# 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

"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

"[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

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

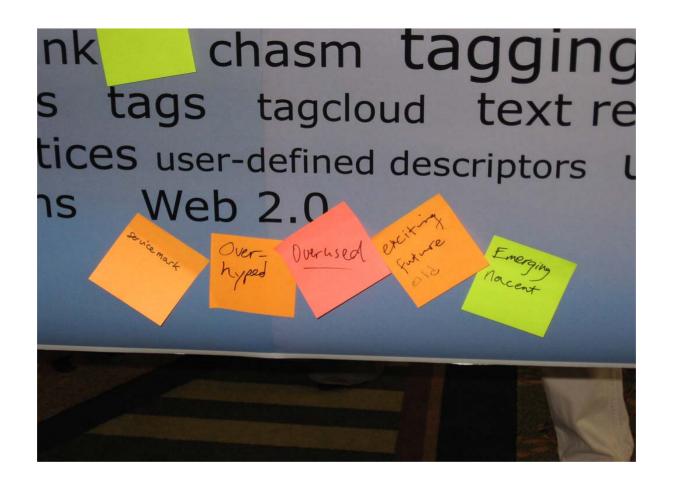
- Morville, 2005

"[T]ags represent an allocation of attention."

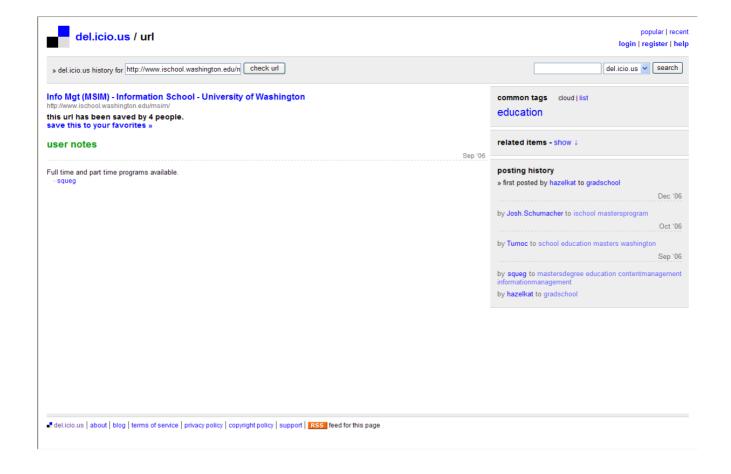
- Mejias, 2005

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

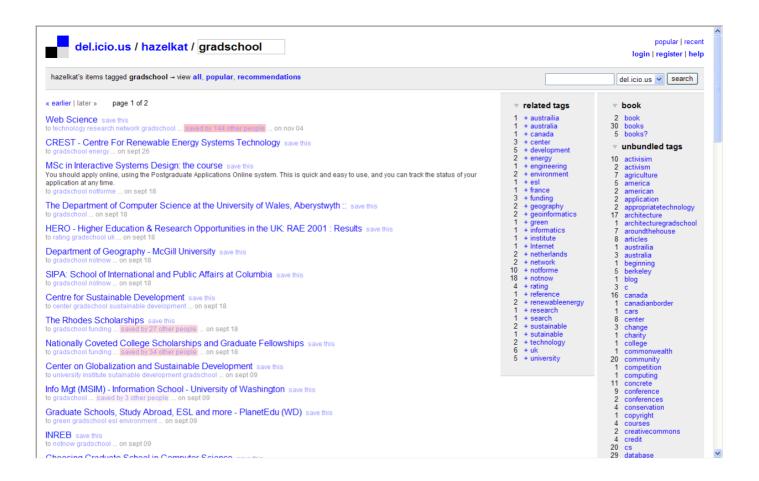
## Tagging: Tagged Poster



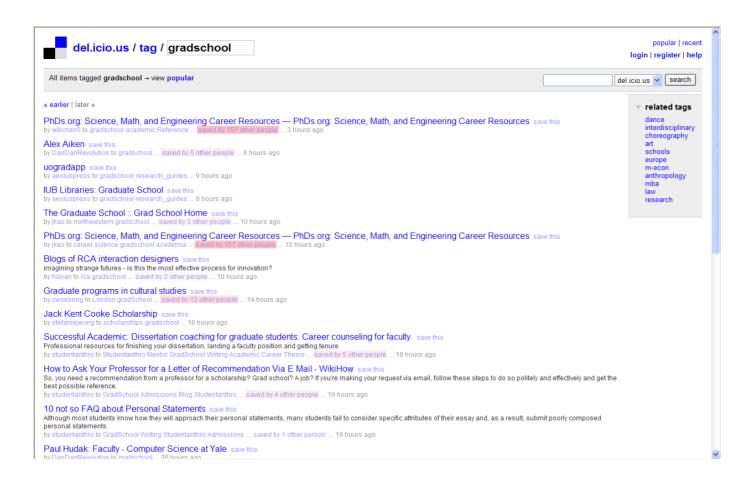
### Tagging: del.icio.us



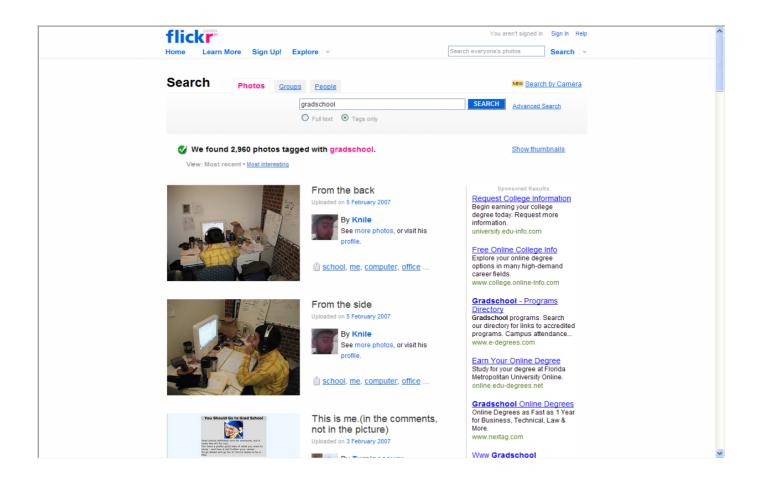
#### Tagging: del.icio.us



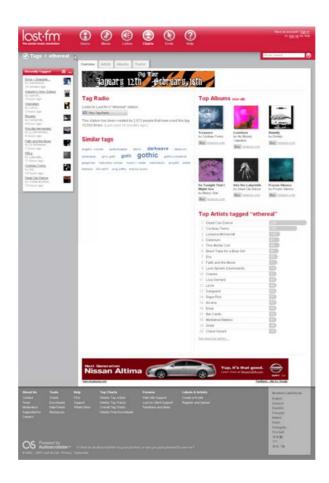
#### Tagging: del.icio.us



#### Tagging: flickr.com



## Tagging: last.fm



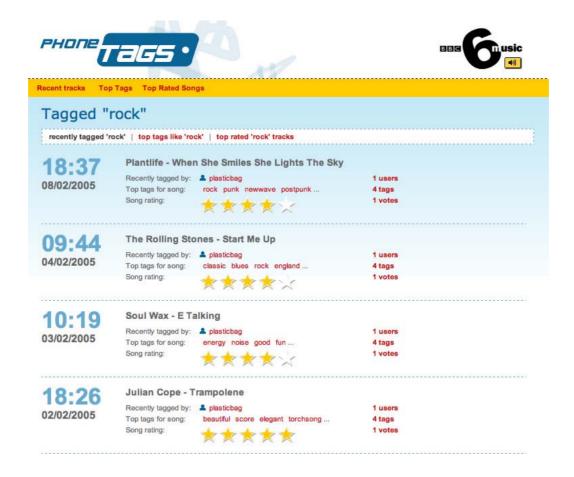
## Tagging: last.fm



# Tagging: steve.museum



## Tagging: BBC Radio 6



#### Tagging: amazon.com



#### Tagging: amazon.com

#### Customers tagged this product with First tag: poser loser (The JuRK on Oct 18, 2006) Last tag: dreck Sort by: Popularity talentless (29) cole slaw (2) garbage (17) death (2) laughable (17) demented (2) idiot (16) depressing (2) music to make you long for the detestable (2) sweet rel... (16) disgusting (2) dreadful (2) hack (13) dull (2) crap (12) loser (11) dumbass (2) terrible (9) dung (2) clown (7) epic fail (2) poser (7) every track ought to be hidden (2) trash (7) fed ex (2) frisbee (2) wigger (7) aka vogon poetry (6) god awful (2) herpes would be more enjoyable (2) moron (5) pathetic (5) hopi indians (2) wannabe (5) horrid (2) stupid (4) i left justin for this (2) wtf (4) ioke (2) awful (3) lol (2) bum (3) makes baby jesus cry (2)

#### 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

- Era1: 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

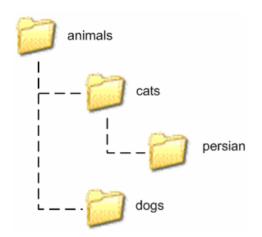


People describe a resource

#### Structural: Taxonomy vs. Tagging

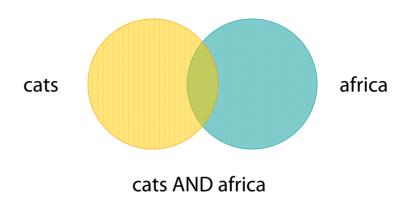
#### **Taxonomy**

- Hierarchical
- Exclusive
- A Folder System

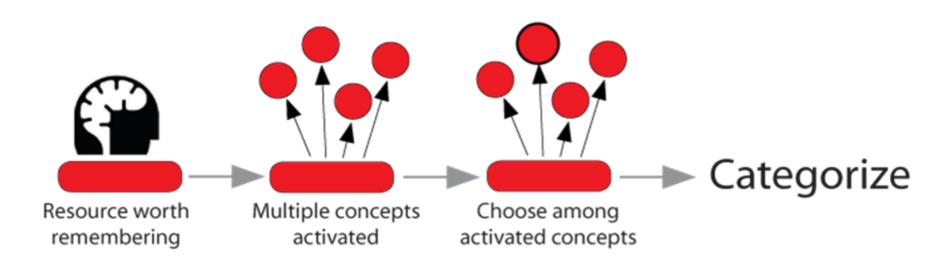


#### **Tagging**

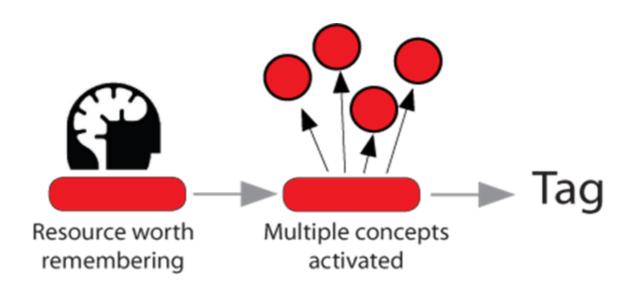
- Flat
- Non-exclusive
- A Filtered Query



#### Cognitive: CV Indexing



#### Cognitive: Tagging



#### Why do individuals tag?

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

# 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







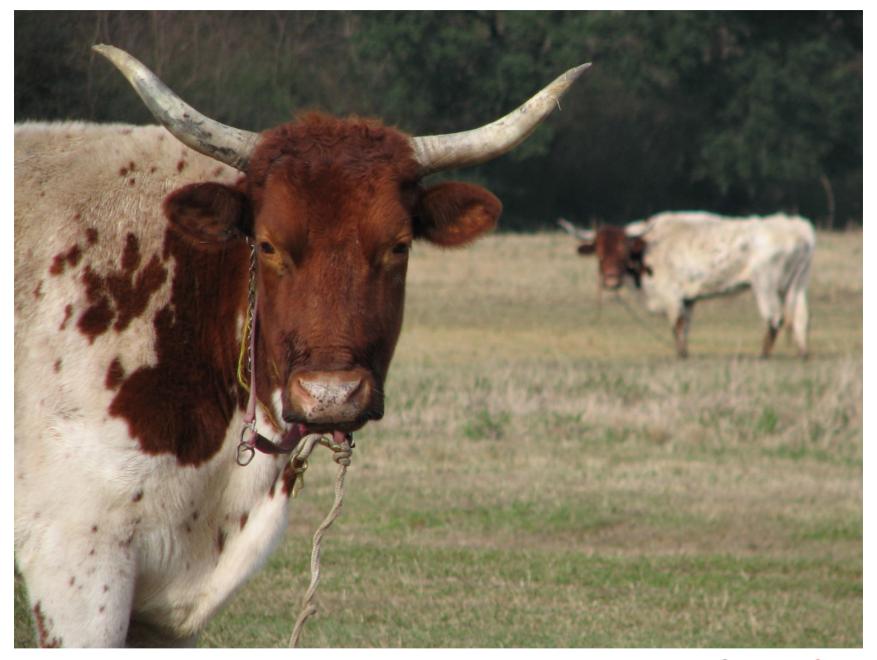


Photo: tjean314 (http://www.flickr.com/photos/tjean314/379212666/)

enterprisetagging.org

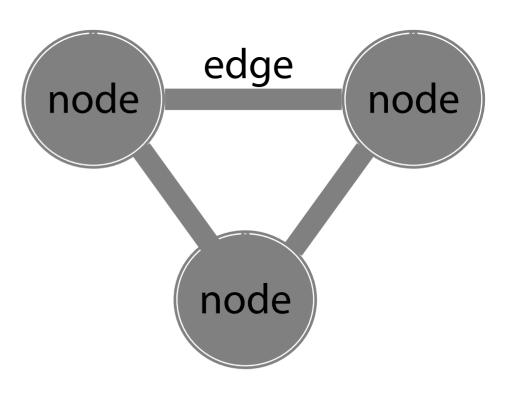
#### Wisdom of Crowds

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

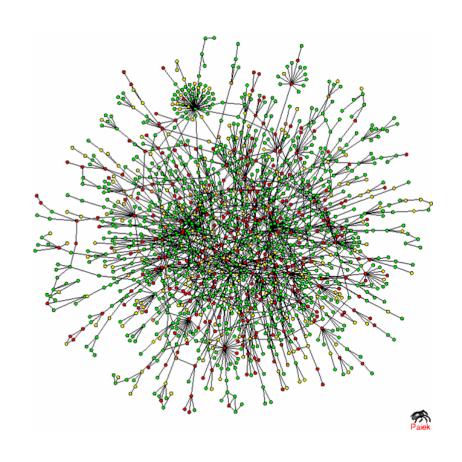
#### Theory of Networks

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

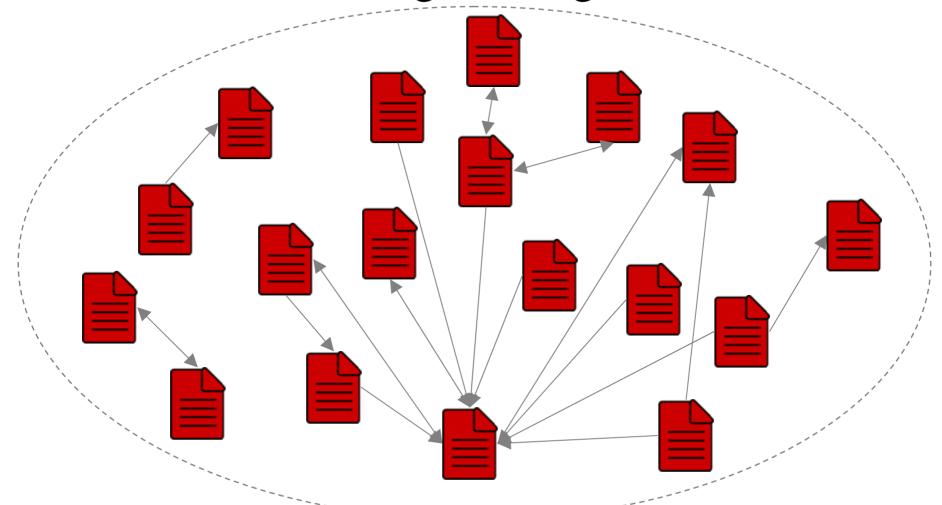


## "New" Theory of Networks

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



# Google Magic



# 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

# 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.

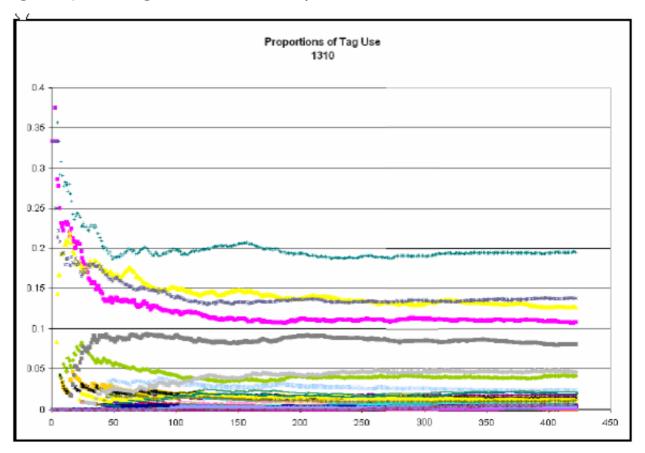
Fraction of red/black balls becomes stable at a random limit.



enterprisetagging.org

# 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



# Microsoft®

### 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

Assist 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

Application-specific CVs

Industry thesauri

WordNet





Search engines

Document repositories

**CMS** 

Wikis

enterprisetagging.org

# How tagging can be used

- Learn the vocabulary of the people ("User Warrant")
- Enhance a controlled vocabulary
- Search enhancement

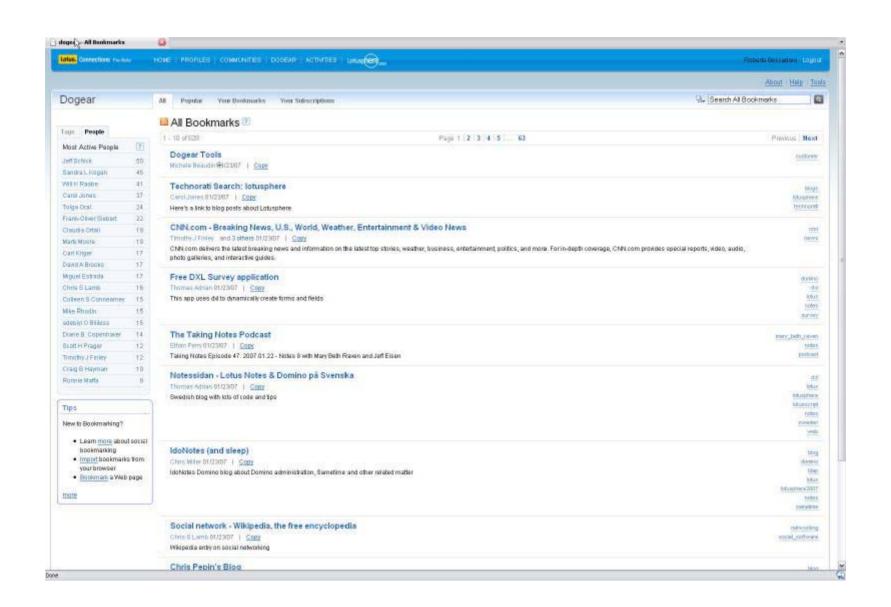
# How tagging can be used

### 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 meaning 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 enterprise tagging.org
```

# 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



#### **Domain**

Small corpus Formal categories Stable entities Restricted entities Clear edges

#### **Participants**

Expert catalogers
Authoritative source of
judgment
Coordinated users
Expert users

#### **Domain**

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

#### **Participants**

Uncoordinated users Amateur users Naïve catalogers No Authority

# Where tagging fits



**Applications** 

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

Wikis Knowledge Base Team Sites Forms Site

## 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.