Tagging

IMT530: Organization of Information Resources
February 10, 2007

Michael D. Braly & Geoffrey B. Froh
Overview

• Definition and history of tagging
• Mechanics of tagging
• Tagging in the enterprise
• Where tagging fits
Section One: Definitions

- Definition and history of tagging
- Mechanics of tagging
- Tagging in the enterprise
- Where tagging fits
What is tagging?

“There is no fixed set of categories or officially approved choices. You can use words, acronyms, numbers, whatever makes sense to you, without regard for anyone else’s needs, interests, or requirements.”

- Shirky, 2005
What is tagging?

“[Tagging] sticks it to The Man, especially if The Man happens to be a traditional taxonomist.”

“It’ll be messy and inelegant and inefficient, but it will be Good Enough.”

- Weinberger, 2005
What is tagging?

“…free tagging, mob indexing, collaborative categorization, ethnoclassification, or whatever you want to call it.”

- Morville, 2005
What is tagging?

“[T]ags represent an allocation of attention.”
- Mejias, 2005

Of all the things in the world, I choose to acknowledge this thing, and I call it x.
Tagging: Tagged Poster

ink, chasm, tagging, tags, tagcloud, text, user-defined descriptors, Web 2.0

Over-used, exciting, future, emerging, mainstream, over-hyped
Tagging: del.icio.us
Tagging: del.icio.us
Tagging: del.icio.us
Tagging: last.fm
Tagging: last.fm
Tagging: steve.museum
Tagging: BBC Radio 6

- **Plantlife - When She Smiles She Lights The Sky**
  - Recent tag: plasticbag
  - Tags: punk, nowave, postpunk...
  - Rating: 4 stars
  - Date: 08/02/2005

- **The Rolling Stones - Start Me Up**
  - Recent tag: plasticbag
  - Tags: classic, blues, rock, england...
  - Rating: 4 stars
  - Date: 04/02/2005

- **Soul Wax - E Talking**
  - Recent tag: plasticbag
  - Tags: energy, noise, good, fun...
  - Rating: 4 stars
  - Date: 03/02/2005

- **Julian Cope - Trampoline**
  - Recent tag: plasticbag
  - Tags: beautiful, score, elegant, torchsong...
  - Rating: 4 stars
  - Date: 02/02/2005
Tagging: amazon.com

Playing with Fire [EXPLICIT LYRICS]
Kevin Federline (Artist)

List Price: $15.98
Price: $10.99 & eligible for FREE Super Saver Shipping on orders over $25. Details
You Save: $4.99 (31%)


Want it delivered Wednesday, January 31? Order it in the next 2 hours and 17 minutes, and choose One-Day Shipping at checkout. See details

42 used & new available from $7.99

Customers tagged this product with

enterpriseTagging.org
Customers tagged this product with

First tag: poser loser (The JuRk on Oct 19, 2006)
Last tag: dreck

Sort by:

<table>
<thead>
<tr>
<th>Tag</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>talentless</td>
<td>29</td>
</tr>
<tr>
<td>garbage</td>
<td>17</td>
</tr>
<tr>
<td>laughable</td>
<td>17</td>
</tr>
<tr>
<td>idiot</td>
<td>16</td>
</tr>
<tr>
<td>music_to_make_you_long_for_the_sweet_rain</td>
<td>15</td>
</tr>
<tr>
<td>hack</td>
<td>13</td>
</tr>
<tr>
<td>crap</td>
<td>12</td>
</tr>
<tr>
<td>loser</td>
<td>11</td>
</tr>
<tr>
<td>terrible</td>
<td>9</td>
</tr>
<tr>
<td>clown</td>
<td>7</td>
</tr>
<tr>
<td>poser</td>
<td>7</td>
</tr>
<tr>
<td>trash</td>
<td>7</td>
</tr>
<tr>
<td>wigger</td>
<td>7</td>
</tr>
<tr>
<td>aka_vegon_poetry</td>
<td>6</td>
</tr>
<tr>
<td>moron</td>
<td>3</td>
</tr>
<tr>
<td>pathetic</td>
<td>5</td>
</tr>
<tr>
<td>wannabe</td>
<td>5</td>
</tr>
<tr>
<td>stupid</td>
<td>4</td>
</tr>
<tr>
<td>wtf</td>
<td>4</td>
</tr>
<tr>
<td>awful</td>
<td>3</td>
</tr>
<tr>
<td>burn</td>
<td>2</td>
</tr>
<tr>
<td>cole_slaw</td>
<td>2</td>
</tr>
<tr>
<td>death</td>
<td>2</td>
</tr>
<tr>
<td>demented</td>
<td>2</td>
</tr>
<tr>
<td>depressing</td>
<td>2</td>
</tr>
<tr>
<td>detectable</td>
<td>2</td>
</tr>
<tr>
<td>disgusting</td>
<td>2</td>
</tr>
<tr>
<td>dreadful</td>
<td>2</td>
</tr>
<tr>
<td>dull</td>
<td>2</td>
</tr>
<tr>
<td>dumbass</td>
<td>2</td>
</tr>
<tr>
<td>dung</td>
<td>2</td>
</tr>
<tr>
<td>epic_fail</td>
<td>2</td>
</tr>
<tr>
<td>every_track_ought_to_be_hidden</td>
<td>2</td>
</tr>
<tr>
<td>fed_ex</td>
<td>2</td>
</tr>
<tr>
<td>fnshee</td>
<td>2</td>
</tr>
<tr>
<td>god_awful</td>
<td>2</td>
</tr>
<tr>
<td>herpes_would_be_more_enjoyable</td>
<td>2</td>
</tr>
<tr>
<td>hopi_indians</td>
<td>2</td>
</tr>
<tr>
<td>horrid</td>
<td>2</td>
</tr>
<tr>
<td>i_left_justin_for_this</td>
<td>2</td>
</tr>
<tr>
<td>joke</td>
<td>2</td>
</tr>
<tr>
<td>lol</td>
<td>2</td>
</tr>
<tr>
<td>makes_baby_jesus_cry</td>
<td>2</td>
</tr>
</tbody>
</table>
Precursors of Tagging

• Formal: Dialog’s user-suggested free-text keywords
  – User suggestions were incorporated in CV change management process
  – Users were mostly experts (reference librarians)

• Informal: User generated content
  – eBay listing lingo and buyer/seller ratings
  – Amazon stars and reviews
Precursors: Déjà vu all over again

Rowley’s 4 eras in indexing

- Era 1: Pre-computer access
- Era 2: Online age
- Era 3: Full-text vs. subject indexing
- Era 4: Tests with real users instead of controlled experiments

Tagging vs. Controlled Vocabulary Debate Is Somewhere Early in Era 3…
tag != keyword
Tags and Controlled Vocabulary

- Process Perspective
- Structural Perspective
- Cognitive Perspective
Process: CVs vs. Tags

Expert determines acceptable list of keywords

People apply term to resource

Resource is described using an expert’s language

Folksonomy evolves

List of tags viewed in aggregate

People describe a resource
Structural: Taxonomy vs. Tagging

**Taxonomy**
- Hierarchical
- Exclusive
- A Folder System

**Tagging**
- Flat
- Non-exclusive
- A Filtered Query

Golder & Huberman, 2005
Cognitive: CV Indexing

Sinha, 2005
Cognitive: Tagging

Resource worth remembering → Multiple concepts activated → Tag

Sinha, 2005
Why do individuals tag?

- Future retrieval
- Contribution and sharing
- Attract attention
- Play and competition
- Self presentation
- Opinion expression

Marlow et al, 2006
How are we doing? Questions or comments so far?
Section Two: Mechanics

- Definition and history of tagging
- Mechanics of tagging
- Tagging in the enterprise
- Where tagging fits
Is Phlat tagging?
Is Phlat social tagging?
Three components of social tagging

Person

Tag

Resource
Wisdom of Crowds

Preconditions for Emergence of Collective Intelligence

- Cognitive diversity: many opinions
- Independence: individuals can freely offer opinions
- Decentralization: opinions are evaluated equally (more or less)
- Easy Aggregation: all opinions can be utilized

Surowiecki, 2005
Theory of Networks

- Formal methods for modeling relationships in systems (“graphs”)
- Applied in social science as Social Network Analysis
- Milgram: Six degrees of separation
“New” Theory of Networks

- Systems that are highly dynamic
  - Molecular state change
  - Neuronal connections
  - North American power grid
- Watts, Strogatz, Barabási

Image: Hawoong Jeong
Social Tagging as Dynamic Network?

- Social tagging presents many features of a dynamic network system
  - Users are joining the system
  - Users are submitting new tags and new bookmarks
  - New web resources are continually emerging to be bookmarked
  - Language is constantly changing
- del.icio.us easy to model as a network and provided access to data

Golder & Huberman, 2006
Stable patterns in dynamic systems

Polya-Eggenberger Urn Model

1. Place a red and a black ball in the urn.
2. Select a ball.
3. Place the ball back in urn with one more of same color.
4. Repeat $N$ times.

Fraction of red/black balls becomes stable at a random limit.
Patterns in tag proportions

At about 100 bookmarks, the proportion of each individual tag to the total group of tags is statistically stable.
Stability in tag proportions

• Why is this happening?
  – Imitation
  – Shared knowledge

• What are the implications?
  – Collective meaning can emerge from individual tagging
  – Lesser used tags can coexist with more popular tags
  – Less popular URLs can produce useful tag data
Dynamic Structure of Social Tagging: Research

- Exposed sub-communities of users and inferred genre clusters in audioscrobbler and Last.fm (Lambiotte & Ausloos, 2005)
- Used del.icio.us to model concept of lightweight, self-organizing emergent ontology (Mika, 2005)
Questions or comments?
Take a Break!
Section Three: Enterprise

- Definition and history of tagging
- Mechanics of tagging
- Tagging in the enterprise
- Where tagging fits
Enterprise Controlled Vocabularies
Does this look familiar?

• Specific Purposes
  – To provide access to content by subject, through providing hierarchical and associative relationships and synonym control for the terms used in the domain
  – Increase precision in retrieval and display by controlling homographs (words that are spelled the same but have different meanings)

• General Purposes
  – Assist users by conveying meaning, orientation, and structure in a subject area
  – Assist users by providing rich relationships among concepts and terms
Specific Purposes - CV

To provide access to content by subject, through providing hierarchical and associative relationships and synonym control for the terms used in the domain.
Specific Purposes - Tags

To provide access to content by subject, through providing associative relationships between the user’s own terminology with that of other users as well as controlled domain vocabulary.
Specific Purposes - CV

Increase precision in retrieval and display by controlling homographs (words that are spelled the same but have different meanings)
Specific Purposes - Tags

Increase precision and recall in retrieval by providing a greater range of possible subject terms than manual or automatic indexing can produce.
General Purposes - CV

Assist users by conveying meaning, orientation, and structure in a subject area
General Purposes - Tags

Assist users by surfacing patterns of shared meaning, orientation, and structure in a subject area
General Purposes - CV

Assist users by providing rich relationships among concepts and terms
General Purposes - Tags

Assistant users by exposing relationships between concepts, resources and other users
Purposes of social tagging?

• Specific Purposes
  – To provide access to content by subject, through providing associative relationships between the user’s own terminology with that of other users as well as controlled domain vocabulary
  – Increase precision and recall in retrieval by providing a greater range of possible subject terms than manual or automatic indexing can produce

• General Purposes
  – Assist users by surfacing patterns of shared meaning, orientation, and structure in a subject area
  – Assist users by exposing relationships between concepts, resources and other users
Enterprise tagging environment

HR/Directory systems
Identity and profile management
Collaboration platforms

Person

Tag
Application-specific CVs
Industry thesauri
WordNet

Resource
Search engines
Document repositories
CMS
Wikis
How tagging can be used

• Learn the vocabulary of the people ("User Warrant")
• Enhance a controlled vocabulary
• Search enhancement
How tagging can be used

• People
  – Community of Expertise
  – Community of Interest
  – “Assist users by providing rich relationships among concepts and terms.” (Mike C)
  – Assist users by exposing relationships between concepts, resources and other users.
Lotus Connections
How are we doing? Questions or comments so far?
Section Four: Where tagging fits

• Definition and history of tagging
• Mechanics of tagging
• Tagging in the enterprise
• Where tagging fits
Surface Problems

- Problems with Language
- Problems with Users
Problems with language

- Semantic issues (Golder & Huberman, 2005)
  - **Homonymy**: same word with different meanings
    - patient
    - mad
  - **Polysemy**: same word with different senses
    - a thing you walk through
    - a thing you open and close
  - **Synonymy**: different words with the same meaning
    - PC
    - computer
  - **Basic level variation**: different levels of specificity
    - fish
    - salmon
    - sockeye

- Multi-lingual environments
Problems with Users

Unintentional bad acts
- “Sloppiness” (Guy & Tonkin, 2006)
  
  sematic
  PetersonElaine emma_tonkin Marieke-Guy
dogs Dog
imt530b

- Lack of expertise (aka “Stupidity”) (Peterson, 2006)
Problems with users

Intentional bad acts

– Offensive/inappropriate content (“graffiti”)
  WorstBossEver movie_downloadz
– System manipulation (“gaming,” “tagspam”)
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time
  best_website_for_big_sexy_time

Deep problems

- Inattention
- Authority
Deep problems: Inattention

- How do you get the first tag
  - Tagging is post-hoc process
  - Valuable resources may be hard-to-find
  - Users may not be familiar with the corpus

- What happens if users don’t participate
  - System is difficult to use
  - Knowledge domain is not understood
  - Concept of tagging is not understood
Deep problems: Authority

“Merely naming the world creates no actual change, either in the world, or in the minds of potential users who don’t understand the system.”

- Shirky, 2005
Deep problems: Authority

• Naming is about rules, control and politics and it is important

• Construction of meaning
  – **CVs**: Fixed; negotiated in advance; authority is an inherent property
  – **Tagging**: Dynamic; post-hoc; flows upward from multiple sources

• In organizations, source of authority sometimes must be predetermined and fixed
  – Product lists, position titles, navigation labels
Where Ontology fits

Better

Domain

Small corpus
Formal categories
Stable entities
Restricted entities
Clear edges

Participants

Expert catalogers
Authoritative source of judgment
Coordinated users
Expert users

Worse

Domain

Large corpus
No formal categories
Unstable entities
Unrestricted entities
No clear edges

Participants

Uncoordinated users
Amateur users
Naïve catalogers
No Authority

Shirky, 2005
Where tagging fits

Worse

Applications
- Accounting
- Payroll
- Health Records
- Supply Chain Mgmt

Better

Applications
- Wikis
- Knowledge Base
- Team Sites
- Forms Site

enterprise-tagging.org
Thank You!
Discussion and Questions
References

- Del.icio.us
- Flickr.com
- Last.fm
- Amazon.com
- Golder
- Shirky
- Steve.museum
- Marlow et al
- Sinha
- Crandall
- Peterson
- Linked
- Emergence
- Wisdom of Crowds
- Morville
Appendices
Tips for tag designers (Rashmi Sinha, SXSW06)

• How are you serving the individual’s motives?
• Does the individual understand and really want to fulfill that goal?
• What is the relationship between the social and the personal?
• Is it too easy to mimic the tags of others?
Tips for tag designers (Rashmi Sinha, SXSW06)

- Don’t make navigation all about the most popular, most tagged…
- Enable discovery, exploration, finding new things
- Don’t force users to do things differently than what comes naturally. Take user input as it comes naturally.
- Solve problems by ensuring good findability.