WEBS6201 Foundations of Web Science

Week 1
Introduction
Mark Weal
Course Structure

• Two Lectures each week
  – Monday 2pm, B54/10031
  – Friday 12pm, B58/1025

• One seminar each week
  – Friday 3pm, B2/3041

• Assessments
  – Coursework 1 – platform biography
    • (due 30th Nov, worth 50%)
  – Coursework 2 – web science report
    • (due 14th Dec, worth 50%)
<table>
<thead>
<tr>
<th>Week Date</th>
<th>Monday 14:00</th>
<th>Friday 12:00</th>
<th>Friday 15:00</th>
<th>Friday 16:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1/10</td>
<td>Module Introduction</td>
<td>Technology and Modernity</td>
<td>(AOIR) Seminar Skills</td>
<td></td>
</tr>
<tr>
<td>2 8/10</td>
<td>Digital Data Infrastructures</td>
<td>Coursework Discussion</td>
<td>Reading/Discussion</td>
<td></td>
</tr>
<tr>
<td>3 15/10</td>
<td>Social Construction of Technology</td>
<td>Value and the Web</td>
<td>Reading/Discussion</td>
<td></td>
</tr>
<tr>
<td>4 22/10</td>
<td>The Web and Data</td>
<td>Privacy and the Web</td>
<td>Reading/Discussion</td>
<td></td>
</tr>
<tr>
<td>5 29/10</td>
<td>Social Networks and Actors</td>
<td>Gender, Feminism and the Cyborg</td>
<td>Reading/Discussion</td>
<td></td>
</tr>
<tr>
<td>7 12/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 19/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 26/11</td>
<td></td>
<td></td>
<td>CW1 Deadline</td>
<td></td>
</tr>
<tr>
<td>10 3/12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 10/12</td>
<td>coursework tutorial</td>
<td>no session</td>
<td>no session</td>
<td>CW2 Deadline</td>
</tr>
<tr>
<td>XMAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 7/1</td>
<td>No session</td>
<td></td>
<td>No session</td>
<td></td>
</tr>
</tbody>
</table>

**Legend**
- Introduction and process
- Sociotechnical Theories
- Reading Seminars
- Guest Lecture Topics

**WEB6201**
Where did the Web come from?

Dr Mark Weal

(adapted from slides by Professor Les Carr)
Where’s This, and What Has it to Do with The Web?
What’s This?
Are we all going to die next Wednesday?

Two nightmare scenarios, two ends of the world. In the first, there is little warning. For maybe a month there would be no sign that life was about to come to an abrupt and nasty end for all living things on Earth.

Then, earthquakes would start unexpectedly, alerting geologists that something terrible, unimaginable, was amiss. After a few days, these seismic disturbances would reach catastrophic proportions.

Cities would be levelled, the oceans would rise and wash in a series of mega-tsunamis that would attack the world’s coasts, killing millions.

Is the end of the world nigh? Doom-mongers fear the consequences of scientists replicating the Big Bang

The fact that the earthquakes were striking randomly, not along well-known geological faultlines, would be proof that something was going wrong in the world. As the days passed, the number of earthquake hits would increase, reaching a crescendo when the world’s coasts would be subjected to a series of tsunamis of gigantic size.

Remember These Headlines?

- Saved from the brink of death by bison and eagles among made a remarkable comeback.
- China lifts ban on Facebo living in a 17 square mile.
- Going for gold too? Samsung Galaxy S4 hand Apple launched 'champ'
- The map that shows who from air pollution (and it's in Asia and Eastern Europe).
- Have scientists acciden lightsaber? New form of energy has only ever been seen.
- Scousers have the 'least trustworthy' accent - while friendliest.
- Mystery of the 2,500-year-old found in Bulgaria that were buried standing up.
- Retail therapy really shop could be triggered by a song.
- 'Jesus was the first proponent of Twitter because Christ's sermon meaning'.
- How to put your finger on a patting someone on the cheek is a simple equation: Science can predict historical trend.
- Hate laundry? World’s first make it a thing of the past.
- out £650
- The clues found near that suggest Jews and worshipped together.
- How size matters even in Fossils of penis bones from the future.
Birth of the Web

• From an underground nuclear bunker on the Swiss border, something did escape
• In 1989 the Web took over the academia, industry and the rest of the world
Society is Diverse. One Web Fits All?

The development of society as a whole (nuanced and structured and refined) is inextricably related to the technology of information provision, consumption and dissemination (e.g. writing, reading, printing, education). Different parts of society have different objectives and hence incompatible Web requirements, e.g. openness, security, transparency, privacy.
Expansion of the Web

• The Web spread the conditions of its initial creation throughout the whole of society as it underwent an initial inflationary phase.

• The academy has
  – government patronage
  – large-scale co-operation
  – sharing of intellectual property with others

This is not like other areas of society – e.g. media, industry, commerce.
## Historic Attempts at Webs

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>System</th>
<th>Scope</th>
<th>Real</th>
<th>Date</th>
<th>Important Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance / Press</td>
<td>Reuters</td>
<td>Professional, centralised</td>
<td>✔</td>
<td>1850</td>
<td>News &amp; stock information (originally carrier pigeon and subsequently telegraph)</td>
</tr>
<tr>
<td>Private Institution</td>
<td>Mundaneum</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1920</td>
<td>Based on indexing technology (the library card)</td>
</tr>
<tr>
<td>Military</td>
<td>Memex</td>
<td>Scholarly, individual,</td>
<td>✗</td>
<td>1945</td>
<td>Aimed at Scientists and Technologists in WWII</td>
</tr>
<tr>
<td>Media</td>
<td>Xanadu</td>
<td>Public, decentralised</td>
<td>✗</td>
<td>1960</td>
<td>Focused on DRM, reuse and writing for “creatives”</td>
</tr>
<tr>
<td>Media</td>
<td>CEEFAX</td>
<td>Public, national,</td>
<td>✔</td>
<td>1970</td>
<td>Broadcast, linked, not participatory</td>
</tr>
<tr>
<td>Government</td>
<td>Minitel</td>
<td>Public, national,</td>
<td>✔</td>
<td>1980</td>
<td>Commercial services and information</td>
</tr>
<tr>
<td>Academy (CS &amp; HEP)</td>
<td>FTP / Archie /</td>
<td>Public, decentralised</td>
<td>✔</td>
<td>1985</td>
<td>Download resources (papers, reports) to hard drives and print them on LaserWriters.</td>
</tr>
<tr>
<td></td>
<td>Anarchie</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commerce</td>
<td>Hypercard, HyperTIES</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1988</td>
<td>Personal applications, sometimes tied to multimedia resources on CDROMs / video disks</td>
</tr>
<tr>
<td>Academy (HEP)</td>
<td>WWW</td>
<td>Public, global,</td>
<td>✔</td>
<td>1990</td>
<td>Universal naming, linking, interoperability, participative. However no writing, no indexing in public version.</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>Microcosm</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Sophisticated linking and openness for personal information stores</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>HyperG</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Extension of Web for with support for writing, indexing and consistency management.</td>
</tr>
<tr>
<td>Commerce</td>
<td>AOL, CompuServ</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Dialup walled garden access to email, forums, chat rooms and information resources</td>
</tr>
</tbody>
</table>
Paul Otlet, 1868-1944

- Belgian lawyer
- Introduced US 3"x5" library card to Europe
- Traité de Documentation (1934)
  - the systematic organisation of all knowledge and thought

**Mundanaeum**: 15 million index card bibliographic index, 1 million documents and images, classified and searchable. *Query became part of the bibliographic record. Content was interlinked.*
Available Technology

The library card, used for the mundaneum

- Encyclopaedias of the past sufficed for the needs of a cultivated minority
  - universal education was unthought of
  - gigantic increase in recorded knowledge
  - more gigantic growth in the numbers of human beings requiring accurate and easily accessible information
- Discontent with the role of universities and libraries in the intellectual life of mankind
- Universities multiply but do not enlarge their scope
  - thought & knowledge organization of the world
- **No obstacle** to the creation of an efficient index to *all* human knowledge, ideas and achievement
Available Technology

The microfilm (modern example), proposed for the World Brain
Memex

- Vannevar Bush was never directly involved with the creation or development of the Internet. He died before the creation of the World Wide Web. Yet many consider Bush to be the Godfather of our wired age often making reference to his 1945 essay, “As We May Think”
Memex

• Associative trails
• Microfilm based
• the ability to connect, annotate, and share both published works and personal trails would profoundly change the process by which the "world's record" is created and used.
Vint Cerf

• A “father of” the Internet (1970s & 1980s)
  – A network of computers and routers and cables and clients and servers. In fact, lots of networks all networked together. An inter-net. A worldwide network of computers sending and receiving messages

FTP (file transfer protocol)
• Upload a document (plain text or PostScript printer instructions) to a server
• Download document from server to workstation hard disk
• Send it to printer
Available Technology

The network that forms a (very very very small) part of the Internet
Ceefax

- Started in 1974
- Started off as BEEBFAX
- Were originally trying to solve the problem of subtitles for the deaf
- Alternative system called ORACLE
- Evolved into the World System Teletext
Hypercard

- Came with the Apple Macintosh.
- Card based metaphor with database support.
- Allowed you to build applications, ‘app development!’
Microcosm

- Open Hypermedia system
- Links were kept separately from the documents
- Could add links to read-only documents (originally laserdiscs)
- Could put links in images, video, etc.
- Company spun out as Active Navigation and technology evolved to become more Web based
- Generic links perhaps the most novel and interesting feature
What the Web Was

- Tim Berners-Lee, had one big vision, with 3 big ideas
  - **URI** uniform resource identifiers (or URLs locators) – everything has an identifier, or name
  - **HTTP** – once you know the ID, a copy of anything can be obtained by a conversation between a client computer and a server on the Internet
  - **HTML** – the format of the information obtained doesn’t just handle the way that the information should be laid out on the printed page, but the way that information is laid out across the Internet. A document with embedded links that readers can interact with.
What is Hypertext?

The reaction of the hypertext research community to the World Wide Web is like finding out that you have a fully grown child.

And it’s a delinquent.

What the Web Became

• The first Web browser & server were opened up to CERN and the World to use.

• And then the world took it on, and hundreds of people in different sectors of society, gradually turned this “online document library” into...

  – Online shopping
  – Internet TV
  – Social media
  – Internet porn
  – Web journalism

  – Search engines
  – E-health services
  – Streaming music services
  – Computer dating
  – Internet piracy...
Complexity

- The Web has many aspects
- It is not just algorithms, networks, graphs, media ...
- It is all of these and more
Examples
## Historic Attempts at Webs

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>System</th>
<th>Scope</th>
<th>Real</th>
<th>Date</th>
<th>Important Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance / Press</td>
<td>Reuters</td>
<td>Professional, centralised</td>
<td>✔</td>
<td>1850</td>
<td>News &amp; stock information (originally carrier pigeon and subsequently telegraph)</td>
</tr>
<tr>
<td>Private Institution</td>
<td>Mundaneum</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1920</td>
<td>Based on indexing technology (the library card)</td>
</tr>
<tr>
<td>Military</td>
<td>Memex</td>
<td>Scholarly, individual, centralised</td>
<td>❌</td>
<td>1945</td>
<td>Aimed at Scientists and Technologists in WWII</td>
</tr>
<tr>
<td>Media</td>
<td>Xanadu</td>
<td>Public, decentralised</td>
<td>❌</td>
<td>1960</td>
<td>Focused on DRM, reuse and writing for “creatives”</td>
</tr>
<tr>
<td>Media</td>
<td>CEEFAX</td>
<td>Public, national, centralised</td>
<td>✔</td>
<td>1970</td>
<td>Broadcast, linked, not participatory</td>
</tr>
<tr>
<td>Government</td>
<td>Minitel</td>
<td>Public, national, centralised</td>
<td>✔</td>
<td>1980</td>
<td>Commercial services and information</td>
</tr>
<tr>
<td>Academy (CS &amp; HEP)</td>
<td>FTP / Archie / Anarchie</td>
<td>Public, decentralised</td>
<td>✔</td>
<td>1985</td>
<td>Download resources (papers, reports) to hard drives and print them on LaserWriters.</td>
</tr>
<tr>
<td>Commerce</td>
<td>Hypercard, HyperTIES</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1988</td>
<td>Personal applications, sometimes tied to multimedia resources on CDROMs / video disks</td>
</tr>
<tr>
<td>Academy (HEP)</td>
<td>WWW</td>
<td>Public, global, decentralised</td>
<td>✔</td>
<td>1990</td>
<td>Universal naming, linking, interoperability, participative. However no writing, no indexing in public version.</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>Microcosm</td>
<td>Private, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Sophisticated linking and openness for personal information stores</td>
</tr>
<tr>
<td>Academy (CS)</td>
<td>HyperG</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Extension of Web for with support for writing, indexing and consistency management.</td>
</tr>
<tr>
<td>Commerce</td>
<td>AOL, CompuServ</td>
<td>Public, centralised</td>
<td>✔</td>
<td>1990</td>
<td>Dialup walled garden access to email, forums, chat rooms and information resources</td>
</tr>
</tbody>
</table>
Any Questions?