounter was the exploration of the critical challenges of the future of the World Wide Web (Hendler, 2005). This was the first global formal discussion on Web Science.

Subsequent workshops were held exclusively on this subject at renowned conferences, such as the Seventeenth[6] and Eighteenth[7] International World Wide Web Conferences. During these workshops, participants were asked to look at the future of the web and discuss how this web may be observed, analysed and influenced. Similarly, Hypertext 2008 (HT2008)[8] included the Web Science workshop: Collaboration and Collective Intelligence,[9] which primarily discussed new types of collaboration on the network, its structure and user behaviour.

Many other talks and presentations have been held between these workshops, which has helped the WSRI and its members to start spreading their ideas (Weitzner *et al.*, 2007; Hall, 2008;

2. Not everyone likes this name. For example, Tim Berners-Lee does not think it appropriate. In any case, it seems to be widely accepted generally by the community.
3. See: <http://www.internetworldstats.com/stats.htm>.
4. See: <http://www.bcs.org/>.
5. See: <http://www.cs.umd.edu/~hendler/2005/WebScienceWorkshop.html>.
6. See: <http://www2008.org/>.
7. See: <http://www2009.org/>.
8. See: <http://www.sigweb.org/ht08/>.
9. See: <http://users.ecs.soton.ac.uk/dem/workshops/webscience08/>.

Hendler *et al.*, 2008). In addition, some of the founding fathers of the WSRI have been interviewed or have written for general interest newspapers (Cellan-Jones, 2008; Lohr, 2008) or for scientific publications (Schneiderman, 2007; Hendler *et al.*, 2008b). At the Web Science Curriculum Workshop in September 2008, the debate concentrated on the objectives of the initiative that are associated with the teaching of Web Science, and the First International Web Science Conference, which is to be held in Athens in March 2009, has been planned for the near future.

A quick search on Google for "Web Science" brings up around three and a half million pages on the subject. Apart from the congresses, articles and interviews that have already been mentioned, these pages show blog entries in a number of languages (Saravanan, 2007; Peña, 2007; Ragget, 2008), news and articles in online journals and newspapers (Biever, 2006; LaMonica, 2006), and some definitions in wikis. Most of these articles and blog entries are from November 2006 and only relate to the initial WSRI announcement.

One of the last (and first) texts published is the provocative book *The Web's Awake* (Tetlow, 2008), which presents the web as a new way of life that we can not control anymore. It provides a description of new web characteristics and types of behaviour that we have not created, but which have emerged of their own accord.

Interestingly, we should point out that the Wikipedia[10] definition of Web Science is rather poor. In line with the spirit of the web 2.0, it is users who should complete this content.

A Framework for Web Science

Given that this is a young field, little has been written about Web Science. Of the few references that we have, *A Framework for Web Science* (Berners-Lee *et al.*, 2006) is possibly the most complete. It is the book[11] with which the WSRI has tried to establish the foundations of the idea of Web Science. It deals with a number of core issues in the development of the web without going into excessively technical details. Given the quantity and range of knowledge in the web environment, it does not appear easy to bring it together under a single, homogeneous heading. For this reason, the aim of the book is not to be a collection of this technical knowledge, but rather a more informative introduction to the main topics that would constitute Web Science.

Consequently, after a brief introduction, the science and the engineering behind it, analysis of the web, social aspects involved and finally the legal, governance, security and standards issues are presented. A great effort is made to establish the connection between very wide-ranging subjects which so far have been analysed in isolation, or, if not, at least independently. Although these subjects are not fully linked, we are beginning to see a structure of the new science proposed.

Some of the aspects emphasised in the book are related to web organisation, from both the social and content, or infrastructures and resources points of view. This underlines the importance of well-designed web architecture. There is also a commitment to organisation and intelligent access to content (or information) through the use of labels, the Semantic Web, etc. It describes the importance of using mathematical models to analyse the web, which has to lead to continuous improvement. In addition, the implication of the social, cognitive and moral contexts of the web on its engineering decisions are also noted. Finally, it asks what type of regulation should be applied to the web and its users. Primarily, this includes areas related to user security (e.g. privacy and identity), improper behaviour, the importance of everyone accepting standards and the search for the web's governance.

The main shortcoming of the text is that the heterogeneity of the topics that it deals with makes a homogeneous study of the discipline itself difficult. At times, the text seems to be simply a review of a collection of well chosen articles. It would possibly have been more interesting if greater effort had been made to coordinate the different texts. Despite this, it is a valuable work that sets out the bases of Web Science. It covers a range of subjects that are fairly representative of what the web now is and the directions it is taking. However, it does seem strange that despite covering so many subjects there is not the slightest mention of the web's influence on teaching, or concepts such as e-learning or distance education when this is one area where the development of the web has had the most influence.

Academic contributions

One of the objectives of the WSRI, if not in the short term then in the medium term, is the joint design between the MIT and the University of Southampton of a Web Science degree. To date, this degree has not been offered and there is no indication as to when it will be up and running or the progress made so far.

An initiative closely associated with Web Science is iSchools (or Schools of Information),[12] presented in September 2005 at Pennsylvania State University. This proposal differs from Web Science in that it focuses more on relations between information, technology and people, irrespective of the platform (which

10. See: <http://en.wikipedia.org/wiki/Web_science>.
11. It is in fact not strictly a book, but the first edition of the journal on Web Science: *Foundations and Trends in Web Science*.
12. See: <http://www.ischools.org/oc/>.

does not necessarily have to be the web). The Oxford Internet Institute[13] also offers training on the web, though this is aimed primarily at Masters and Doctoral level. Likewise, there is the thread concept as proposed by Georgia Tech,[14] which offers students the possibility to define a curriculum, which, although some way away, resembles to some extent a Web Science degree. In any event, these threads are more akin to what is now seen as a specialisation within a degree, and not a degree in itself.

Finally, the proposed future IT Degree from the UOC includes an optional subject devoted to Web Science. Work is also underway to prepare postgraduate teaching in this field.

Conclusions

On the basis of the WSRI proposal and the most easily accessible information to date, it seems that Web Science is still at an early stage of development. Its members continue to strive to make the proposal more widely known and give the impression that the core discussion subject is still somewhat disperse. Clearly, it is not easy to give a precise definition of what comes under this new area, what is left out, and what interactions there are among the sub-areas.

All this does not make designing the new associated degrees easy. At present, MIT and the University of Southampton might supposedly be preparing the first Web Science degree, while the remaining universities await it. The curriculum of this new degree has to be widely accepted by a significant number of university departments, which complicates even more the design.

Despite this, the proposal is only two years old and significant efforts seem to be underway to bring it into being and provide a solid definition of the new field. The importance that the web needs to take on in twenty-first-century society could be compared to the significance of penicillin, the printing press, the steam engine or electricity in previous centuries. Web Science is essential to ensure the creation of professionals who are capable of understanding and influencing the web and that which surrounds it.

Reference

BERNERS-LEE, T.; HAENDLER, J.; LASSILA, O. (2001). "The Semantic Web" [online article]. *Scientific American*. [Accessed: 20/08/2008]. <http://www.sciam.com/article.cfm?id=the-semantic-web>

BERNERS-LEE, T.; HALL, W.; HENDLER, J. A. [*et al.*] (2006). "A Framework for Web Science". *Foundations and Trends in Web Science*, vol. 1, no. 1.

BIEVER, C (2006). "Web pioneers call for new 'web science' discipline" [online article]. *New Scientist Tech.* [Accessed: 20/08/2008]. <http://technology.newscientist.com/article/dn10438-web-pioneers-call-for-new-web-science-discipline.html>

CELLAN-JONES, R. (2008). Article and interview with Dr Wendy Hall [online]. BBC [Accessed: 20/08/2008] <http://www.bbc.co.uk/blogs/technology/2008/03/what_is_web_science.html>

HALL, W. (moderator) (2008). *Does the Semantic Web Need Web Science?* [online] Debate panel at the 5th European Semantic Web Conference, Tenerife. [Accessed: 20/08/2008]. <http://videolectures.net/eswc08_hall_dsw/>

HENDLER, J. (2005). *Web Science Workshop. The Emerging Science of the Web* executive summary [online]. [Accessed: 20/08/2008]. <http://www.cs.umd.edu/~hendler/2005/WebSciReport.pdf>

HENDLER, J.; BERNERS-LEE, T.; HALL, W. [et al.] (2008). Presentation of the *Tetherless World Constellation: The Future of the World Wide Web* [online]. [Accessed: 20/08/2008]. <http://www.rpi.edu/news/events/tw/index.html>

HENDLER, J.; BERNERS-LEE, T.; HALL, W. [*et al.*] (2008b). "Web Science: An interdisciplinary approach to understanding the World Wide Web". *Communications of the ACM*, 51, 7 (pp. 60-69).

LaMONICA, M. (2006). "Berners-Lee, universities launch 'Web Science' initiative" [online article]. *c/net News.com*. [Accessed: 20/08/2008] <http://news.cnet.com/Berners-Lee,-universities-launch-Web-science-initiative/2100-1025_3-6132016.html?hhTest=1 , 2 November 2006>

LOHR, S. (2006). "Group of University Researchers to Make Web Science a Field of Study" [online article]. *The New York Times*. [Accessed: 20/08/2008]. <http://www.nytimes.com/2006/11/02/technology/02compute.html>

MARKOFF, J. (2006). "Entrepreneurs See a Web Guided by Common Sense" [online article]. *New York Times*. [Accessed: 20/08/2008]. <http://www.nytimes.com/2006/11/12/business/12web.html>

PEÑA, I. (2007). *Last Thoughts About Web Science And Academic Blogging Or Why Did Not Academia Came Up With Wikipedia. And Some Acknowledgments Too* [online article (blog)]. [Accessed: 20/08/2008].

13. See: <http://www.oii.ox.ac.uk/>.
14. See: <http://www.gatech.edu/>.

<http://ictlogy.net/review/20070729-oii-sdp-2007-epilogue-last-thoughts-about-web-sciences-and-academic-blogging-or-why-did-not-academia-came-up-with-wikipedia-and-some-acknowledgments-too/>

RAGGET, D. (2008). *Web Science and link rot, a way forward* [online article (blog)]. [Accessed: 20/08/2008]. <http://people.w3.org/~dsr/blog/?p=7>

SARAVANAN (2007). *Web Science: Web is a network of people* [online article (blog)]. [Accessed: 20/08/2008]. <http://saran945.com/blog/web-science-web-is-a-network-of-people/>

SHNEIDERMAN, B. (2007). "Web science: A provocative invitation to computer science". *Communications of the ACM*, 50, 6 (pp. 25-27).

TETLOW, P. (2008). *The Web's Awake: An Introduction to the Field of Web Science and the Concept of Web Life.* Piscataway: IEEE Press.

WEITZNER, D.; ALONSO, J. M. (2007). Summary of the *Toward More Transparent Government Workshop on eGovernment and the Web Workshop* [online]. [Accessed: 20/08/2008]. <http://www.w3.org/2007/06/eGov-dc/summary>

Recommended citation

RIERA, Daniel (2008). "Science, the Web and Web Science". In: Julià MINGUILLÓN (coord.). "Web Science" [online dossier]. *UOC Papers*. Iss. 7. UOC [Accessed: dd/mm/yy].
<http//www.uoc.edu/uocpapers/7/dt/eng/riera.pdf>
ISSN 1885-1541

Daniel Riera
Director; IT Engineering Programme; IT, Multimedia and Telecommunications Department; UOC.
drierat@uoc.edu

PhD in IT Engineering from the Autonomous University of Barcelona (UAB) in 2006, Masters in Advanced Process Automation Techniques in 2002 from the UAB and degree in IT Engineering, also from the UAB, in 1998.

As a Ministry scholarship holder with a Research Personnel Training grant, he spent a year at Imperial College London, as a visiting scholar, working in the ICParc planning and resource control department. It was here that he received training in constraint programming and optimisation, and collaborated on projects for the RAC and Railtrack.

He has taught at the UAB from September 1998 to August 2005 and since September 2007. In September 2005, he became a lecturer in the UOC's IT, Multimedia and Telecommunications Department. Since September 2006, he has been working as the Academic Director of the Postgraduate course in Bio-computing currently offered by the UOC. His teaching work focuses primarily on programming and bio-computing. In January 2007, he was named director of the department's IT Engineering Programme.

His main areas of research include discrete systems modelling using Petri nets and optimisation with constraint programming techniques. He is the author of a number of articles and papers for national and international journals and conferences. He has participated in research projects and agreements as a member of the research team. He has worked as a researcher with the LOGISIM, the Centre for Logistics Systems Simulation and Optimisation, part of the CIDEM's Network of Technological Innovation Support Centres. He is currently a member of the UOC's software engineering research group (GRES-UOC), where he is researching verification of UML/OCL models using constraint programming.