The Science of the World Wide Web

James Hendler, Rensselaer Polytechnic Institute

Wendy Hall, Southampton University

Ten years ago, in an article in this magazine [1], we were two of the authors arguing for the need to create a field that explored the underlying science, the engineering principles and the social impacts of the World Wide Web. At that time, Wikipedia was still in its infancy (and totally dismissed by the establishment), Facebook was still restricted to university users, Twitter was in beta and improved search dominated Web conference research agendas. There were virtually no smart phones, on-line surveillance was largely unknown beyond security services, and no one knew that being a data scientist was one day going to be “the sexiest job in the world” [2].

Over the last ten years, the exponential Web growth has continued and a deep understanding of the technologies involved is even more vital to a society that has become increasingly dependent on Internet technology for all aspects of communication. The 2006 article and its follow-ons helped fuel the creation of an emerging interdisciplinary field now known as Web Science with academic programs in multiple countries and with conferences and journals now existing for researchers in the area.

In the past decade, new uses of the Web in both scientific and public discourse have grown, and shared data on the Web increasingly has become both a critical research resource and a challenge to manage technically and socially. Web Science researchers have focused on several different aspects of the Web and of its emergence. Much of the interest in data science today is coming about because of the vast stores of information, both structured and unstructured, that have become available on the Web. The information available through mobile web platforms, such as geolocation information from smart phones, now powers a growing segment of new industries such as uber, lyft, airBnB and other parts of what has become known as the sharing economy. Social media analysts look to understand, mathematically and socially, the trends being seen on the Web as reflected through information shared on social networking web sites and mobile applications. Governments in cities and countries around the world now release data on the Web in open formats [3], open publishing is increasingly making journal and conference papers freely available to researchers, and online forums, such as PatientsLikeMe.com, provide new sources of information that researchers can tap.

The emerging Internet of Things promises to yield even more data including much real-time data on the movement not just of people, but of energy and other resources needed for modern society. The growth of available information is also leading to increasing use of data analytics in many fields, and the intersection of network-, data- and web- sciences is helping to bring new technologies to scientists and engineers working on large scale problems [4].

One of the goals of Web Science is to be able to track and explore trends and usages of the information space that abounds. To this end, a number of research groups in Web Science laboratories around the world have begun a project called the “Web Observatory” to collect and share data about the Web and its use [5]. Increasingly, the move is from static analyses to tracking change in real-time, and to improve predictive models for understanding the impacts of information use across the network. Repository metadata and lightweight standards have been developed, and information is now being tracked from twelve countries on five continents. The continued development of a global Web observatory and data repository is enabling researchers to track and analyze past patterns of web usage and growth, and in the future predictive models may help us better understand the impacts of emerging technologies.

Researchers are also studying crowd-sourcing, collective intelligence and citizen science and looking specifically at what can be learned from the collected data these technologies can generate. Web sites such as Wikipedia are powered by an interaction of many people, in many roles, and increasingly we see the growth of citizen science sites, like the “zooniverse” of scientific applications derived from the astronomical GalaxyZoo site [6,7]. These sites, harnessing the cognitive capacities of many millions of people, create powerful “social machines.” Developing the principles for the successful design and governance of these sites, and lowering the barriers of entry for scientists and others in creating them, remains an active area of Web Science research [8,9].

As the information on the web continues to grow, so increasingly will the tension between the privacy desires of data providers and potential use of the information for the social good. Issues such as internet governance and policy, re-decentralization, digital identity, digital literacy, and the ethical issues our society faces as more and more of our lives are lived on-line are major analytic challenges for the future [10]. Web Scientists must grasp these issues and tackle them using the tools and methodologies we have begun to develop over the last ten years in a truly interdisciplinary way.

However, there is a challenge to the emerging field. As we see the increasing emphasis by politicians and funding bodies on the economic advantages of more students entering the STEM area, there has been a concomitant tendency to treat the social sciences and humanities as if they are somehow of lesser value. However, not only is the Web a network of machines, it’s a network of billions of people throughout the world interacting together in never-before-possible ways. Appreciating the impact of the Web’s science and engineering on our society, and conversely the impact society has on the development of the Web and the Internet, requires a deeper understanding of socio-technical systems than ever before. Web Science therefore is, and must remain, an interdisciplinary pursuit, uniting the talents of a wide swath of researchers from many fields of physical and mathematical sciences, engineering, humanities and social sciences.

The need for Web Science is manifest, and will become ever greater over the next decades. In some ways we have just been preparing the groundwork for the future research needed to ensure that this amazing eco-system we have collectively built remains fit for purpose and especially for the betterment of humanity in the years to come.

[1] Berners-Lee, Tim, et al. "Creating a Science of the Web." *Science,* 313.5788 (2006): 769-771.

[2] Patil, T. H., and D. J. Davenport. "Data Scientist: The Sexiest Job of the 21st Century." *Harvard Business Review,* (2012).

[3] Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. "Benefits, adoption barriers and myths of open data and open government." *Information Systems Management,* 29.4 (2012): 258-268.

[4] O'Hara, Kieron, et al. "Web Science: understanding the emergence of macro-level features on the World Wide Web." *Foundations and Trends in Web Science,* 4.2-3 (2013): 103-267.

[5] Tiropanis, Thanassis, et al. "The web science observatory." *IEEE Intelligent Systems,* 28.2 (2013): 100-104.

[6] Lintott, Chris J., et al. "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." *Monthly Notices of the Royal Astronomical Society,* 389.3 (2008): 1179-1189.

[7] Simpson, Robert, Kevin R. Page, and David De Roure. "Zooniverse: observing the world's largest citizen science platform." *Proceedings of the 23rd International Conference on World Wide Web*, ACM, 2014.

[8] Hendler, Jim, and Tim Berners-Lee. "From the Semantic Web to social machines: A research challenge for AI on the World Wide Web." *Artificial Intelligence,* 174.2 (2010): 156-161.

[9] Shadbolt, Nigel R., et al. "Towards a classification framework for social machines." *Proceedings of the 22nd International Conference on World Wide Web*, ACM, 2013.

[10] Cranor, Lorrie, et al. "Towards a Privacy Research Roadmap for the Computing Community." *arXiv preprint arXiv:1604.031*(2016).