

Tagging research: A survey and agenda

Jonathan Furner
GSE&IS, UCLA

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

February 22, 2008

Gosh

- “There is a **revolution** happening on the internet that is alive and building momentum with each passing tag.” (Kroski 2005, quoted in Abbas & Turner 2006).

“sky metadata” image by woowoowo: <http://www.flickr.com/photos/leprecon/57830899/> Some rights reserved.

Oops

- “Despite a considerable amount of attention in academic circles, as represented in various blog posts ..., little academic research work has been invested in tagging systems to date.” (Marlow et al. 2006).
- “The literature of tagging is largely **opinion-based** [rather than evidence-based] ... and the topic is largely absent from academic literature.” (Speller 2007).

The history of tagging research ...

- ... predates tagging
- tagging = **indexing / metadata-generation**

Mid-90s precursors

- Pauline Brown [**Rafferty**] & Rob Hilderley (Central England): “democratic indexing”
- Brian C. **O’Connor** (North Texas): “user-generated descriptors”
- James M. **Turner** (Montréal): “user-assigned terms”

What's in a name?

distributed collaborative
social inclusive ethno
 mob wikified community
 folk(sonomic) democratic
 dynamic cooperative
 (end-)user-generated, -supplied
 reader-assigned
 audience-derived
 public-created, -contributed
 visitor-produced, -defined

recordkeeping
 documentation
 ordering arrangement
 description metadata
 representation
 taxonomy bookmarking
 organization classification
tagging indexing ranking
 annotation cataloging
 categorization

Tagging and user tagging

- **tagging**
 - assigning **tags** to **resources**
 - aim: (i) **description** of resources
(ii) **access** to resources
- **user tagging**
 - tagging by “users”!
 - i.e., the population of taggers (describers)
= the population of searchers (access-seekers)

Characteristics of user tagging

- it's user-oriented
- it's empowering
- it's democratic
- it's **cheap**
- it's collaborative
- it's distributed
- it's dynamic
- it's instructive

Some clarification

- [archetypically] user tagging = indexing that is done ...
 - by people (who are non-professional taggers)
 - using computers
 - not using vocabulary control
 - to digital objects, e.g., web pages, digital images, videos; or to digital records of concrete objects, e.g., books, paintings, people, or of abstract objects, e.g., editions, texts, works ...
 - producing links between (tagger,) tag, and tagged
 - for resource discovery/rediscovery

Some clarification, cont'd

- **folksonomy** (coined by Vander Wal in 2004 [Smith 2004]) = a vocabulary ...
 - that emerges from multiple acts of tagging
 - that can be used as a kind of search thesaurus ...
 - allowing searchers to recognize rather than requiring them to recall
 - “without [their] needing to know an often outdated, Anglo-centric controlled vocabulary that librarians and users alike do not always know or understand” (Abbas & Turner 2006)
 - [archetypically] imposing little structure other than a popularity ranking: non-hierarchical

Summary

- proposal of framework for use in ...
 - reviewing previous work
 - suggesting directions for future work

Research fronts

1. **practices:** *How* do people tag?
2. * **contexts; motivations; meanings; incentives:** *Why* do (/should) people tag?
3. **networks; communities; cultures:** *What social structures* shape and are shaped by tagging practices?
4. **designs; interfaces:** *How might* people tag?

Research fronts, cont'd

5. **effects** (descriptive): What *kinds of relationships* among taggers, tags, and tagged resources are expressed as a result of tagging?
6. * **effects** (evaluative): Does tagging *work*?
7. * **methods**: In what ways might (/should) we *study* tagging?
8. **institutions; fields; paradigms**: How does tagging change *LIS*?

1. Practices

- *How do people tag?*
 - cognitive processes
 - social processes
- Do patterns of tagging activity and behavior change as systems grow?
- “Will the masses ever tag?” (Furnas et al. 2006)

2. Contexts; motivations; meanings; incentives

- *Why* do (/should) people tag?
 - explanatory/predictive
 - In what **contexts** do/don't (/should/shouldn't) taggers tag, and why?
 - What are (/should be) their **motivations**, objectives, goals?
 - What **meanings** do tagging practices (their own and those of others) have for taggers?
 - What systems of **incentives** should be in place for a tagging system to succeed?

2. Contexts; motivations; meanings; incentives, cont'd

- In what **contexts** do/don't (/should/shouldn't) designers offer tagging opportunities, and why?
- What are (/should be) their **motivations**, objectives, goals?
- What **meanings** do tagging practices have for designers?

3. Networks; communities; cultures

- What *social structures* shape and are shaped by tagging practices?
 - issues of trust, authority, representation, diversity, privacy
 - e.g., Are frequent taggers the “new experts”?

4. Designs; interfaces

- How *might* people tag?
- How might folksonomies be visualized?
- How might distributed indexing work in tandem with vocabulary control?
 - integrating both in hybrid systems
 - creating thesauri from statistical analysis of co-occurrences of tags
- How might distributed indexing work in tandem with social network analysis?
 - e.g., addressing retrieval-effectiveness problems by weighting tags that members of a particular social network use more heavily

4. Designs; interfaces, cont'd

- How might users be modeled through analysis of their tagging activity (e.g., to support recommendation systems)?
- Does system design influence ...
 - level of user participation
 - tag content
 - rate of convergence on coherent folksonomy? (Furnas et al. 2006)

Some dimensions of variation in implementations

- type of tagged resource: books, journal articles, web pages, blog entries, encyclopedia entries, photos, videos, museum objects ...
- multiple taggers per resource (Del.icio.us) vs. single tagger per resource (Flickr, YouTube)
- online resources (Del.icio.us, Flickr) vs. offline resources (LibraryThing, Steve)
- user-authored resources (Flickr) vs. non-user-authored resources (Steve, Last.fm)
- ... etc.

Toward a taxonomy of user tagging systems

- cf. Hammond et al. 2005; Marlow et al. 2006
- type of parent institution
- type of user
- type of resource
- type of access
- functionality of tagging service
- functionality of search service
- goals of implementers
- motivations of end-users

Marlow et al.'s taxonomy

- tagging rights: self-tagging vs. free-for-all tagging
- tagging support: blind tagging vs. viewable tagging vs. suggestive tagging
- aggregation: bag-model vs. set-model (denying repetition)
- type of object
- source of material: participants vs. system vs. open
- resource connectivity: allowing resources to be grouped
- social connectivity: allowing users to be grouped

5. Effects (descriptive)

- What kinds of relationships among taggers, tags, and tagged resources are expressed as a result of tagging? (cf. Golder & Huberman 2006)
- What kinds of tags, and what kinds of sets of tags, are generated ...
 - by different kinds of tagger?
 - for different kinds of resource?
 - for different kinds of purpose?
 - in different kinds of context?
 - over different periods of time?
- What kinds of folksonomic structure emerge from tagging activity?

Golder & Huberman's study of Del.icio.us

- a few people use a lot of tags; a lot of people use a few tags
- the rate of growth in people's tag sets varies from person to person
- tags can identify what a resource is about, what it is, who created it, what the tagger's opinion of it is, etc. (cf. Kipp 2007)
- a few tags are used by a lot of people; a lot of tags are used by a few people
- the first tags in any bookmark tend to represent basic levels

Golder & Huberman's study of Del.icio.us, cont'd

- most resources attract most of their bookmarks quickly, but some attract bursts in popularity
- well-connected individuals can influence “burstiness”
- each tag’s frequency as a proportion of the total frequency of all tags assigned to a given resource remains stable over time: explainable by ...
 - the “**Imitation**” argument: actions are viewed as correct to the extent that one sees others doing them
 - the “**Shared knowledge**” argument: the consensus is real!
- “a significant amount of tagging, if not all, is done for personal use rather than public benefit”

6. Effects (evaluative)

- Does tagging *work*?
 - **outputs**-based:
 - e.g., To what extent do *tags* support resource discovery?
 - Are the precision/recall scores “Good Enough”?
 - **outcomes**-based:
 - e.g., To what extent does *tagging* support taggers’/designers’ goals?

7. Methods

- In what ways might (/should) we *study* tagging?
 - e.g., What new **evaluation** frameworks are necessary?

8. Institutions; fields; paradigms

- How does tagging change *LIS*?
 - Does the “new ecology” of tagging represent a paradigm shift?
 - Is the new search culture one of satisficing rather than optimizing?

Motivations to offer the opportunity to tag

- to improve resource **discovery** (aka information retrieval): i.e., to allow people to discover resources tagged by others
- to improve resource **rediscovery**: i.e., to allow people to organize resources so that they can be found again (cf. Flickr: “to enable new ways of organizing photos”)
- to help people **share** resources, and knowledge about resources (cf. Flickr: “to help people make their photos available to the people who matter to them”)
- to help people **collaborate** with others, and/or participate in community/group activities

Motivations to offer the opportunity to tag, cont'd

- to help people attract attention to their own resources
- to help people identify others with shared interests
- to help people identify popular resources and topics
- to help people engage with resources
- to help people make meaning/sense
- to help people express and communicate their opinions
- to help people construct and present their identity

Motivations to offer the opportunity to tag, cont'd

- to help people demonstrate their knowledge
- to help people feel like they're part of a community
- to help people feel like they're empowered
- to help people play and compete with others
- to provide people with entertainment
- to reduce costs and/or make money

Motivations of taggers

- cf. Marlow et al. 2006
- individual users are likely to have multiple motivations simultaneously
- some will consciously repurpose available systems to meet their own needs
- some will be persuaded by the norms of their friends

Mental models of tagging services' functions

- kinds of primary motivation
 - **individualist** vs. **social**
- kinds of primary usage
 - (tagging-) **intrinsic** vs. (tagging-) **instrumental**
- kinds of ultimate goals
 - (resource-) **intrinsic** vs. (resource-) **instrumental**

Justifications of tagging

- the “It’s easy and fun!” argument
- the “Information overload” argument
- the “Where’s the text?” argument
- the “Tagging be nimble” argument
- the “Only users know what users want” argument
- the “Wisdom of crowds” argument
- the “Sticking it to The Man” argument

The “It’s easy and fun!” argument

- tagging has a low “participation barrier”
- “Using a carefully crafted taxonomy or categorization requires an expert to design the category structure and training to those who assign categories to resources. Conversely, tagging is a much lighter task cognitively that requires no previous consideration or training ...” (Furnas et al. 2006).
- whatever, have you ever *played* the ESP Game? (Von Ahn & Dabbish 2004).

The “Information overload” argument

- there’s too much content, and traditional [manual] indexing practices don’t scale up
- [an argument for distributed indexing, but not necessarily for relaxed vocabulary control]
- [anyway, what’s wrong with automatic indexing?]

The “Where’s the text?” argument

- tagging is good in cases where automatic full-text indexing isn’t possible or useful
 - e.g., visual resources; music; fiction
- so, there are particular environments where tagging provides a flexible way of organizing stuff

The “Tagging be nimble” argument

- folksonomies adapt quickly to reflect changes in terminology
- [but what happens if taggers don't retag older resources after those changes?]

The “All the right words” argument

- cf. Furnas 1987
- the “vocabulary problem” in IR: different people use different words to describe the same things (→ reduced recall) and the same words to describe different things (→ reduced precision)
- so let’s just assign *all* the index terms that people could possibly come up with as search terms
- “thoughtful sociotechnical design of tagging systems may uncover ways to overcome the Vocabulary Problem without requiring either the rigidity and steep learning curve of tightly controlled vocabularies, or the computational complexity and relatively low success of purely automatic approaches to term disambiguation” (Marlow et al. 2006)

The “Only users know what users want” argument

- traditionally-controlled vocabularies actually hinder retrieval, because they do not match users’ vocabularies
- if index terms are created by users, other users are more likely to find what they need
- “An author is an authority when it comes to what she intended her work to be about, but not about [what] it means to others.” (Weinberger 2005).
- [(cuts both ways for systems like Flickr)]

The “Only users know what users want” argument, cont’d

- [traditional vocabularies are based on twin principles of literary/user warrant]
- [importance of user-centered indexing, and indexer–user consistency, is well-established in LIS literature: cf. Fidel 1994; Bates 1998]
- [Is the real problem the implicit assumption that user groups are homogeneous?]

The “Wisdom of crowds” argument

- cf. Galton 1907:
 - 787 people at a country fair guessing the weight of an ox
 - median guess was within 0.8% of the correct weight
 - mean of guesses was within 0.01% of the correct weight
- i.e., most expertise can be replicated through repeated non-expert input

The “Wisdom of crowds” argument, cont’d

- [to become wise, a crowd needs:
 - diverse opinions
 - independent decision-making
 - decentralization of power ...
- otherwise it might impose a “**tyranny of the majority**” or “digital Maoism” (Lanier 2006)]

The “Sticking it to The Man” argument

- “The Man” = Melvil Dewey
- “The idea that to know a field is to see its **structure** is coextensive with Western history. We have spent an inordinate amount of time encouraging **experts** and authorities to construct huge structures of classification, from trees of life to trees of knowledge. The urge to **tidy up** has shown itself most recently on the Web as the Semantic Web. The tagging movement says, in effect, that we’re not going to wait for the experts to deliver a taxonomy from on high. We’re just going to build one **ourselves**. It’ll be messy and inelegant and inefficient, but it will be **Good Enough**. And, most important, it will be *ours*, reflecting our needs and our ways of thinking.” (Weinberger 2005).

The “Sticking it to The Man” argument, cont’d

- “... [T]agging repudiates one of the deepest projects our culture has undertaken over and over again: The rendering of all knowledge into a **single, universal** framework. The rendering has been assumed to be a process of **discovery**: The universe has an inner order that experts and authorities can expose. But in a networked world we know better than ever that such order is a myth of rationality. We can’t even agree even on basic issues such as what constitutes a ‘major’ religion or a ‘legitimate’ state. Order and categorization, we are learning, depend on **context** and project. The semi-chaotic state of the ‘tagosphere’ represents the nature of our shared world better than the cool marble columns of the old mono-order ever could.” (Weinberger 2005).

The “Sticking it to the man” argument, cont’d

- distributed indexing creates a bottom-up consensus view of the world
- which is [inherently, necessarily] more valid than any one view imposed from top-down (cf. Shirky 2005)
- i.e., it’s better to give most of the power to some [opinionated] people [often young ones who “don’t care about the common good”], than to give all the power to an educated few

Problems for tagging

- lack of **synonym** control → reduced recall
- lack of **homonym/polyseme** control → reduced precision
- lack of compliance with standards guiding taggers' decisions as to ...
 - tags' **form**: singular vs. plural; multi-word tags; misspellings
 - tags' **specificity**: the “bird”/“robin” problem
 - tags' **exhaustivity**: How many tags per resource?
- the “**Matthew effect**” → the tyranny of the majority

Problems for folksonomies

- lack of **hierarchical, faceted** structure
- lack of support for **browsing**

Where do *you* stand on vocabulary control?

- “We face a paradox. Ostensibly, the need and the opportunity to apply **thesauri** to information retrieval are greater than ever before. On the other hand, users **resist** most efforts to persuade them to apply one.” (Aitchison & Dextre Clarke 2004, quoted in Gilchrist 2006).
- the strongly-con: And so they should!
- the weakly-con: Understandably so.
- the weakly-pro: Well, it’s their loss ...
- the strongly-pro: Instruction is required!

The strongly-pro response

- search has become increasingly disintermediated
 - first, users were left to *search* for themselves
 - now, users are asked to *index* the documents they create
- “This is a very real **problem** that calls for action in the area of information literacy.” (Gilchrist 2006).

The weakly-pro response

- encourage taggers by ...
 - allowing taggers to **explain** why they've tagged in a particular way (bringing bias out into the open)
 - allowing taggers to give **feedback** on others' tagging decisions
 - automatically suggesting different **facets** of the resource for taggers to describe
 - automatically suggesting standard **forms** of tags (e.g., through co-occurrence analysis)

The strongly-con response

- So what?
- “In a world of information abundance, one usually just needs **good enough** responses. After all, if you find 250,000 photos of London, does it really matter that you missed another 50,000?” (Weinberger 2005).

Rationalism vs. empiricism

- user-tagging services may well be cheap and empowering, etc. ...
- ... but, **do they work?**
- **empirical** evaluations of **effectiveness** of retrieval from collections of user-tagged resources are rare ...
- ... then again, are such evaluations the most appropriate?

IR-oriented research questions

- cf. NIST MINDS workshop (Callan et al. 2007)
- “The scientific community’s responsibility is to discover **new knowledge** about how people seek, use, and organize information, and to develop **new tools** that assist people in achieving their **goals**.”
- “IR is an **empirical** science; the field cannot move forward unless there are means of **evaluating** the innovations devised by researchers.”
- but we need to develop new experimental evaluation **methodologies** to cope with new realities

New realities

- **scale**: the web is big
- information **heterogeneity**: more than formal texts
- information **quality**: often low
- **ubiquity**: IR tools are now everyday tools; their users are now everyone
- web search is still **unsuccessful**: “up to 50% of searches don’t result in a single click”
- **task diversity**: “Search is not the end goal. It is a tool that can help people accomplish other tasks ...”

Motivations of *searchers*

- “ ‘Finding a needle in a haystack’ can mean:
 - A *known* needle in a known haystack
 - A *known* needle in an unknown haystack
 - An *unknown* needle in an unknown haystack
 - *Any* needle in a haystack
 - The *sharpest* needle in a haystack
 - *Most* of the sharpest needles in a haystack

Motivations of *searchers*, cont'd

- *All* the needles in a haystack
- Affirmation of *no* needles in the haystack
- Things *like* needles in any haystack
- Let me know *whenever* a new needle turns up
- *Where* are the haystacks
- Needles, haystacks—whatever” (Koll 2000)

New goals, new methods

- we need to understand ...
 - how people use **everyday** IR tools (which we still know surprisingly little about)
 - how **task-contexts** shape these practices
 - how **individual** searchers' practices vary
 - how people **analyze/organize** already-found information

New goals, new methods, cont'd

- we need to develop new retrieval models that ...
 - incorporate **multiple sources of evidence** (e.g., about tags added incrementally by users over time)
 - support a wide range of solutions **tailored** for different genres of information, different groups of users, different types of task, different individuals
 - define “retrieval” to include **use** as well as discovery

New goals, new methods, cont'd

- we need to develop new evaluation methods that allow us to judge the success of systems that implement tagging services

Evaluating system success

- What are the **factors** that determine whether or not user-tagging services will be successful?
- **success:**
 - **popularity?**
 - the level at which a system **performs its functions?**
- “Does the success of the system depend only on participation in sheer numbers, or are there other factors [sic] such as group coherence and homogeneity that are required?” (Furnas et al. 2006)

Toward a framework for evaluation

- On what **criteria** may the performance of systems be judged?
 - What are the **goals** of system users?
 - What are the **functions** of systems?
 - What are the different **kinds** of user-tagging system?
- What methods may we use to **measure** the extent to which systems meet those criteria?

Criteria for evaluation

- **effectiveness:** relevance-based (recall/precision)
- **efficiency**
- **cost-effectiveness**
- **usability**

- i.e., how well, how quickly, how cheaply, how easily users can get their jobs done
- objective measures vs. **user satisfaction**

The impact of indexing quality on effectiveness

- What is the relationship between resource-description quality and retrieval effectiveness?
- What is the effect of ...
 - automatic vs. manual generation of terms?
 - derivation vs. assignation?
 - specificity and exhaustivity?
 - vocabulary control?
 - faceting and hierarchy?
- **indexer--searcher consistency** is key

Folksonomies

- most implementations of tagging are based on folksonomic model
- folksonomies = vocabularies that **emerge** from taggers' **uncontrolled** choices of resource descriptors
- reflect more accurately the actual consensus of opinion of end-users?

When taggers are searchers ...

- taggers are drawn from the same population from which searchers are drawn?
- taggers tend to use the same terms to tag resources as searchers use to look for resources?
- high levels of **tagger--searcher consistency** (and thus retrieval effectiveness) are assured?

Conclusions

- different kinds of user have different **motivations** for making use of tagging services
- different kinds of user have different perceptions of the **functions** of tagging services
- assessment of the success with which systems perform any of their multiple functions may be based on a variety of different **criteria**
- designs of evaluative tests must take these complexities into account

Thank you.

Jonathan Furner

Graduate School of Education & Information
Studies

University of California, Los Angeles

furner@gseis.ucla.edu