

Lecture Overview

CS6120 Lecture 10: Web 2.0

A. O'Riordan, 12/3/08

- Web 2.0
 - Principles
 - Examples
 - Software Development Issues
 - Technologies - Tagging, Remixes, Rich Clients
 - Economics
 - Criticisms
 - Towards Web 3.0
- Blogging and Syndication
 - RSS and Atom
- Social Networking Services

1

2

Origin of Web 2.0

- dot-com bubble in the fall of 2001 marked a turning point for the web; shakeout; e-commerce didn't collapse – actually it greatly increased
- The concept of "Web 2.0" began with a brainstorming session between O'Reilly and MediaLive International; Web 2.0 Conference in October 2004
- Disagreement about just what Web 2.0 means; many features of Web 2.0 trend pre-date '00s
- Buzzword intensive; Companies/products of first dot-com boom now dubbed Web 1.0; Some people now talking about Web 3.0

3

Sense of Web 2.0 by example [O'Reilly]

Web 1.0 -> Web 2.0

DoubleClick --> Google AdSense
Ofoto --> Flickr
Akamai --> BitTorrent
mp3.com --> Napster
Britannica Online --> Wikipedia
personal websites --> blogging
domain name speculation --> search engine optimization
page views --> cost per click
screen scraping --> web services
publishing --> participation
content management systems --> wikis
directories (taxonomy) --> tagging ("folksonomy")
Stickiness --> syndication

4

Harnessing Large User Base [O'Reilly]

- Peer-to-peer – decentralization
 - Every downloader becomes a server, e.g. Napster, BitTorrent
 - More popular data is, the faster it will be served
- Product based on link structure (e.g. Google), collective activity (e.g. eBay), added value of user input/comment (e.g. Amazon)
- User Engagement – reviews, edits, recommendations, tagging, invitations to participate
 - Low-cost viral marketing
 - Collaborative categorization (tagging)

9

Technologies I

- Rich Internet application techniques, e.g. Ajax, Adobe Flash, Adobe Flex, Microsoft Silverlight
- Semantically valid XHTML/HTML markup and Cascading Style Sheets to aid in the separation of presentation and content
- Microformats extending pages with additional semantics e.g. XFN (social networking sites), hCard (contacts)
- Folksonomies, e.g., by means of tags or tagclouds

11

Architecture of Participation [Bricklin]

- Three ways to build a large database [from the Cornoponia of Commons, Dan Bricklin]
 1. Pay people – experts or employees construct/ organize data
 - e.g. Yahoo! Directory; Encarta
 - limited visitors to viewing and whose content only the site's owner could modify
 2. Recruit volunteers – inspired by open-source software ethic
 - e.g. Open Directory Project
 3. Users do it – thousands of informal participants
 - e.g. P2P File Sharing

10

Technologies II

- APIs
 - REST (Representational State Transfer)
 - SOAP (no longer an acronym)
 - JSON (JavaScript Object Notation)
- Syndication, aggregation and notification of data in RSS/Atom feeds
- Mashups, merging content from different sources, client- and server-side
- Weblog-publishing tools ... Blogger, WordPress, Userland, MovableType, TypePad, LiveJournal
- wiki or forum software, etc., to support user-generated content

12

Data Driven

- Enhance the data - Amazon a pioneer of this
 - Original database came from ISBN registry provider R.R. Bowker
 - Adding publisher-supplied data such as cover images, table of contents, index, and sample material
 - Even more importantly, they harnessed their users to annotate
 - Data now unique!
- Google Maps - user annotations on satellite images
- But countervailing Free Data movement - Wikipedia, Creative Commons

13

Remixes/Mashups

- Many based on Google Maps, e.g. Housing Maps — <http://housingmaps.com/>
- Other examples – Music/Video Remixing



14

Tagging

- Short human-readable labels similar in form and function to "categories", but differ in that they typically are not organized
 - Allow users to generate content and classify that content in their own way - reflects the vocabulary of users
 - Metadata - data about data
 - Renaissance of Manual Indexing? [Vos, J., *Proc. Intrnl. Sym. of Info. Sci.* p234–254, 2007.]
- Folksonomy - collaboratively creating and managing tags
 - term attributed to Thomas Vander Wal (www.vanderwal.net)
 - flat system – as opposed to (Yahoo!) hierarchy
- Problems: inaccurate tagging; synonym control

15

Tag Cloud

- Tag cloud is a visual depiction of user-generated tags used to describe the content of web site/page – popularized by Flickr
- Example (Tag Cloud of Candidate Speech in 2008 Democratic Debate, Pollster.com BETA)

Senator Clinton

access administration america american attack believe brian
 care chance college congress corporate course determine doing fact government
 guns having health hope important invisible Iraq Iraq leader listen middle
 mistakes money obviously pay policy president ready
 republican responsibility security senator serious something support system
 though tried war willing working worried years

16

Example Sites that Use Tagging

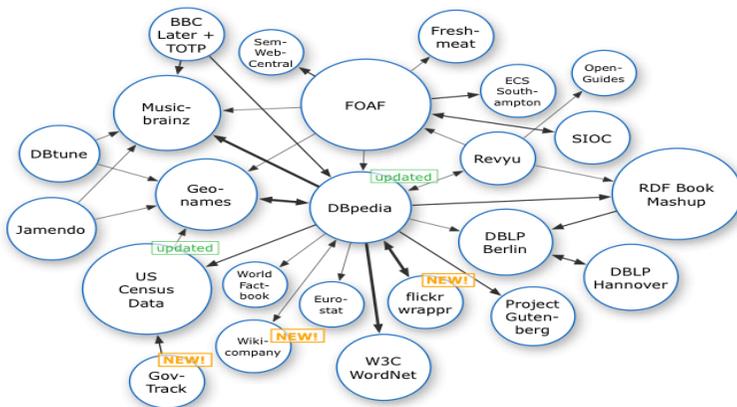
- [Flickr](#), photo sharing 
- [del.icio.us](#), the collaborative bookmarking network 
- [Digg](#), a technology news site 
- [43things](#) social networking 
- [StumbleUpon](#) discover and rate webpages 

17

- “I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers.” [Berners-Lee and Fischetti, *Weaving the Web*, 1999]
- Questions of feasibility
- Resource Description Framework (<http://www.w3.org/RDF/>)
 - metadata model
 - abstract model with several serialization formats
 - Knowledge representation in Ontology
- The Web Ontology Language (OWL)

18

Example: Linking Open Data Project



19

Lightweight Development Approach

- “End of the Software Release Cycle” [O’Reilly]
- Operations become more important
- Monitoring of user behavior
- Popularity of scripting languages, e.g. Perl, Ruby, Python, PHP and rich client software, Ajax
- Open source development practices; Agile development process
- Sites with millions of users may be expected to bear a “Beta” logo for years

20

Rich Interaction and Services for Multiple Devices

Rich Interaction

- Web apps such as Gmail now have rich user interfaces and PC-equivalent interactivity
- Ajax a key technology (XMLHttpRequest)
- Data interchange formats (XML and XSLT)

Multiple Devices

- iTunes is the exemplar
 - User's handheld device to a massive web back-end, with the PC acting as a local cache and control station
- Podcasting

21

Economics of Web 2.0

- Products and services for free or very cheap; added value
- Development software usually free and open source, e.g. LAMP stack; "Release early, release often"; Agile development process
- Low startup and operations costs
- Internet based distribution, sales, support, marketing, recruiting etc.
- Businesses can locate anywhere (as long as broadband available)
- Customers region specific – e.g. Orkut, big in Brazil, not in US

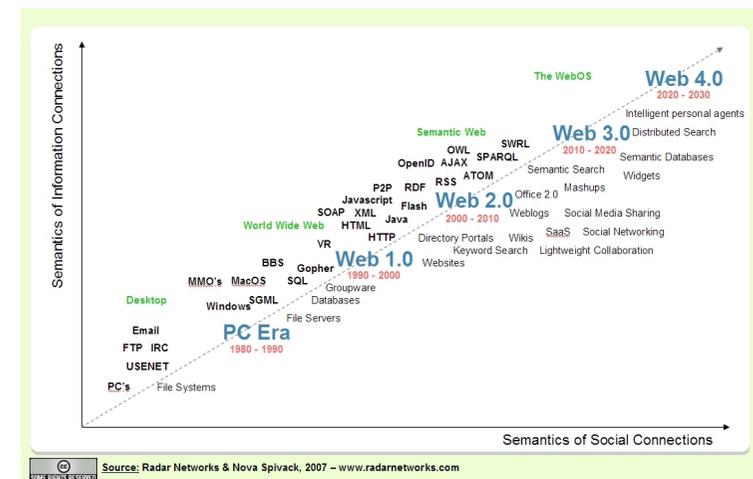
22

Criticism of Web 2.0

- Nothing new – many trends predate 2004, e.g. Web as Platform, Amazon user reviews, workflow and collaborative software as in Lotus Notes
- "nobody really knows what it means" [Tim Berners-Lee, quoted in IBM developerWorks, 2006]
- "Bubble 2.0" [Economist, 2006]
- Web 2.0 trademarked for live events – CMP Media sent a cease-and-desist demand to the Irish non-profit organization IT@Cork on in 2006 but retracted it two days later
- Users who don't contribute

23

Web 1.0 -> Web 2.0 -> Web 3.0



24

What about Web 3.0 then?

- Web 3.0 (2007/10 - ?)
- Web data in reusable and remotely queryable formats e.g. XML and XQuery
- AI and greater interactivity – monitoring user behaviour; data mining
- More semantics – Semantic Web; reasoning about content
- 3D, e.g. Second Life
- scalable vector graphics [Tim Berners Lee, quoted in *International Herald Tribune, 2006*]
- ubiquitous connectivity
- open technologies, open APIs and protocols, open data
- Who knows?

25

Blogging and Syndication

- Blog (web log) - more than a personal home page in diary format
- Enable users to browse or to subscribe to it with notifications
 - “incremental Web”; “live web”
- Enable talk/conversations about each post; 2-way;
 - link directly to a particular post; “permalinks” should remain unchanged indefinitely –e.g. # symbol
- Can be conceptualized as giant peer-to-peer bulletin board (“blogosphere”)
- Trackbacks – step towards two-way links
 - Remember HTML only has one-way links

26

The Blogosphere [CACM]

From articles in Communications of the ACM, Dec 2004, Andrew Rosenbloom editor

- “Mainstream personal communication like email” - tens of millions of bloggers ???
- “Self-published online journals ... bypass media gatekeepers”
- “tight-knit active communities”
- “rich and complex social environment”
- “quirky highly personal”
- “global phenomenon”
- “community forum”
- “shape democracy”

27

Example Trend in 2007: Micro-blogging

- Form of blogging with brief messages
- Examples: Twitter (launched Oct 2006), Pownce (2008), Jaiku – Google owned (2007)

The Twitter logo, consisting of the word "twitter" in a lowercase, blue, sans-serif font.

- Twitter allows users to send updates (or tweets; text-based posts, up to 140 characters long) to Twitter website
- Via SMS, IM, or a third-party apps such as Twiterrific

28

Syndication

- Most periodically updated web sites have syndication files called feeds
 - Orange button for RSS feed 
- XML-based feed formats allow users to get updated site content in their favorite reader software (web-based or not web-based)
- Used for portals, news, blogs, podcasting (streaming audio) and community-based sites, e.g. del.icio.us
- Share or aggregate content from multiple sources

29

Syndication Technologies

- RSS was born in 1997 Userland's "Really Simple Syndication" for blog updates, and Netscape's "Rich Site Summary"
- Really Simple Syndication (RSS) -is a family of feed formats used to publish frequently updated content
 - File format .xml or .rss
 - Really Simple Syndication (RSS 2.0), RDF Site Summary (RSS 1.0 and RSS 0.90), Rich Site Summary (RSS 0.91)
 - There are incompatibilities between the different versions
 - Some versions support podcasting
 - Several BitTorrent-based peer-to-peer applications also support RSS
- Atom – alternative to RSS designed to overview shortcoming

30

RSS - Some detail

- The RSS 2.0 specification copyright Harvard University and is frozen. No significant changes can be made (although the specification is under a Creative Commons licence)
- RSS APIs - MetaWeblog and Blogger APIs
- Libraries – FeedParser and Rome
- RSS 2.0 requires feed-level title, link, and description
- RSS 2.0 can contain plain text or escaped HTML
- RSS 2.0 is not in an XML namespace
- RSS no schema in specification
- does not specify the handling of relative URI references
- RSS autodiscovery – application/rss+xml MIME type in IE7 (not standardized)

31

Atom

- Adopted as IETF Proposed Standard RFC 4287 in 2005 (see <http://atompub.org/rfc4287.html>)
- Feeds are composed of a number of items, known as entries
- An Atom Feed Document is a representation of an Atom feed, including metadata about the feed, and some or all of the entries associated with it
- Atom Elements - Container Elements (atom:feed atom:entry atom:content) and Metadata elements (atom:author atom:category atom:contributor atom:generator atom:icon atom:id atom:link atom:logo atom:published atom:rights atom:source atom:subtitle atom:summary atom:title atom:updated)
- Atom uses IRIs [\[RFC3987\]](#); every URI [\[RFC3986\]](#) is also an IRI

32

Atom feed Example – single entry

```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <title>Example Feed</title>
  <link href="http://example.org/" />
  <updated>2003-12-13T18:30:02Z</updated>
  <author>
    <name>John Doe</name>
  </author>
  <id>urn:uuid:60a76c80-d399-11d9-b93C-0003939e0af6</id>
  <entry>
    <title>Atom-Powered Robots Run Amok</title>
    <link href="http://example.org/2003/12/13/atom03" />
    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
    <updated>2003-12-13T18:30:02Z</updated>
    <summary>Some text.</summary>
  </entry>
</feed>
```

33

Research Example I

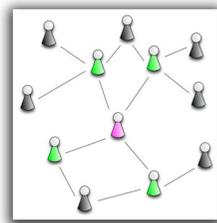
- Feed Aggregator Used for Meta-Search (OpenSearch API <http://www.opensearch.org>)



34

Social Network Services

- Social networks are explicit representations of the relationships between individuals and groups in a community
- “Friend of a Friend” (FOAF)
 - Six Degrees of Separation
 - Six Degrees of Kevin Bacon
 - Erdős number
- Social Network Services exploit this phenomenon, e.g. MySpace
- Issues: Privacy, Security, Social Impact



35

Social Networking Sites with 30M+ Users

[various sources, March, 2008]

Rank	Site	Popular here	Approx. # Million users	Focus
1.	MySpace	Worldwide	110	Music/General
2.	Facebook	English-speaking	98	General
3.	Habbo	South America	86	Chatrooms
4.	orkut	Brazil (and India)	67	General
5.	Friendster	Southeast Asia	58	General
6.	Live Spaces	Worldwide	40	Blogging
7.	Classmates	USA	40	Colleagues
8.	Xanga	USA	40	Blogging
9.	Flixster	Worldwide	36	Movies
10.	Netlog	Europe	31	General
11.	Tagged.com	Worldwide	30	General
14.	Bebo	UK, Ire and NZ	21	General

36