Web2.0 in the light of HCI

Visualizing Tags over Time

CS 260
October 25, 2006
Hannes Hesse
What?

Web 2.0 is the **network as platform**, spanning **all connected devices**; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a **continually-updated service** that gets better the more people use it, consuming and **remixing** data from multiple sources, including individual users, while **providing their own data** and services in a form that allows remixing by others, creating network effects through an "**architecture of participation**," and going **beyond the page metaphor** of Web 1.0 to deliver rich user experiences.
Web 2.0 Meme Map

- Flickr, del.icio.us: Tagging, not taxonomy
- PageRank, eBay reputation, Amazon reviews: user as contributor
- Blogs: Participation, not publishing
- BitTorrent: Radical Decentralization
- Gmail, Google Maps and AJAX: Rich User Experiences
- Google AdSense: customer self-service enabling the long tail
- Wikipedia: Radical Trust

**Strategic Positioning:**
- The Web as Platform

**User Positioning:**
- You control your own data

**Core Competencies:**
- Services, not packaged software
- Architecture of Participation
- Cost-effective scalability
- Remixable data source and data transformations
- Software above the level of a single device
- Harnessing collective intelligence

- "An attitude, not a technology"
- The Long Tail
- Data as the "Intel Inside"
- The perpetual beta
- Software that gets better the more people use it
- Play
- Rich User Experience
- Hackability
- The Right to Remix "Some rights reserved"
- Emergent: User behavior not predetermined
- Granular Addressability of content

- Trust your users
- Small Pieces Loosely Joined (web as components)
Tags vs. Taxonomies
Taxonomies

hierarchical, exclusive

/articles/cats all articles on cats
/articles/africa all articles on Africa
/articles/africa/cats all articles on African cats
/articles/cats/africa all articles on cats from Africa
Categories

- Business to Business®
- Consumer Products and Services®

- ActiveX®
- Beginner's Guides (38)
- Books®
- Browsers®
- Caching (12)
- Chat (162) NEW!
- Conferences (7)
- Cookies (17)
- Databases and Searching (12)
- Domain Registration®
- Evaluation (9)
Tags
non-hierarchical, inclusive

“cats”

“cats” AND “africa”

“africa”
A model of tagging systems

(Marlow, Naaman, Boyd, Davis)
A simple taxonomy of tagging systems
The Vocabulary Problem

Probability of two people applying the same term to something typically below .2

Results in high failure rates in many common situations

(Furnas et al., 1987)
Figure 2. The number of tags in each user’s tag list, in decreasing order.

(Golder and Huberman)
Figure 2. Distribution of distinct tag collections, represented as the probability that a r

(Marlow, Naaman, Boyd, Davis)
Figure 7a,b. The stabilization of tags’ relative proportions for two popular URLs (#1310, #1209). The vertical axis denotes fractions and the horizontal axis time in units of bookmarks added.

(Golder and Huberman)
“Leveraging collective intelligence”
Human Computation
(von Ahn)
WORD: LAPTOP

It contains a KEYBOARD
Representing data that changes over time
Physics Balsa Bridge Building Winning Strengths
(three year moving average)
SmartMoney treemap
Martin Wattenberg
LifeLines
Brett Milash, Catherine Plaisant, Anne Rose, 1996
UMD
ThemeRiver
S. Havre et al., 2002
German economic and monetary union

NATO to redefine military strategy

OPEC agrees to raise oil price

Iraq invades Kuwait

ThemeRiver
S. Havre et al., 2002
A Brief History of Time

http://research.yahoo.com/taglines
Interestingness

\[
\frac{\text{# occurrences in interval}}{\text{# occurrences in entire collection} + C}
\]
compare **TF-IDF**
(term frequency * inverse document frequency)

\[ TF := \frac{\text{# of occurrences of term}}{\text{# of words}} \]

\[ IDF := \log \left( \frac{\text{# of all documents}}{\text{# of documents containing term}} \right) \]

\[ \text{TF-IDF} := TF \times IDF \]
The Dataset

86.8 M tag-photo instances
1.26 M unique tags
=> each tag used 70 times on average

data spans 472 days
=> 1.2 M incoming tags / week
Bulk Uploads
Interesting Tags

Events
Valentine’s Day, Thanksgiving, Xmas, Hanukkah, Republican National Convention
Web_2.0, Lunar eclipse, NYC Marathon, GoogleFest ...

Personalities
Jeanne Claude, Pope ...

Social media tagging
What’s in your fridge?, Faces_in_holes, dilodec04, What’s in your bag?, Badge, Do your worst ...
Discussion

Is the “interestingness” measure appropriate?

Do the results make sense to you?

What are the strengths and weaknesses of the visualization?