STI · INNSBRUCK

Semantic Web Services

Web Science

Lecture II – 12th March 2009

Dieter Fensel

Where are we?

STI · INNSBRUCK

#	Date	Title
1	5th March	Introduction
2	12th March	Web Science
3	19th March	Service Science
4	26th March	Web Services (WSDL. SOAP, UDDI, XML)
5	2nd April	Web 2.0 and RESTful services
6	23rd April	WSMO
7	30th April	WSML
8	7th May	WSMX
9	14th May	OWL-S and others
10	28th May	WSMO-Lite, MicroWSMO
11	4th June	SWS Use Cases
12	18th June	seekda: the business point of view
13	25th June	Mobile services
14	2nd July	Exam Preparation

Outline

STI · INNSBRUCK

- What is The Web?
- Web Evolution
- What is Web science?
- Web science process/methodology
- Web science challenges

STI · INNSBRUCK

What is The Web?

The Web

- The **World Wide Web** ("**WWW**" or simply the "**Web**") is a system of interlinked, hypertext documents that runs over the Internet. With a Web browser, a user views Web pages that may contain text, images, and other multimedia and navigates between them using hyperlinks. - wikipedia
- The Web was created around 1990 by Tim Berners-Lee working at CERN in Geneva, Switzerland.

The Web

- A distributed document delivery system implemented using application-level protocols on the Internet
- A tool for collaborative writing and community building
- A framework of protocols that support e-commerce
- A network of co-operating computers interoperating using HTTP and related protocols to form a 'subnet' of the Internet
- A large, cyclical, directed graph made up of Web pages and links

WWW Components

- Structural Components
  - Clients/browsers – to dominant implementations
  - Servers – run on sophisticated hardware
  - Caches – many interesting implementations
  - Internet – the global infrastructure which facilitates data transfer
- Semantic Components
  - Uniform Resource Identifiers (URIs)
  - Hyper Text Transfer Protocol (HTTP)
  - Hyper Text Markup Language (HTML)
    - eXtensible Markup Language (XML)

Uniform Resource Identifiers (URIs)

- Uniform Resource Identifiers (URIs) are used to name/identify resources on the Web
- URIs are pointers to resources to which request methods can be applied to generate potentially different responses
- Resource can reside anywhere on the Internet
- Most popular form of a URI is the Uniform Resource Locator (URL)

Hypertext Transfer Protocol (HTTP)

- Protocol for client/server communication
  - The heart of the Web
  - Very simple request/response protocol
    - Client sends request message, server replies with response message
  - Provide a way to publish and retrieve HTML pages
  - Stateless
  - Relies on URI naming mechanism

HTTP Request Messages

- GET – retrieve document specified by URL
- PUT – store specified document under given URL
- HEAD – retrieve info. about document specified by URL
- OPTIONS – retrieve information about available options
- POST – give information (eg. annotation) to the server
- DELETE – remove document specified by URL
- TRACE – loopback request message
- CONNECT – for use by caches

HTML

- Hyper-Text Markup Language
  - A subset of Standardized General Markup Language (SGML)
  - Facilitates a hyper-media environment
- Documents use elements to “mark up” or identify sections of text for different purposes or display characteristics
- Mark up elements are not seen by the user when page is displayed
- Documents are rendered by browsers

HTML

- HTML markup consists of several types of entities, including: elements, attributes, data types and character references
  - DTD (Document Type Definition)
    - <!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
  - Element (such as document (<html>…</html>), head elements (<title>…</title>)
  - Attribute: HTML
  - Data type: CDATA, URIs, Dates, Link types, language code, color, text string, etc.
  - Character references: for referring to rarely used characters:
    - "水" (in hexadecimal) represents the Chinese character for water

From the World Wide Web to the Semantic Web

- Overabundance of Information
  - Highly scattered and distributed
  - Pressure to search and integrate information
  - Cost for locating relevant information and deriving value from it is prohibitively expensive
- Reduce costs by:
  - Interconnecting workflows and business processes
  - Enable data and service sharing
- Enable collaboration and information sharing
  - Data and services sharing between diverse scientific groups such as genomic and biological sciences, and geosciences
  - Across industry wide consortiums and standards bodies
  - Across ordinary internet users (RSS – Blogs)

From Web 1.0 to 2.0 (updated from O'Reilly)

	Web 1.0	Web 2.0
Platforms	Netscape, Internet Explorer	Google Services, AJAX, Flock
Web Pages	Personal Websites	Blogs
Portals	Content Management Systems	Wikis
Encyclopediæ	Britannica Online	Wikipedia
Talk	Netmeeting	Skype, Asterisk
Knowledge	Directories, Taxonomies	Tagging, Folksonomies
Referencing	Stickiness	Syndication
Content	Akamai	BitTorrent, P2P
Events	Evite	Upcoming.org

From Web 1.0 to Semantic Web

Web 1.0	Web 2.0	Semantic Web
Personal Websites	Blogs	**Semantic Blogs:** semiBlog, Haystack, Semblog, Structured Blogging
Content Management Systems, Britannica Online	Wikis, Wikipedia	**Semantic Wikis:** Semantic MediaWiki, SemperWiki, Platypus, dbpedia, Rhizome
Altavista, Google	Google Personalised, DumbFind, Hakia	**Semantic Search:** SWSE, Swoogle, Intellidimension
CiteSeer, Project Gutenberg	Google Scholar, Book Search	**Semantic Digital Libraries:** JeromeDL, BRICKS, Longwell
Message Boards	Community Portals	**Semantic Forums and Community Portals:** SIOC, OpenLink DataSpaces
Buddy Lists, Address Books	Online Social Networks	**Semantic Social Networks:** FOAF, PeopleAggregator
…	…	**Semantic Social Information Spaces:** Nepomuk, Gnowsis

Web Evolution

- Traditional Web (Web1.0)
  - Normal User: browsing
  - Communication style: one-direction communication (e.g. reading a book)
  - Data: web data (string and syntactic format)
  - Data contributor: webmaster or experienced user
  - How to add data: compose HTML pages
- Social Web (Web2.0)
  - Normal User: browsing + publishing and organizing web data
  - Communication style: human-human (sharing)
  - Data: web data + tags
  - Data contributor: normal user – revolution!
  - How to add data: tagging
- Semantic Web
  - Normal User: interacting (human-machine)
  - Communication style: human←→machine
  - Data: web data + tags + metadata (in SW Language)
  - Data contributor: normal user, machine
  - How to add data: machine generate or user publish

What is Web Science?

Which science explains the Web?

- Is it:
  - Computer Science
  - Web Engineering
  - Artificial Inteligence
  - Economics
  - Low
  - Sociology
  - …
  - ?

Web Science definition

A new science that focuses on how huge decentralized Web systems work.

"The Web isn't about what you can do with computers. It's people and, yes, they are connected by computers. But computer science, as the study of what happens in a computer, doesn't tell you about what happens on the Web."
Tim Berners-Lee

"A new field of science that involves a multi-disciplinary study and inquiry for the understanding of the Web and its relationships to us"
Bebo White, SLAC, Stanford University

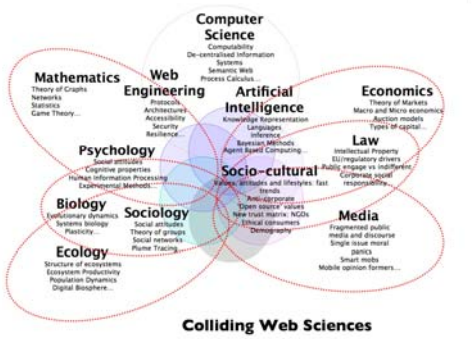
Shift from how a single computer works to how huge decentralized Web systems work

Endorsements for Web Science

"Web science represents a pretty big next step in the evolution of information. This kind of research likely to have a lot of influence on the next generation of researchers, scientists and, most importantly, the next generation of entrepreneurs who will build new companies from this."
Eric E. Schmidt, CEO Google

"Web science research is a prerequisite to designing and building the kinds of complex, human-oriented systems that we are after in services science."
Irving Wladawsky-Berger, IBM

Web science – multi-disciplinary approach



http://webscience.org/images/collide.jpg

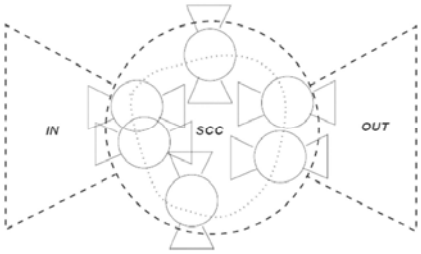
Web science – multi-disciplinary approach

- To understand the Web we need:
  - To understand society
  - To understand computer engineering
  - To understand users
  - To understand economics
  - To understand network effects at high scale
- Different solutions may be required
  - New systems
  - New interfaces
  - New tools
  - New protocols

The Goals of Web Science

- To understand what the Web is
- To engineer the Web's future
- To ensure the Web's social benefit

Shape of the Web



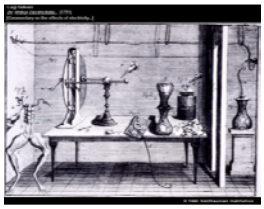
What Could Scientific Theories for the Web Look Like?

- Some simple examples:
  - Every page on the Web can be reached by following less than 10 links
  - The average number of words per search query is greater than 3
  - Web page download times follow a lognormal distribution function (Huberman)
  - The Web is a "scale-free" graph
- Can these statements be easily validated? Are they good theories? What constitutes good theories about the Web?

http://webcast.bibalex.org/Presentations/Bebo91108.ppt

Food For Thought

Electricity : 1800



Electricity Now

$$\oint \vec{E} \cdot d\vec{A} = \frac{q}{\varepsilon_0}$$

$$\oint \vec{B} \cdot d\vec{A} = 0$$

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_B}{dt}$$

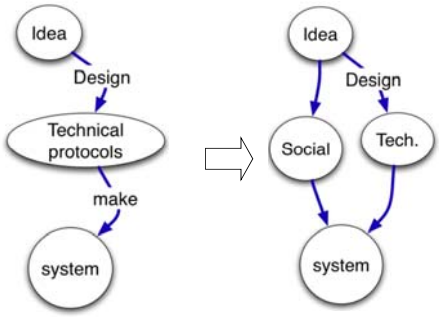
$$\oint \vec{B} \cdot d\vec{s} = \mu_0 i + \frac{1}{c^2}\frac{\partial}{\partial t}\int \vec{E} \cdot d\vec{A}$$

What are the analogies for Web Science and Design? Is our understanding of the Web like that of 1800 electricity?

http://webcast.bibalex.org/Presentations/Bebo91108.ppt

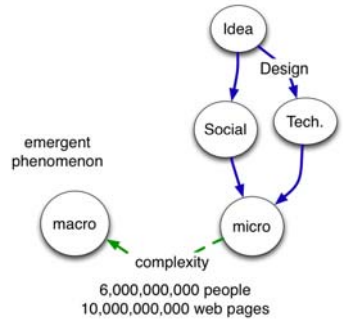
Web Science process/methodology

Social convenience when designing a system

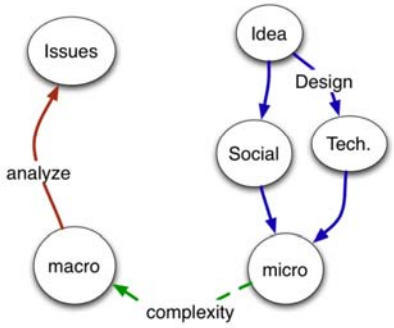


Slides 28-36 adapted from: Tim Berners Lee – The Two Magics of Web Science
http://www.w3.org/2007/Talks/0509-www-keynote-tbl/

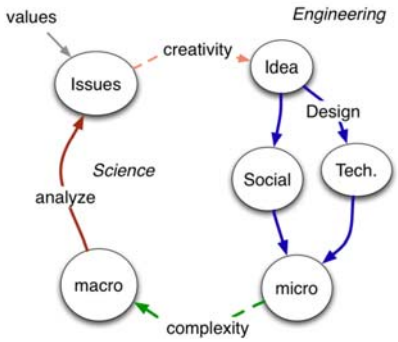
Macroscopic effect



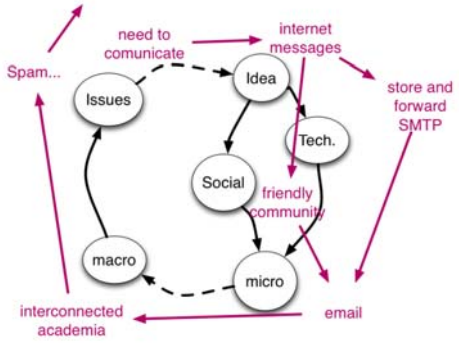
Analyze the result



Science and Engineering in a Cycle

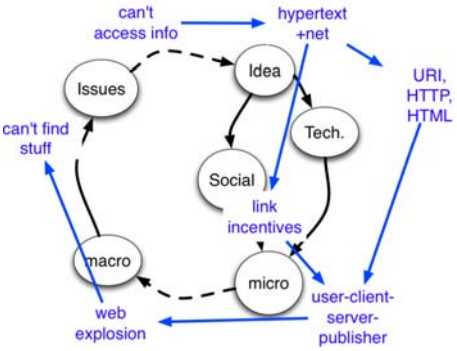


Example: E-Mail

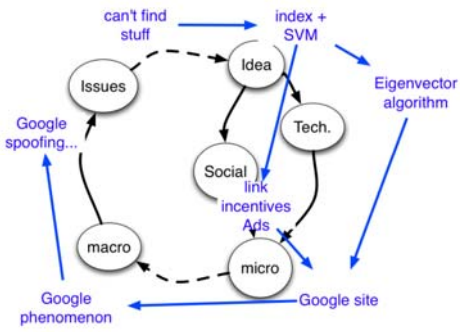


Example: The Web as envisioned

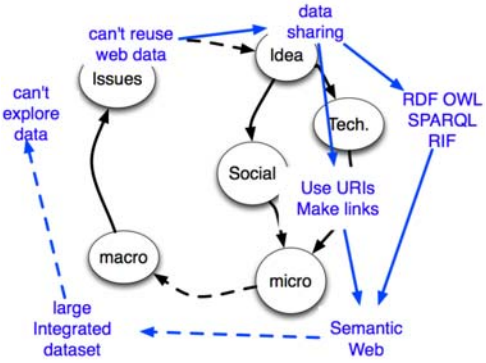
Example: WWW



Example: building on WWW: Google



Example: Semantic Web



Web Science challenges

Challenges

- User Interfaces
  - generality, explore new dimensions, analyze, visualize, ….
- Information Policy
  - Identity, privacy, transparency, trust, …
- Resilience
  - Internet breakage, Web breakage, Spam, phishing, etc.

Challenges (contd.)

- Collective quality assessment
  - new web-enables processes
- New devices
  - capabilities, dimensions, …
- Collective creativity

Web Science – why this matters

- Web is an evolving system and an essential part of humanity
- Understanding the Web is a major challenge as big as any other global cause

Relevant links

- Web Science Research Initiative http://webscience.org/
- Berners-Lee, Tim, Hall, Wendy, Hendler, James, Shadbolt, Nigel, Weitzner, Danny (2006): Creating a science of the Web. http://eprints.ecs.soton.ac.uk/12615/
- Berners-Lee, Tim, Hall, Wendy, Hendler, James, O'Hara, Kieron, Shadbolt, Nigel, Weitzner, Danny (2006): A Framework for Web Science. http://eprints.ecs.soton.ac.uk/13347/
- Shadbolt, Nigel. Web Science Research Initiative Seminar November 2008 http://www.ecs.soton.ac.uk/podcasts/video.php?id=153

Next Lecture

#	Date	Title
1	5th March	Introduction
2	12th March	Web Science
3	**19th March**	**Service Science**
4	26th March	Web Services (WSDL. SOAP, UDDI, XML)
5	2nd April	Web 2.0 and RESTful services
6	23rd April	WSMO
7	30th April	WSML
8	7th May	WSMX
9	14th May	OWL-S and others
10	28th May	WSMO-Lite, MicroWSMO
11	4th June	SWS Use Cases
12	18th June	seekda: the business point of view
13	25th June	Mobile services
14	2nd July	Exam Preparation

Questions?