Web Science

Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh

May 2009
Chercheurs Sans Frontières

Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh

May 2009
Congratulations to ERCIM

Professor Wendy Hall
University of Southampton
www.ecs.soton.ac.uk/~wh

May 2009
ACM Sans Frontières

Wendy Hall
ACM President
Plans for ACM Europe

ACM has 16,000+ members in Europe

Plans to launch ACM Europe Council chaired by Fabrizio Gagliardi

Represented on ACM Council

ACM Europe activities – research meetings, European chapters and SIGs, help nominate European ACM members for fellowship etc.

Similar plans for ACM India and ACM China
ACM Europe Launch event (to be confirmed)

Plans for pre-dinner reception at Informatics Europe Symposium
October 8th, Paris
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

the reputations, experience and skills of our Directors enables us to work closely alongside academia, government, industry and donors to realize our aims

research / thought leadership / insight
Introduction

our aims

promote and encourage multidisciplinary collaborative research
to study the development of the Web
provide a global forum to enable academia, government and industry to understand the scientific, technical and social factors that drive the growth of the Web and enable innovation
devise curricula for the new discipline of Web Science so as to train future generations of Web Scientists
Web Science is interdisciplinary
Web Science is about emergence

- simple principles and protocols (e.g. links) create complex structures - the Web
- simple behaviors (e.g. Blogging) create complex phenomena - the Blogosphere
- we will anticipate new principles and behaviors
Web Science
A Process

- creative innovation
- design and engineering
- the social and the technical
- interpretation and analysis
Web Science
A Process

We can understand after the fact, e.g. the Google effect

Can we predict, anticipate before the fact?
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

research / thought leadership / insight
Web Science
The Blogosphere

values

self expression

creativity

Self Publish

Blog

Design

Self Expression

Blogsite

Trackback

Blogosphere

analyze

Spoofing and splogs

Social

Micro

Tech.

Citation

complexity

research / thought leadership / insight
Web Science - Examples
Wikipedia - Collective Intelligence

• What is its structure?
• How stable is it?
• Why do people contribute?
• What lessons does it offer?
the social web

- social networking sites
- (inter-linked) blogs + comments + aggregators
- community-edited news sites, participatory journalism
- content-sharing sites
- discussion forums, newsgroups
- wikis, Wikipedia
- services that allow sharing of bookmarks/favorites
- ...and mashups of the above services

“democratic”, participatory, conversational

Material from Ciro Cattuto

http://isiosf.isi.it/~cattuto

research / thought leadership / insight
experimental data

- large-scale del.icio.us snapshot from distributed crawl
- from beginning of 2004 up to November 2006
- full hypergraph structure + timestamps

~ 650,000 users
~ $2 \cdot 10^7$ resources
~ $5 \cdot 10^7$ posts
~ $3 \cdot 10^6$ distinct tags
networks of tag co-occurrence

\[ S_{nn}(t) = \frac{1}{k_i} \sum_{j=1}^{k_i} s_t \]

Material from Ciro Cattuto
http://isiosf.isi.it/~cattuto
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
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

Music
- MusiBrainz
- Last.fm
- Last.fm Torch
- TuneRanger

Online Activities
- Flicker Exporter
- LinkedIn
- Google Scholar

Geographic
- Geonames
- OpenCalais
- LinkedMDB

Publications
- ACM
- RAE 2001
- Centre-Uri

General
- DBpedia
- Linked Data
- OSIRIS

Life Sciences
- KEGG
- Gene Ontology
- UniProt

4.5 billion triples
180 million data links

As of March 2009
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.
Web Science
Information Accountability

Issues of security, privacy, trust
New privacy framework
Technical architecture for usage tracking and policy explanation
Legal rules that concentrate on permissible/impermissible use
Analysis of technical/social interplay

research / thought leadership / insight
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

S
ince 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 who depend on the Web - Media, banking and health care - 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 much more than the sum of its pages. Vast emergent properties have arisen that are transforming society. E-mail led to 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, discovered by libraries creating online communities that share everything from comic strips to proper tips.

But few investigators are studying how such emergent properties have actually blossomed, how we might harness them, what new phenomena may be coming or what any of this might mean for humanity. A new branch of science - Web Science - aims to address such issues. The timing for this new venture is ideal: computers were built first, and computer science followed, which subsequently computing science. Web Science was launched as a discipline at the University of Southampton in England; the dream of a Web Science Research Initiative (WSRI) to bring together researchers from 16 of the world's major universities was realized.

This new discipline will model the Web's structure, articulate the architectural principles that have fueled its phenomenal growth, and enable us to create more useful and engaging environments.

Insights Already

Although Web science as a discipline has its roots in earlier research, it has revealed the potential of such work. At the 1990s, for instance, the Web was just beginning to take off. Today, with the explosion of content and the growth of social networking, the Web is a tool that we can use to understand the world around us. And the Web Science Research Initiative is just one example of how this new field can help us to make sense of the world.
Outreach and Thought Leadership

- Refined Research Agenda with Sci Council (MIT Nov 2008)
- Influence on funding agencies – UK EPSRC (Digital Economy), European Commission, NSF?, China?
- Workshops – WebEvolve2008 @WWW2008
- More workshops in the pipeline for 2009 – scholarship on the Web, critical infrastructure, transparency in journalism, e-government II (Washington, October) and lots more to come
- Web Science 2009, Athens, 18-20 March 2009
- Web Science 2010, Rayleigh-Durham
Education

• Doctoral Summer School, joint with the Oxford Internet Institute, July 2008
• Curriculum workshop, September 2008
• Curriculum wiki launched
• Curriculum workshop in Athens @ WebSci’09, more planned
• Doctoral Training Centre @ Southampton
WSRI Affiliation Activities

- WSRI Affiliated Labs (WAL’s)
- Wider network of Web Science research groups
- Curriculum Development
- WSRI Ambassadors and Evangelists
WSRI Affiliated Web Science Labs

- Developing a network of Web Science Labs around the world
- Pursuing a coordinated programme of work
  - Research – annual meeting of research directors
  - Doctoral Summer Schools (21-28 July 2009 @RPI, 2010 @Tsinghua
  - Curriculum Development
  - Technology Transfer
Web Science

why this matters

- the Web matters
- an essential part of humanity
- understanding the Web is a major challenge as big as any other global cause
- www.webscience.org