

**Margaret E.I. Kipp**  
**College of Information and Computer Science**  
**Long Island University**  
**margaret.kipp@gmail.com**  
**<http://publish.uwo.ca/~mkippp/>**

# Patterns in Tagging:

## Collaborative Classification Practices in Social Bookmarking Tools

ASIST 2007, Milwaukee, Wisconsin

# Social Bookmarking Sites

## For anyone

- del.icio.us
  - bookmark anything (e.g. web page, PDF, video clip, audio, etc)
  - owned by Yahoo!
  - <http://del.icio.us/>

## For academics

- citeulike
  - focus on journals and academic works
  - <http://citeulike.org/>
- connotea
  - accepts non academic links
  - owned by Nature
  - <http://connotea.org>

# Study 1: Del.icio.us Study

- study tag usage for highly tagged sites
- examine convergence of tag usage
- co-occurrence analysis for co-used tags

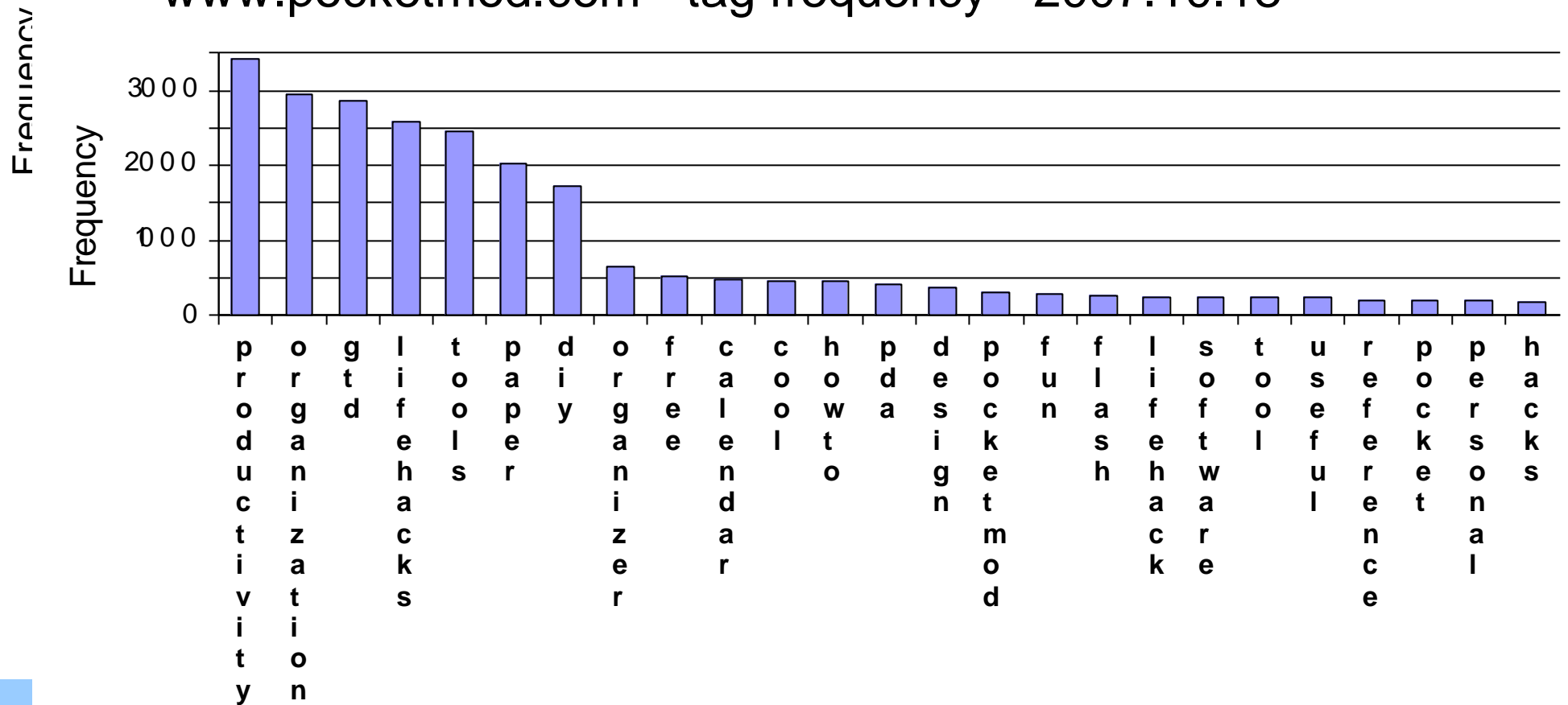
The screenshot shows the del.icio.us website in a Mozilla Firefox browser window. The page features the del.icio.us logo, navigation links (popular, recent, login, register, help), and a 'hotlist' section. The 'hotlist' section displays a table of trending items with their titles, save counts, and tags.

| Category      | Item  | Save Count | Tags   |
|---------------|---|------------|--|
| hot right now | 1. <a href="#">Linux.com   CLI Magic: Searching with find</a> save this   | 110 people | linux find tutorial cli search tags          |
|               | 2. <a href="#">Free Science and Video Lectures Online! Even more physics videos and video lectures!</a> save this | 88 people  |  |
|               | 3. <a href="#">TheTypeTrust</a> save this   | 71 people  |  |
| 1 hour ago    | 1. <a href="#">Strobist: How To: DIY \$10 Macro Photo Studio</a> save this  | 106 people | photography diy lighting tutorial howto tags |
|               | 2. <a href="#">Run a Secure Web Proxy in 10 Steps - Pure Mhz Community</a> save this                              | 89 people  |  |
|               | 3. <a href="#">Technophilia: Find great podcasts - Lifehacker</a> save this                                       | 91 people  |  |

# Tag Convergence (frequency)

www.pocketmod.com - tag frequency - 2006.02.01

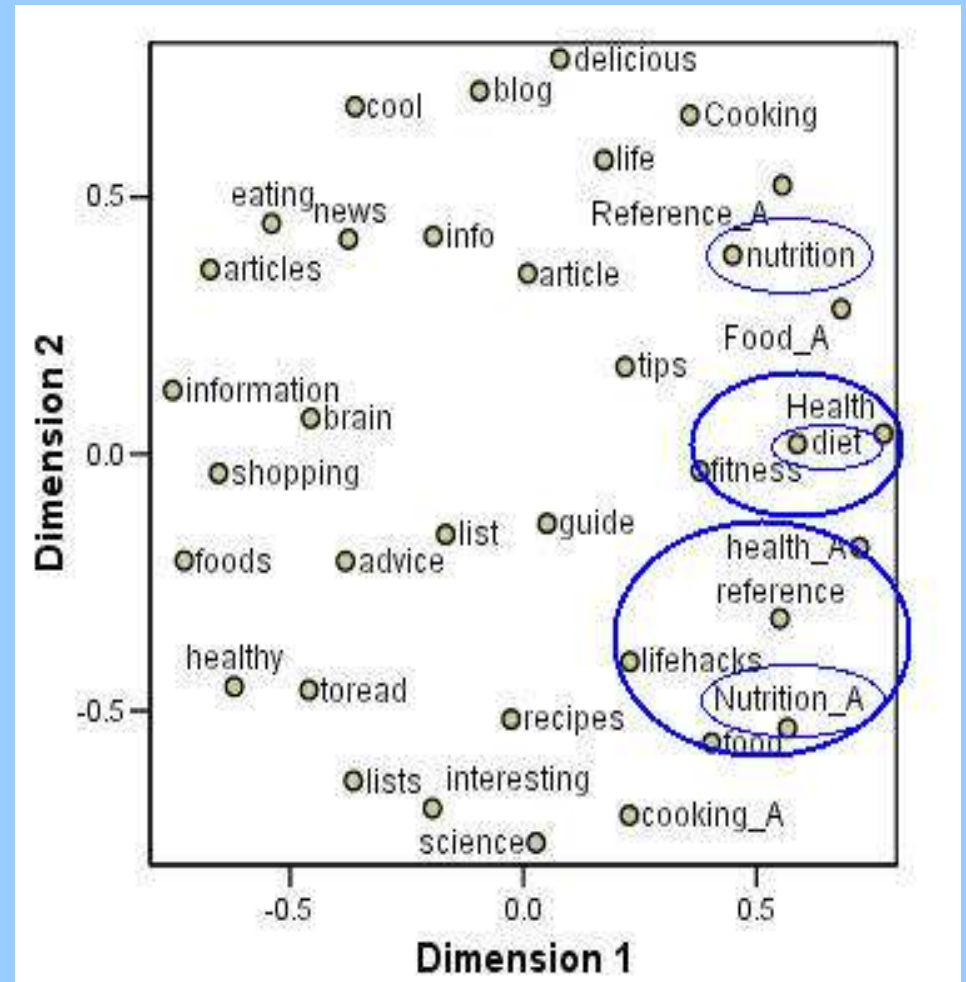
www.pocketmod.com - tag frequency - 2007.10.18



.first 7 tags have remained constant over time

# Co-tag Graphs (MDS)

- many user groups; many differing priorities
- some co-word graphs in del.icio.us showed clusters of what might be user groups



# Study 2: CiteULike Study

CiteULike: A free online service to organise your academic papers - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.citeulike.org/

Register | Log in | FAQ |

Article title  Search

## CiteULike

### Navigation

- Home
- Log in
- Register
- Discussion list

### Journals

- Browse current issues

### Groups

- View group

### Experimental Features

- Import from BibTeX

## All about CiteULike

Read all about it

CiteULike is a free service to help academics to share, store, and organise the academic papers they are reading. When you see a paper on the web that interests you, you can click one button and have it added to your personal library. CiteULike automatically extracts the citation details, so there's no need to type them in yourself. It all works from within your web browser. There's no need to install any special software.

Because your library is stored on the server, you can access it from any computer. You can share your library with others, and find out who is reading the same papers as you. In turn, this can help you discover literature which is relevant to your field but you may not have known about.

You're currently looking at a list of the last few papers submitted by all the CiteULike users. Why not [register](#) for a free account today and start organising your collection and see just the articles you're interested in? All we need is your email address, a username, and a password. It should take less than fifteen seconds.

Alternatively, you can read [more about CiteULike](#).

## Everyone's library

Some recent papers posted to CiteULike - all mixed together.

- [Homogeneous fluorescent assay for RNA polymerase.](#)

### Everyone's Tags

Most active tags on CiteULike

Filter:

adaptation agent ai  
algorithm algorithms  
analysis architecture  
attention bayesian  
bioinformatics biology book  
brain business cancer  
classification clustering  
cognition collaboration  
collaborative  
communication  
community complexity  
culture database design  
development digital  
disease dynamics  
economics education  
evolution expression  
fMRI folksonomy games  
gender gene genetics  
geometry graph hci  
history human  
information interaction  
knowledge language

- study tag use and types of tags on articles compared to author keywords and subject headings

# Author, User, Intermediary Term Comparison

- comparison done using standard thesaurus terms (synonym, NT, BT, RT, etc.)
- most common relationship between three groups of terms was "related but not in the thesaurus"
- next most common RT and then equivalence

# User Vocabulary Length

- measure of how many unique tags each user used
- highest number of unique tags used for an item: 18 (min. 1, median 2)
- highest number of unique tags used by a single user: 66 (min. 1, median 4)
- generally connection between high user vocabulary and heavy posting (> 25 articles)

# Commonalities Between Studies

- similarities in common tag characteristics (e.g. spelling variations, sesquipedalians, acronyms, etc)
- use of affective tags (e.g. cool, interesting) and time and task related tags (e.g. @toread, todo) in both studies
  - ~ 16% of tags in Del.icio.us study
  - ~ 1 tag per article in CiteULike study not directly subject related

# Tags that are like Keywords

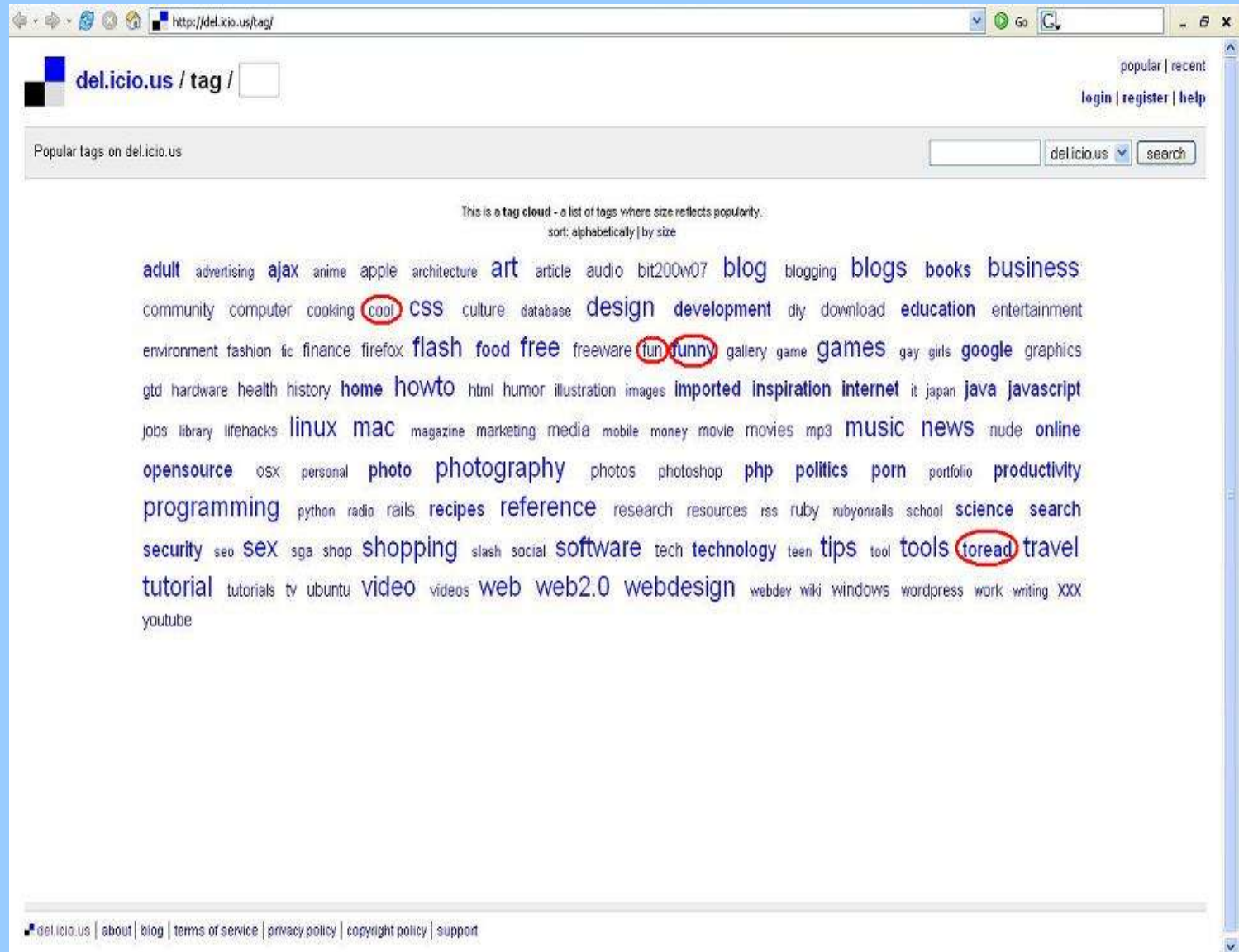
- Subject
  - information\_seeking\_behaviour
  - copyright
  - information\_literacy
- Type/Format
  - conference
  - article
- Geographic/spatial
  - country names

# Tags that are not like Keywords

- time and task management tags
  - toread, 2read, to-read, alire (French for toread)
  - todo, to-do
  - lis510 (course/project codes)
  - toprint
  - maybe
- affective tags
  - fun
  - cool
  - interesting
- other tags
  - new
  - no-tag, system:unfiled
  - the, and

# Study 3: Non Subject Tag Study

- study non subject related tags
- how are such tags used with subject tags?



# Non Subject Tags

- some time and task or affective tags are very popular
  - cool, fun, funny, toread appeared in main del.icio.us tag cloud
- ToRead and fun are popular tags on all three sites
- affective terms appear more frequent on Citeulike and Connotea than expected
  - biology articles more often listed as toread; math and physics as fun

# Discussion

- tagging has all the problems of free text search/indexing
- distributed user tagging demonstrates need for examination of how users associate groups of documents (time and task management)
- tag groups tend to converge on a useful set of terms after a threshold number of users (what is the threshold?) but new terminology may alter this profile

# Discussion 2

- users use some terminology which is rare or completely absent from subject heading lists (e.g. time and task tags)
- user terms often not part of formal thesaurus
- user vocabulary length is often directly related to the number of items bookmarked (frequent users use more tags?)

# Discussion 3

- What is the effect of personal and subjective terms such as cool, fun and toread in a social bookmarking system?
- What happens when these terms are aggregated?
- Amazon and Google use personal information to generate popularity or relevance indicators, do non subject tags offer any similar advantages?

**Margaret E.I. Kipp**  
**College of Information and Computer Science**  
**Long Island University**  
**margaret.kipp@gmail.com**  
**<http://publish.uwo.ca/~mkipp/>**

**Thank you!**

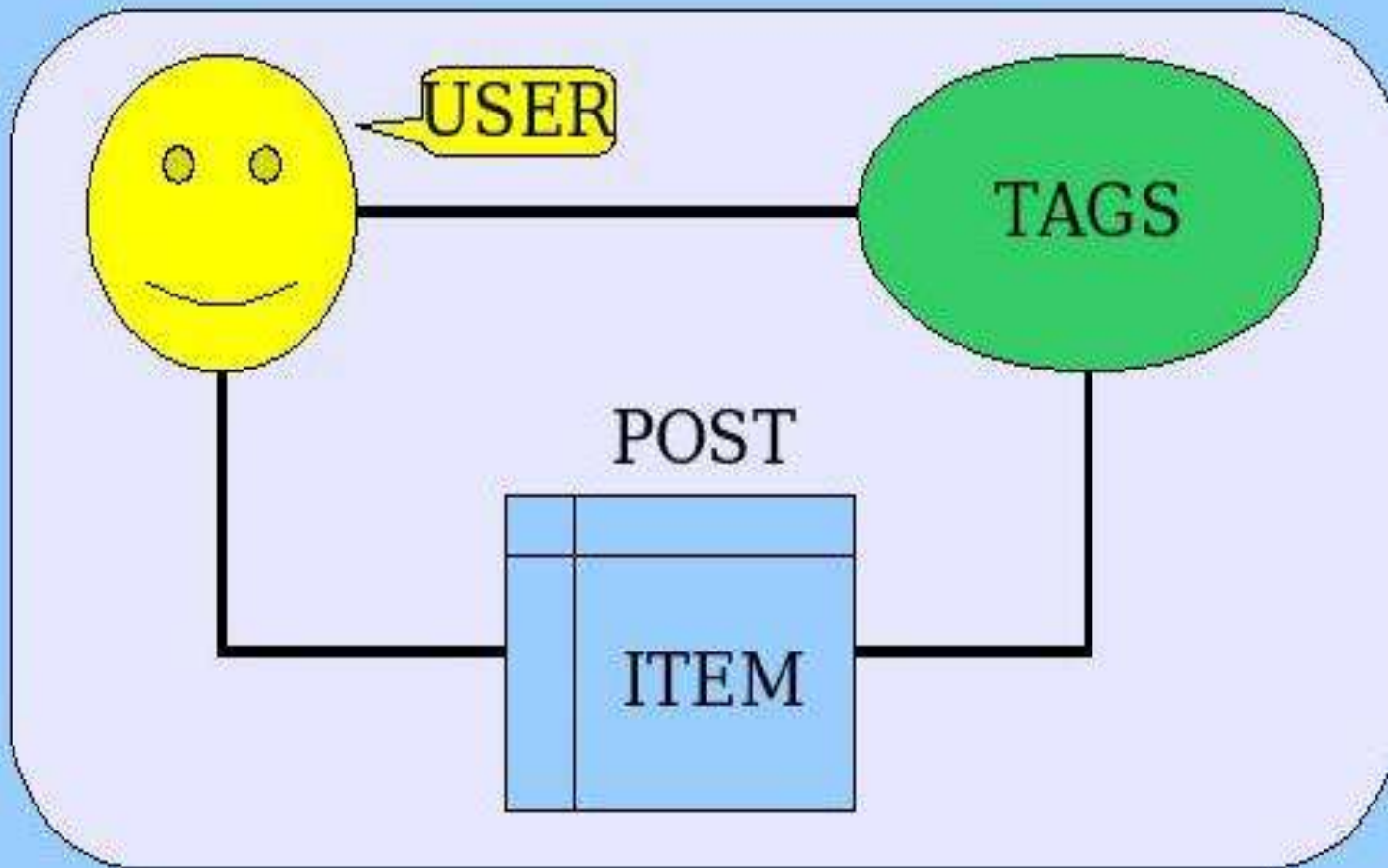
ASIST 2007, Milwaukee, Wisconsin

# Supplemental Material

# Social Bookmarking and Tagging

- Social Bookmarking:
  - sharing bookmarks, articles, etc.
  - association of tags (keywords) with links
  - tags and articles joined into networks of related terms
  - users encouraged to share bookmarks and tags with others
- Tagging:
  - associating a term with a link or article
  - labelling or classifying for personal use

# Social Bookmarking



Social bookmarking creates a relationship between a user, an item and a set of tags.

# Motivation

- Are categories emerging in social tagging that will complement those developed through professional methods?
- What does tag convergence and co-word usage suggest about the utility of tagging?
- What implications do the use of affective or time and task related tags have for the organisation of information?