The Emerging Science of the Web:
And Why it is Important

Professor Dame Wendy Hall
23 June 2010
Inspiration
“As we may think”
Vannevar Bush
Atlantic Monthly
July 1945
Everything is deeply intertwingled

Ted Nelson and Doug Engelbart

Augmenting human intellect
What do India, the Earl Mountbatten of Burma, the University of Southampton and my research career have in common?
The Mountbatten archive moved to Southampton in 1987.
Microcosm: Mountbatten archive application

Viceroy's House, New Delhi, at 11.15 hours on Wednesday, 2nd April 1947

The 2nd miscellaneous meeting of His Excellency the Viceroy was held.

The item discussed was:-

The Indian National Army

Present

The Viscount Mountbatten of Burma, Viceroy of India
The Hon'ble Pandit Nehru, Member for External Affairs and Commonwealth Relations

*The Hon'ble Patel, Member for Home Affairs, Information and Broadcast
The Hon'ble Liaquat Ali Khan, Member for Finance
The Hon'ble Sardar Baldev Singh, Member for Defence
Field Marshal Sir Claude Auchinleck, C in C [CINC] India

Analysing Document

[Transcript]

Day of typescript minutes for the miscellaneous meeting, held on 2 April 1947.

Members of the Indian National Army imprisoned for Indians, 2 April 1947; contemporary copy typescript minutes of the Viceroy's second
Linkbases in Microcosm

Note the direction of this arrow!

Link database

Separable hyperstructure

Documents
Links in Microcosm

- source, destination, description
- source: object | concept | context
- We generated links based on metadata description of documents in docuverse and “it all falls out”
ECHT’90
ACM Hypertext’91
Lessons learnt

• Big is beautiful: the network is everything
• Scruffy works: let the links fail to make it scale
• Democracy rules: open, free and universal

• But we lost (for a time) conceptual and contextual linking, and link descriptions – the Web is a strangely linkless world
• Missing links – search engines fill the gap
Internet Growth - Usage Phases - Tech Events

Read Only Web | Read/Write Web | Social Web

Mark Schueler, PhD student

Note – events shown relate to the time axis only.
Web 2.0

- Wiki’s
- Blogs
- Flickr
- YouTube
- MySpace
- Facebook
- Second Life
- Twitter
Welcome to Wikipedia,
the free encyclopedia that anyone can edit.
3,102,741 articles in English

Today's featured article

Grover Cleveland (1837–1908) was both the 22nd and 24th President of the United States. Cleveland is the only President to serve two non-consecutive terms (1885–1889 and 1893–1897) and therefore is the only individual to be counted twice in the numbering of the presidents. He was the winner of the popular vote for President three times—in 1884, 1888, and 1892—and was the only Democrat elected to the Presidency in the era of Republican political domination that lasted from 1860 to 1912. Cleveland's admirers praise him for his honesty, independence, integrity, and commitment to the principles of classical liberalism. As a leader of the Bourbon Democrats, he opposed imperialism, taxes, subsidies and inflationary policies, but as a reformer he also worked against corruption, patronage, and bossism. Critics complained that he had little imagination and seemed overwhelmed by the nation's economic disasters—depressions and strikes—in his second term. Even so, his reputation for honesty and good character survived the troubles of his second term.

(more...)

Recently featured: Werner Mölders – Electron – Han Dynasty
We’re hungry to share data and get answers to questions.

Welcome to Galaxy Zoo, where you can help astronomers explore the Universe.

New, more detailed images added - see here for details.

The Galaxy Zoo files contain almost a quarter of a million galaxies which...
The Semantic Web
A Web of Data
The Semantic Web

Semantic Web LayerCake (Berners-Lee, 99; Swartz-Hendler, 2001)

RDF triples: Subject, Predicate, Object
Tim Berners-Lee on the next Web

Filmed Feb 2003, Posted Mar 2009

About this talk

20 years ago, Tim Berners-Lee invented the World Wide Web. For his next project, he's building a web for open, linked data that could do for numbers what the Web did for words, pictures, videos: unlock our data and reframe the way we use it together.

About Tim Berners-Lee

Tim Berners-Lee invented the World Wide Web. He leads the World Wide Web Consortium, overseeing the Web's standards and development. Full bio and more links

About our sponsor

A dive into the rich tradition of Rolex innovation, revealing unprecedented watchmaking achievement

What to watch next
Content, Emergence and Unanticipated Reuse

The four micro principles of the Semantic Web

1. All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references.
2. URI references should be dereferenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.
3. Data should be provided using the RDF/XML syntax.
4. Data should be interlinked with other data.
Linked Data on the Web: May 2007

500 Million RDF Triples
120,000 RDF links between data sets
Linked Data on the Web: April 2008

23 billion RDF Triples
3 million RDF links between data sets
LOD Datasets on the Web: March 2009

4.5 billion triples
180 million data links
Tipping points?
Obama’s Groundbreaking use of the Semantic Web

by David Peterson

In a revolutionary move, Obama’s administration is set to utilise next generation web technologies to bring an unprecedented level of transparency to government. In this case it will shed light on how the roughly US $800 billion dollar economic stimulus will be spent. The recently launched recovery.gov website (powered by nothing other then Drupal) brought with it the promise that citizens would be able to view where the money was going and how it was going to be spent.

To enable the citizen masher to do their wizardry, the administration will be opening up a veritable candy store of goodies: Semantic Web, RDF, Linked Data, SPARQL, RDFa, SIoC, ATOM, RESTful APIs, JSON, Widgets, Wikis, XForms, P2P Networks. Wow. They only forgot the lions and tigers and bears oh my... This is an unbelievable stack of technology. I didn’t think the government even knew what an RSS feed was :)

Does your system cater for SMS alerts?

SMS enable your system alerts and enjoy the mobility of push technology, simple set-up and rapid integration.

SMS alerts can reach over 712 mobile networks using a range of API connections including HTTP/S, FTP, SMPP, SMTP, XML, Com Object & SOAP connections.

FIND OUT HOW TO SMS ENABLE YOUR SYSTEM IN 5 EASY STEPS:

Download the free developer guide now.

PM welcomes Sir Tim Berners-Lee to Downing Street

Tuesday 15 September 2009

PM welcomes Sir Tim Berners-Lee to Downing Street

The Prime Minister welcomed the creator of the World Wide Web, Sir Tim Berners-Lee, and Professor of Artificial Intelligence at the University of Southampton, Nigel Shadbolt, to Downing Street this morning.

Mr Berners-Lee and Mr Shadbolt presented an update to Cabinet on their work advising the Government on how to make data more accessible to the public.

Gordon Brown has already spoken publicly about his aim of making the UK a world leader in opening up government information on the internet, an important element of Building Britain's Future.
Launched 21st Jan 2010

Unlocking innovation

Working with UK Public Sector information and data

Advised by Sir Tim Berners-Lee and Professor Nigel Shadbolt and others, government are opening up data for reuse. This site seeks to give a way into the wealth of government data and is under constant development. We want to work with you to make it better.

We’re very aware that there are more people like you outside of government who have the skills and abilities to make wonderful things out of public data. These are our first steps in building a collaborative relationship with you.
A GLOBAL MOVEMENT TO DEMOCRATIZE DATA

Most Popular Datasets

1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. OSHA Data Initiative - Establishment...
5. IT Dashboard - Federal IT Spending (major...

Search our catalogs

DATA.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to...
Tomorrow the Web of Linked Data

- We’re eager to share data for all sorts of reasons
- Can we develop a theory of tipping points?
- What are the implications for business?
- What are social and policy implications?
- What will we do with it?
- Web 3.0?
Introduction

our motivation

• the Web has been transformational
• we need to understand it
• anticipate future developments
• identify opportunities and threats
• we have established a new discipline: Web Science
Web Science Research Initiative - launched in November 2006

- Research
- Thought leadership
- Education

research / thought leadership / insight
Web Science is {inter|multi|trans}disciplinary
Web Science is about additionality, not the union of the disciplines but more than their intersection.
Web Science - Examples

Web Structure

Scale-free
The Web has a fractal nature

Power laws
Over the Web the numbers of links into and links out of any Web page obey a Power Law

Small worlds
The average distance (or diameter) is much smaller than the order of the graph.
Web Science - Examples

The Blogosphere

- The Blogosphere
  - Why did it take off?
  - What structure does it have?
  - What drives its evolution?
- Web Science aims to understand the scientific, technical and social factors that drive the growth of the Web
Web Science - Examples
Wikipedia - Collective Intelligence

• What is its structure?
• How stable is it?
• Why do people contribute?
• What lessons does it offer?
Web Science - Examples
Linked data

- Moving from a Web of documents to a Web of data
- Methods for linking data
- Role of the Semantic Web
- Unanticipated reuse

Figure 8: Browsing the Structured Data Web for Proteomics
Web Science Emerges

Studying the Web will reveal better ways to exploit information, prevent identity theft, revolutionize industry and manage our ever growing online lives

By Nigel Shadbolt and Tim Berners-Lee

Since the World Wide Web blossomed in the mid-1990s, it has exploded to more than 15 billion pages that touch almost all aspects of modern life. Today more and more people’s jobs depend on the Web. Media, banking and healthcare are being revolutionized by it. And governments are even considering how to tax their countries with it. Little appreciated, however, is the fact that the Web is more than the sum of its parts. Vast emergent properties have arisen that are transforming society. E-mail and instant messaging, which has led to social networks such as Facebook. The transfer of documents led to file-sharing sites such as Napster, which have led to user-generated portals such as YouTube. And tagging content with labels is creating online communities that share everything from concert news to parenting tips.

But how investigators are studying how such emergent properties have actually blossomed, how we might harness them, what new phenomena may be emerging, or what any of this might mean for humankind. A new branch of science—Web science—aims to address such issues. The missing link, however, computers were built first, and computer science followed, which subsequently improved computing significantly. Web science was launched as a formal discipline in November 2008 when the two of us and our colleagues at the Massachusetts Institute of Technology and the University of Southampton in England announced the beginning of a Web Science Research Initiative. Leading researchers from 16 of the world’s top universities have since expanded on that effort.

This new discipline will model the Web structure, articulate the architectural principles that have fueled its phenomenal growth, and discover how online human interactions are driven by and can change social conventions. It will help us to identify the principles that can ensure that the network continues to grow productively and set complex issues such as privacy protection and intellectual property rights. To achieve these ends, Web science will draw on mathematics, physics, computer science, psychology, sociology, law, political science, economics, and more.

Of course, we cannot predict what this new field will accomplish, but we can predict that it will have a profound impact on our lives. The Web is a complex network of people and ideas that are constantly evolving. It is a platform for innovation that is changing the way we live, work and play. And it is a medium for communication that is transforming the way we interact with each other.

Insights Already

Although Web science as a discipline is new, earlier research has revealed the potential value of such work. As the 1990s progressed, searching for information by looking for key words among the mounting number of pages was returning more and more irrelevant content. The founders of Google, Larry Page and Sergey Brin, realized they needed to prioritize the results.

Their big insight was that the importance of a page—how relevant it is—was best understood in terms of the number and importance of the pages linking to it. The difficulty was that part of this definition was measure: the importance of a page is determined by the importance of the pages linking to it.
WST Outreach and Thought Leadership

- Publications e.g. Foundations and Trends in Web Science
- Impact on research agenda of funding agencies
- Summer Graduate Schools - OII July 2008, RPI July 2009, Koblenz July 2010
- Conferences
  - Web Science 2009, Athens, 18-20 March 2009
  - Web Science 2010, Raleigh Durham, 26-27 April 2010 (co-located with WWW2010)
  - Web Science 2011, Koblenz 15-17 June 2011
- Research talks and workshops all over the world
- Curriculum development
- Sponsors Forum

research / thought leadership education
WSTNet announced at WebSci10 in April

Founding Laboratories

- Southampton
- MIT
- RPI
- Oxford Internet Institute
- DERI, Galway
- Tsinghua Graduate School at Shenzhen
- Koblenz
- VU, Amsterdam
- NorthWestern, Chicago
- ANN, USC
Web Science Doctoral Training Centre

Aim – to create a cohort of web scientists

(a) Develop appropriate research skills,
(b) Understand /use different disciplines
(c) Create a coherent community.

80 students over next 8 years

50 fully funded by RCUK Digital Economy Programme
4 year scholarships (1+3)

PhD students are the life blood of a world-class research lab
They are the key to innovation
The future is mobile

Handheld Philanthropy

research / thought leadership / insight
Web Science
why this matters

• the Web matters
• an essential part of humanity but less than 25% of us have access at the moment
• understanding the Web is a major challenge as big as any other global cause
  – nobody owns the Web
  – what would happen if someone did?
  – could we kill it?
  – it has become our cultural legacy, our social heritage
  – we cannot take for granted the freedom to exchange information that is at the heart of the Web
• For more information see
  www.webscience.org