Web 2.0, Tagging, Multimedia, Folksonomies, Lecture, Important, Must Attend, …

Martin Halvey

Lecture Overview

- Introduction to Web 2.0
- Overview of Tagging Systems
  - Overview of tagging
  - Design and attributes
  - Case Studies
- Other Web 2.0 Technologies
- Conclusions

Web 2.0 Definition

"Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform"

Tim O'Reilly (2006-12-10). Web 2.0 Compact Definition: Trying Again

Web 2.0 Definition

"An idea in people's heads rather than a reality. It's actually an idea that the reciprocity between the user and the provider is what's emphasised. In other words, genuine interactivity if you like, simply because people can upload as well as download"

Stephen Fry: Web 2.0

Web 2.0 Definition

"Participatory Web"

Bart Decrem (2006-06-13). Introducing Flock Beta 1. Flock official blog

Web 2.0

"piece of jargon"

"nobody really knows what it means"

"if Web 2.0 for you is blogs and wikis, then that is people to people. But that was what the Web was supposed to be all along"

-Tim Berners-Lee

Web 2.0

- Allow users to do more than retrieve information
- "Network as platform" computing, allowing users to run software applications entirely through a browser
- In contrast to systems which categorise users into roles with varying degrees of functionality
- Re-use of existing technologies, some new

Web 2.0 Feature/Techniques

- Rich Internet Application techniques, AJAX
- Semantically valid XHTML and HTML
- REST and/or XML- and/or JSON-based APIs
- Cascading Style Sheets
- RSS or Atom feeds
- Mash Ups
- Weblog publishing
- Wiki or Forum
- Social Networking

Tagging

- A **tag** is a (relevant) keyword or term associated with or assigned to a piece of information (e.g. a picture, article, or video clip), thus describing the item and enabling keyword-based classification of information.
- Usually chosen informally and personally by item author/creator or by its consumer/viewers/community

Folksonomies

- A **folksonomy** is a user generated taxonomy used to categorize and retrieve web content such as Web pages, photographs and Web links, using open-ended labels called tags
- Intended to make a body of information increasingly easy to search, discover, and navigate over time
- Normally used online but they can arise in a number of other contexts

Folksonomies

- Well-developed folksonomy is ideally accessible as a shared vocabulary that is both created by, and familiar to, its primary users
- Folksonomy tools are not part of the WWW protocols
- Arise where special provisions are made
- Particularly useful when no other text is available

Folksonomies and Semantic Web

- The **Semantic Web** is an evolving extension of the WWW in which content is expressed not only in a format that can be read and used by automated tools, as well as natural language
- Folksonomies can be used in conjunction with semantic web technologies to provide rich descriptions, but not quite yet.
- However metadata from folksonomies is not consistent or reliable

Bridging the Semantic Gap

- The difference between low-level data representation of multimedia and the higher level concepts users associate with the same multimedia
- Providing annotations can alleviate this problem
- Need to improve annotations to overcome this problem

Example Tagging Systems

- Del.ioco.us (http://del.ioco.us)
- CiteULike (http://www.citelike.org)
- Flickr (http://www.flickr.com)
- YouTube (http://www.youtube.com)
- Last.fm (http://www.last.fm)
- Technorati (http://www.technorati.com)
- ESP Game (http://www.espgame.org)

Types of Tags

- Identifying What (or Who) it is About
- Identifying What it Is
- Identifying Who Owns It
- Refining Categories
- Identifying Qualities or Characteristics
- Self Reference
- Task Organising

System Design and Attributes

- Tagging Rights
- Tagging Support
- Aggregation
- Type of Object
- Source of Material
- Resource Connectivity
- Social Connectivity

User Incentives

- Future Retrieval
- Contribution and Sharing
- Attract Attention
- Play and Competition
- Self Presentation
- Opinion Expression

Tagging/Folksonomies Pros

- Easy to use
- Intuitive
- Cheap way of getting annotations
- Gives new users quick and simple impression of content
- Can aid browsing and search

Tagging/Folksonomies Cons

- Inaccurate and irrelevant tags
- Lack of stemming
- Freely chosen tags can result in
  - Synonyms
  - Homonyms
  - Polysemy

Tagging Conclusions

- Provide keywords to describe objects
- Collaborative tagging provides a community view of object
- Number of pros and cons to using tagging
- Number of design approaches can be taken
- Number of incentives for users to tag

Tagging Case Studies

- A number of large studies of tagging have taken place
- Focus on different aspects of tagging
- We will focus on three analyses of different systems, the systems analysed are:
  - Del.icio.us
  - Flickr
  - YouTube

Tagging Case Studies

	YouTube	Flickr	Del.ico.us
Rights	Self Tagging	Self Tagging	Free for all
Support	Blind tagging	Blind tagging	Blind tagging
Type of Object	Videos	Photographs	Links
Source	Participants	Participants	Web resource

Del.icio.us – Golder and Huberman

- Provide an analysis of collaborative tagging systems
- They study the del.icio.us system for organising bookmarks
- Analyse the structure of collaborative tagging systems as well as their dynamical aspects

Del.icio.us – Golder and Huberman

- Analyse two del.icio.us datasets
- Dataset 1
  - 212 URL's
  - 19,422 bookmarks
- Dataset 2
  - 229 users
  - 68,668 bookmarks

Del.icio.us – Golder and Huberman

- Users have variety in use of tags, some have many, some have few
- Tags vary in frequency of use
- However, stable patterns emerge in tags
- Adding numbers of infrequently used tags/opinions, does not disrupt the general consensus

Del.icio.us – Golder and Huberman

- Appears that most tags are added for personal use
- Never the less they are still useful to the public
- They believe that consensus choices that emerge may be used on a large scale to describe and organise how web documents interact with one another, and also can be used to make recommendations

Flickr – Marlow et al

- Present a study of the photo sharing and tagging system Flickr
- Looked at the dynamics of the Flickr and hope to expose interesting trends and topics in the Flickr
- Set of 25,000 users for individual analysis
- Set of 2,500 users for network analysis

Flickr – Marlow et al

- Present a model for tagging systems
- Compare tagging on Flickr with that of del.icio.us
- Dynamics of interaction and participation are different to Del.icio.us
- To be expected as they are different models of tagging, and different user incentives

YouTube – Halvey and Keane

- More interested in user goals when using and searching in a tag based system
- Examine why some videos are viewed more often than others
- Investigation of user interactions to see if they vary in Web 2.0
- Analysed104,465 video pages from 57,639 users

YouTube – Halvey and Keane

- Views of pages match distributions found in web search
- Views from navigation match distributions found in web navigation
- Found that in general the more tags that users provide the more likely that a video will be watched, up to a certain point

YouTube – Halvey and Keane

- In general users are consumers rather than creators of resources for the service
- Users of YouTube do not use the social tools unless they gain a benefit
- Videos are recommended because they are popular, not popular because they are recommended
- Videos receive the majority of their views in the first couple of days

Tagging Case Studies

- Briefly seen a number of analyses of tagging systems
- A number of different approaches were investigated, from a number of different aspects
- Although tagging and tag based systems seem random, there are a number of regularities

Tag Clouds

- Visual depiction of user generated tags
- Normally there is a weighting associated with the tags
- Can be alphabetised
- Importance of a tag can be represented by font size or colour
- Tags are usually hyperlinks that lead to further information

Tag Clouds

- There are a number of uses for tag clouds, these include
  - Browsing
  - Search
  - Impression Formation/Gisting
  - Recognition/Matching

Tag Clouds – Rivadeneira et al

- Investigated the use of tag clouds for impression formation
- Experiment 1
  - Influences of tag cloud attributes on low-level cognitive processes
- Experiment 2
  - Effect of font size and word layout on impression formation and memory

Tag Clouds – Rivadeneira et al

- The use of different font sizes had an effect
- Layout has an influence on the effectiveness of the users in performing tasks
- Some of the results can be attributed to westernised reading

Tag Clouds – Halvey and Keane

- Investigated browsing and search
- Participants carried out search tasks on tags presented 6 different formats
- Font size, alphabetisation and layout were varied for each of the layouts where appropriate.
- The effect of each scenario on task completion time was investigated

Tag Clouds – Halvey and Keane

- Alphabetisation can aid users to find information more easily and quickly
- Font size is very important for how quickly and easily users find information
- Position of tags is also very important
- It appears that users scan lists and clouds rather than read them

Other Web 2.0 Technologies

- Rich Internet Applications
- Really Simple Syndication (RSS)
- Cascading Style Sheets (CSS)
- Rest, XML, JSON
- Mash Up
- Wiki
- Weblogs
- Social Networking

Rich Internet Applications

- Web applications that have the features and functionality of desktop applications
- typically transfer the processing necessary for the user interface to the web client
- Keep the bulk of the data back on the application server
- Run in a web browser, and/or do not require software installation
- Run locally in a secure environment called a *sandbox*

Rich Internet Applications

- Can provide additional support for multimedia, which allow more interaction
  - Java Applications
  - User Interface Languages, e.g. SMIL
  - ActiveX
  - Google's GWT Framework

Really Simple Syndication (RSS)

- Web feed formats used to publish frequently updated content, e.g. Blogs, podcasts etc.
- RSS formats are specified using XML, extend the basic XML schema
- RSS first launched in 1999
- Several BitTorrent-based peer-to-peer applications also support RSS
- Media RSS from Yahoo provides an RSS for multimedia

Cascading Style Sheets (CSS)

- Stylesheet language used to describe the presentation of a document written in a markup language
- Separates document content from document presentation
- Improves accessibility, provides more flexibility and controls the specification of presentation characteristics
- CSS specifications are maintained by the W3C, CSS2 supports different media types

REST, XML or JSON API's

- Application Programming Interface (API) is a source code interface that an operating system or library provides to support requests for services
- Representational State Transfer (REST) is a collection of network architecture principles that outline how resources are defined and addressed
- JavaScript Object Notation (JSON) is a lightweight computer data interchange format

Mash Ups

- Combines data from more than one source into a single integrated tool
- Content is typically sourced from a third party via a public interface or API
- Yahoo, Google and Microsoft provide editors
- Three types of mash up
  - Consumer mash up
  - Data mash up
  - Business mash up

Wikis

- Computer software that allows users to easily create, edit and link web pages
- Can provide collaborative websites, power community websites, and effective intranets for use in knowledge management
- Wikipedia is one of the largest, it has approximately 9.1 million articles in 252 languages, comprising a combined total of over 1.41 billion words for all Wikipedias as of November 2007

Weblogs (Blogs)

- Website where entries are written in chronological order
- Combine text, images, and links to other blogs, web pages, and other media related to its topic
- Primarily textual, although some focus on art (artlog), photographs (photoblog), sketchblog, videos (vlog), music (MP3 blog), audio (podcasting)
- Technorati is main blog search engine as was tracking more than 106 million blogs as of September 2007

Social Networking Sites

- Social networks for communities of people, online
- Provide a collection of various ways for users to interact, including the use of multimedia
- Contain directories of some categories
- Contain means to connect with friends
- Use recommender systems linked to trust

Conclusions

- Tags provide cheap, easy metadata to describe objects
- Collaborative tagging (folksonomies) provide a community based description
- Tagging systems can vary in a number of ways
- There are a number of technologies that have emerged and are used as part of Web 2.0