

# Twenty tall tales about tagging

**Jonathan Furner**  
**GSE&IS, UCLA**

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

**Getty/OCLC :: 09.17.09**

## 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).

*Indexed.* Peter Halley. 2007. Acrylic on canvas. 96 x 150 inches. <http://www.peterhalley.com/>.

## 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).

## 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 educational

## 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

## Tall tale #1

- tagging is **good!**
- [actually: we don't know]

## Tall tale #2

- tagging is **bad!**
- [actually: we don't know]

## Tall tale #3

- tagging is **new**
- [actually: it's online tagging services that are (relatively) new]

## Tall tale #4

- tagging is **different** from indexing
- [actually: tagging is a form of indexing]

## Tall tale #5

- folksonomies are **different** from indexing languages
- [actually: a folksonomy is a form of indexing language]

## Tall tale #6

- folksonomies are necessarily **unstructured**
- [actually: a folksonomy can be highly structured]

## Tall tale #7

- tagging is necessarily **uncontrolled**
- [actually: tagging can be tightly controlled]

## Tall tale #8

- tagging is for **resource description**
- [actually: tagging is for resource discovery]

## Tall tale #9

- tagging is **always** for resource discovery
- [actually: tagging is sometimes for learning]

## Tall tale #10

- every tagger is an expert, all the time
- [actually: every tagger is sometimes a novice]

## Tall tale #11

- the crowd is always right
  
- [actually: the crowd is sometimes wrong]

## Tall tale #12

- individual crowdmembers act independently of one another
- [actually: individual crowdmembers are influenced by prior activity]

## Tall tale #13

- good tagging can only be done by **novices**
- [actually: good tagging can be done by experts]

## Tall tale #14

- some taggers have no expertise
- [actually: every tagger is sometimes an expert]

## Tall tale #15

- good tagging is **accurate** tagging
- [actually: good tagging is effective tagging]

## Tall tale #16

- **effective** tagging is accurate tagging
- [actually: effective tagging is consistent tagging]

## Tall tale #17

- the consistency that's important is **inter-tagger consistency**
- [actually: the consistency that's important is tagger–searcher consistency]

## Tall tale #18

- tagger–searcher consistency is **necessarily** high because (i) taggers are searchers, and (ii) people use the same kinds of terms *to tag* as they do *to search*
- [actually: we don't know that people use the same kinds of terms to tag as they do to search]

## Tall tale #19

- **past searching activity** is necessarily the best guide to future searching activity
- [actually: the best guide might be activity of some other kind]

## Tall tale #20

- there are twenty tall tales
- [actually: ... uh ...]

## Who can be taggers?

- catalogers / indexers
- machines
- authors / creators
- readers / viewers
  
- ... all **experts** in some respect

## How do we identify good tags?

hypothesis:

- **good tags** are those supplied by **good taggers**
- good taggers are those deemed to have supplied good tags in the past
- i.e., good taggers are those who have a good **reputation** for tagging

## How can we measure reputation?

- by asking **the crowd**, or ...
- by asking **those with good reputations**
  
- by discriminating among **individuals**, or ...
- by discriminating among **groups**

## 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

## Conclusions, cont'd

- different kinds of user/expert can make different kinds of **contribution** to tagging services
- the best implementations will give the widest range of different user/expert groups the fullest opportunities to make their best contributions
- the big challenge:  
providing sufficient **incentives** to tag ...  
while keeping **costs** down

**Thank you.**

**Jonathan Furner**

Graduate School of Education & Information  
Studies

University of California, Los Angeles

[furner@gseis.ucla.edu](mailto:furner@gseis.ucla.edu)