Web science: a new frontier

Professor Nigel Shadbolt FREng, Professor Dame Wendy Hall FRS, Professor James Hendler and Professor William Dutton

Over the last 20 years humans have built the largest information fabric in history. The World Wide Web has been transformational and has assumed a central role in our lives - people shop, date, trade and communicate using it. We now take the Web for granted but over the past few years there has been a growing recognition that the Web ecosystem needs to be treated as an important and coherent area of study – this is Web science.

It is "science" in the original and broad sense of the term: the quest to build an organised body of knowledge to better understand a fundamental aspect of the world we live in. It will need to embrace engineering – the Web is an engineered construct, a set of protocols. It will need to embrace human and social sciences – the Web is a social phenomenon that has produced transformative behaviours. The Web is a space built and used by people for people. We will not understand it if we simply reduce it to its technological parts.

Emergence is one of the key features of the Web – from the emergence of the "blogosphere" to the appearance of Wikipedia, from the linking of scientific data to the formation of social networks – complex structures emerge from apparently simple principles. We need to understand the feedback loops between engineering solutions, applications and new behaviours. At the moment these are charted only in retrospect.

Nevertheless important discoveries have been made. During the past decade the Web's connectivity has been studied – we have learnt that wherever you look, some pages have many more links to them than others and this distribution looks the same at whatever scale you sample. As the Web is "scale free" even if a majority of the pages were removed, a path from one page to any other is likely to exist.

However, removing a relatively small number of highly connected items would lead to the disintegration of the network.

Furthermore, we have discovered that the Web is a "small world" – despite the billions of pages you can get from one page to any other in 20 or so clicks. The Web has this beneficial structure of short paths because of human behaviour – we group into communities and to understand this we require insights from sociology and psychology.

As the Web grows, changes and expands, researchers are seeking novel ways to explore and navigate content. Progress would give us an understanding of human behaviour in online information seeking and the display of "collective intelligence" – work still in its infancy. Collaborative behaviour leads to the emergence of large-scale resources such as Wikipedia. The challenge is to technically enable such collective resources, and understand what drives people to collaborate. We are trying to understand how to exploit human participation to solve a range of computationally challenging tasks, from classifying astronomical objects to deciphering faded texts and manuscripts.

The Web may appear to be a permanent feature of modern life but there is nothing inevitable about its continued existence. There is work that seeks to ensure that the Web remains distributed yet stable, open and secure. The principles of universal access and open formats underlie the Web's basic protocols. Changes to these formats or to the assumption of content accessibility could have far reaching consequences.

As the Web evolves there is talk of the "experiential Web", which would encompass and encode our daily activities, capable of replaying these experiences in a range of media. It is a Web in which a greater number of

2 billion

people enjoy access to the Web, the majority do not

Main image - Visualisation of round-trip times across a wide cross-section of the Internet. © 2003 The Regents of the University of Cal fornia. All rights reserved.

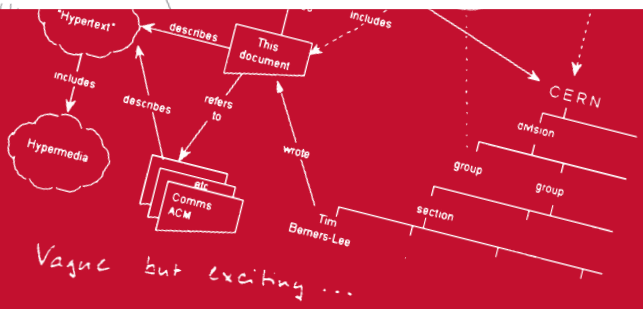
Archive image - Schematic drawn by Tim Berners-Lee FRS in his proposal for an information management system at CERN. Described as 'vague, but exc ting' by his supervisor, the distributed hypertext system described in the proposal eventually became the structure behind the World Wide Web. © CERN, 1989.

objects could have a Web presence – from refrigerators to buses, items of clothing to scientific sensors. In this Web what should we expect in terms of privacy? As we link increasing amounts of data together, as the Web of linked data emerges, new opportunities and new issues will surface.

The Web has had profound social impact; but, whilst two billion people enjoy access to the Web the majority do not. As we see the Web take hold in the developing world, we are witnessing new and exciting behaviours.

Understanding the Web will have wide reaching implications. It is on a par with other great scientific challenges such as understanding the climate, our biological nature or the larger Universe. But it must be interdisciplinary, drawing insights from mathematics, physics, computer science, psychology, ecology, sociology, but also law, political science, economics and more.

If we are to anticipate how the Web will develop we will require insight into our own nature and the dynamics of policy and practice worldwide. Web science is not only a new frontier it is an endeavour that will bring together a new generation of enquiring minds.



From the archive

It is a testament to how useful the World Wide Web is in sharing information that most of us do not think too hard about how it works, but could not imagine living without it. Even in past times there were physical world equivalents in borderless communication and data sharing, driven by humanity's need to think, work and play.

The Royal Society's existence and purpose owe much to the 17th century commonwealth of letters. This was surprisingly international so that scholars from as far afield as China could communicate their ideas, pictures and specimens with co-workers in London. It was an information network of sorts but of course there were drawbacks: without the underpinning agency of the internet, information moved at the speed of the wind and remained at the mercy of tides. More seriously for scientists, years' worth of data, in the form of notes and specimens, could be lost at sea.

19th century communication technologies sped matters up considerably, but the electric telegraph and telephony remained, essentially, messaging systems. By the 20th century, the combination of machine-read microforms and prototype computing led futurologists such as H.G. Wells to speculate on a *World Brain*[12], so that "any student, in any part of the world, will be able to sit with his projector in his own study at his or her convenience to examine *any* book, *any* document, in an exact replica".

Wells thought that this encyclopaedic vision would usher in a social revolution. When Tim Berners-Lee FRS floated the hypertext project which resulted in his creation of an internet-based web client and server at CERN in 1989-90, the social revolution of the internet was the eventual result. The World Wide Web has changed scientific research forever in terms of sharing ideas and results through online publishing and data exchange.