

@toread and Cool: Tagging for Time, Task and Emotion

Margaret E. I. Kipp
Faculty of Information and Media Studies
University of Western Ontario
margaret.kipp@gmail.com

8th Information Architecture Summit, Las Vegas, Nevada. March 22-27, 2007

Background

- My research examines:
 - how people organise things on the web
 - how this compares to traditional library classification techniques
- Specific points of interest:
 - structures and the creation of structures in classification systems
 - relationship between personal information management and classification

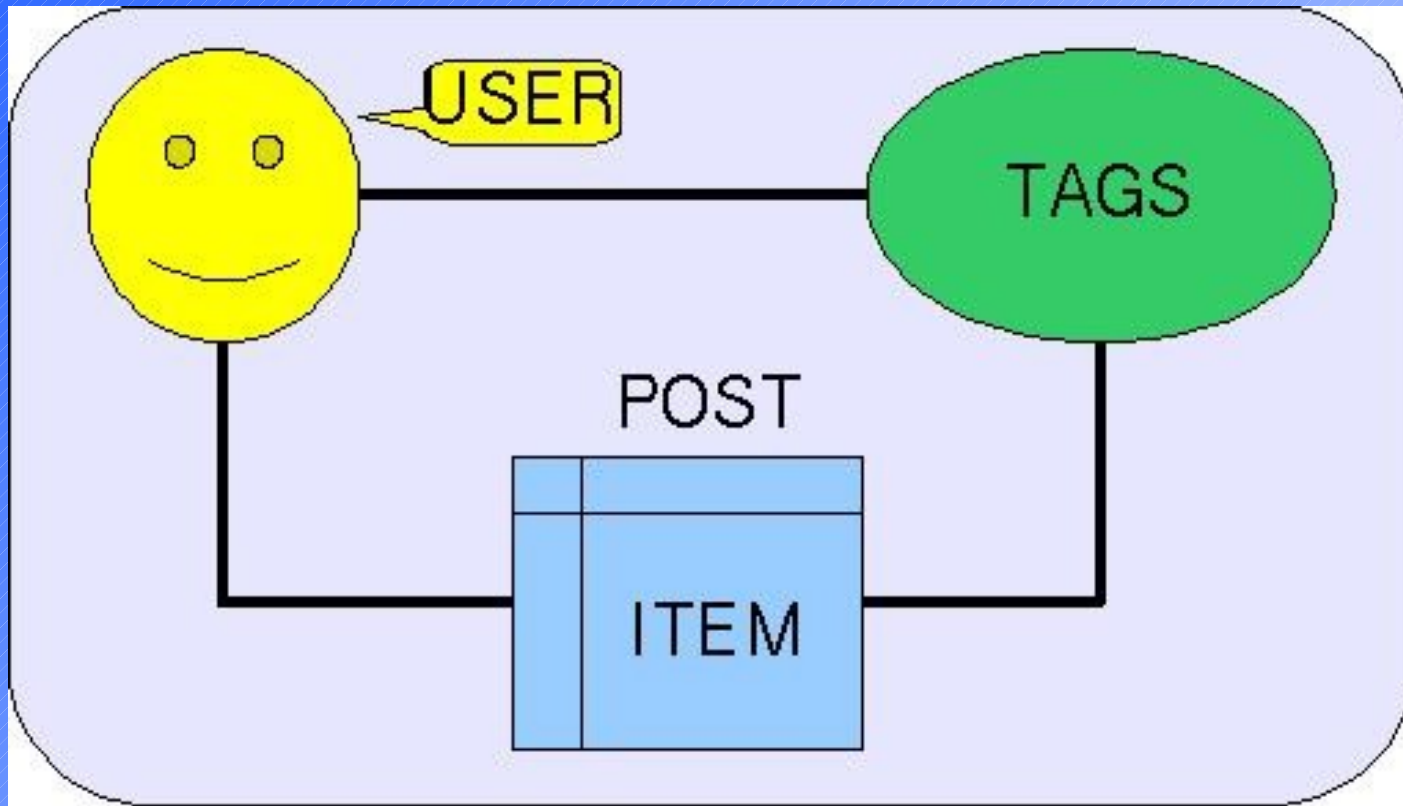
What is Tagging?

- the act of associating a term with a link or article
- labelling or classifying for personal use
- act of generating a dynamic taxonomy or folksonomy
- Related definitions:
 - folksonomy - user generated taxonomy of related tags
 - tag cloud - tag display where size equals popularity

What is Social Bookmarking?

- public sharing of links
 - association of tags (keywords) with links
- network of related links created by users
 - network of related tags created by users
- site for sharing bookmarks, articles, etc.
 - tags and articles are joined into networks of related terms
 - users are encouraged to share bookmarks and tags with others

A Social Bookmarking Post



A post is a relationship between a user, an item and a set of tags.

del.icio.us

popular | recent

login | register | help

- » **keep**
your favorite websites, music, books, and more in a place where you can always find them.
- » **share**
your favorites with family, friends, and colleagues.
- » **discover**
new and interesting things by browsing popular & related items.
[Learn more »](#)

discover favorites:

» sign up now

username

password

password again

email

What's a tag?
A tag is just a word that describes an item saved on del.icio.us.
[Learn more »](#)

Connotea

the del.icio.us hotlist also see: [popular](#) | [recent](#) | [tags](#)

- hot right now
1. [Linux.com | CLI Magic: Searching with find](#) [save this](#) **110** people
first posted by plovs ... on july 10 [linux](#) [find](#) [tutorial](#) [cli](#) [search](#) [tags](#)
 2. [Free Science and Video Lectures Online! Even more physics videos and video lectures!](#) [save this](#) **88** people
 3. [TheTypeTrust](#) [save this](#) **71** people
-
- 1 hour ago
1. [Strobist: How To: DIY \\$10 Macro Photo Studio](#) [save this](#) **106** people
first posted by yeltsinfina ... on july 10 [photography](#) [diy](#) [lighting](#) [tutorial](#) [howto](#) [tags](#)
 2. [Run a Secure Web Proxy in 10 Steps - Pure MHz Community](#) [save this](#) **89** people
 3. [Technophilia: Find great podcasts - Lifehacker](#) [save this](#) **91** people

Site Details

For anyone

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

Specialised for academics

- citeulike
 - focuses on journals and academic books
 - <http://citeulike.org/>
- connotea
 - accepts non academic links
 - <http://connotea.org>

Previous Studies

- Study 1: Del.icio.us
- delicious tag usage on highly tagged sites
- examination of convergence of tag usage
- co-occurrence analysis for co-used tags
- Study 2: Citeulike
- citeulike tag usage compared to author keywords and subject headings from journal databases
- examine types of tags and more traditional index terms

Non Subject Tags

- Study 1: Del.icio.us
- Affective Tags
 - cool: 906 occurrences
- Time and Task Tags
 - toread: 939 occurrences
- 3049 unique tags identified as time and task (16%)

Del.icio.us Tag Cloud

del.icio.us / tag /

popular | recent
login | register | help

Popular tags on del.icio.us del.icio.us

This is a **tag cloud** - a list of tags where size reflects popularity.
sort: [alphabetically](#) | [by size](#)

adult advertising **ajax** anime apple architecture **art** article audio bit200w07 **blog** blogging **blogs** books business
community computer cooking **cool** **CSS** culture database **design** **development** diy download **education** entertainment
environment fashion fic finance firefox **flash** **food** **free** freeware **fun** **funny** gallery game **games** gay girls **google** graphics
gtd hardware health history **home** **howto** html humor illustration images **imported** **inspiration** **internet** it japan **java** **javascript**
jobs library lifehacks **linux** **mac** magazine marketing media mobile money movie movies mp3 **music** **news** nude **online**
opensource OSX personal **photo** **photography** photos photoshop **php** **politics** **porn** portfolio **productivity**
programming python radio rails **recipes** **reference** research resources rss ruby rubyonrails school **science** **search**
security seo **sex** sga shop **shopping** slash social **software** tech **technology** teen **tips** tool **tools** **toread** travel
tutorial tutorials tv ubuntu **video** videos web web2.0 **webdesign** webdev wiki windows wordpress work writing XXX
youtube

del.icio.us | [about](#) | [blog](#) | [terms of service](#) | [privacy policy](#) | [copyright policy](#) | [support](#)

Non Subject Tags 2

- Study 2: Citeulike
 - average of 3.5 per article, not all from tags
- Categories:
 - Time and Task Management (@toread)
 - Geographic Tags (berkeley)
 - Specifics (pubmed-mining)
 - Generalities (information science)
 - Emergent Vocabulary (folksonomy)
 - Other (system defined placeholder no-tag)

Motivations

- Previous studies showed use of non subject tags (in del.icio.us study they were at least 16% of the sample!)
- What is the purpose of these tags?
- Are they only individual or is there any collective useful information here?

Research Questions

- What patterns of user tagging activity emerge on examination of affective or time and task related tags?
- How do users use time and task related tags or affective tags to indicate the value they see in a document?
- What implications do the use of affective or time and task related tags have for the organisation of information?

Methodology

- Data source
 - Del.icio.us, Citeulike, Connotea
- Collection times
 - October 20-31st, 2006
- Collection method:
 - python scripts
- Data collected
 - all items tagged with a specific list of 83 affective and time and task related tags

Examples of Tags Collected

- Affective Tags

- cool
- boring
- exciting
- important
- funny
- strange
- favorite/favourite

- Time and Task Tags

- @toread
- todo
- tobuy
- toblog
- tovisit
- @pending
- todescribe

Data Collected

- all posts tagged with one of 83 specific tags
 - 48 tags were time and task related, 30 were affective tags and the remaining 5 consisted of prepositions and conjunctions
 - 78 tags were in English
 - 5 tags were in French (lire, alire, @lire, acheter, amusant)
 - non English tags do not yet appear frequently in the popular tag clouds of these sites

Data Collected

- total of 203 352 posts collected
 - 1831 posts were collected from Citeulike
 - 2891 from Connotea
 - 198630 from Del.icio.us
- without zeros, total of 107 584 posts
 - 1805 for Citeulike
 - 2462 for Connotea
 - 107 584 for Del.icio.us

General Results

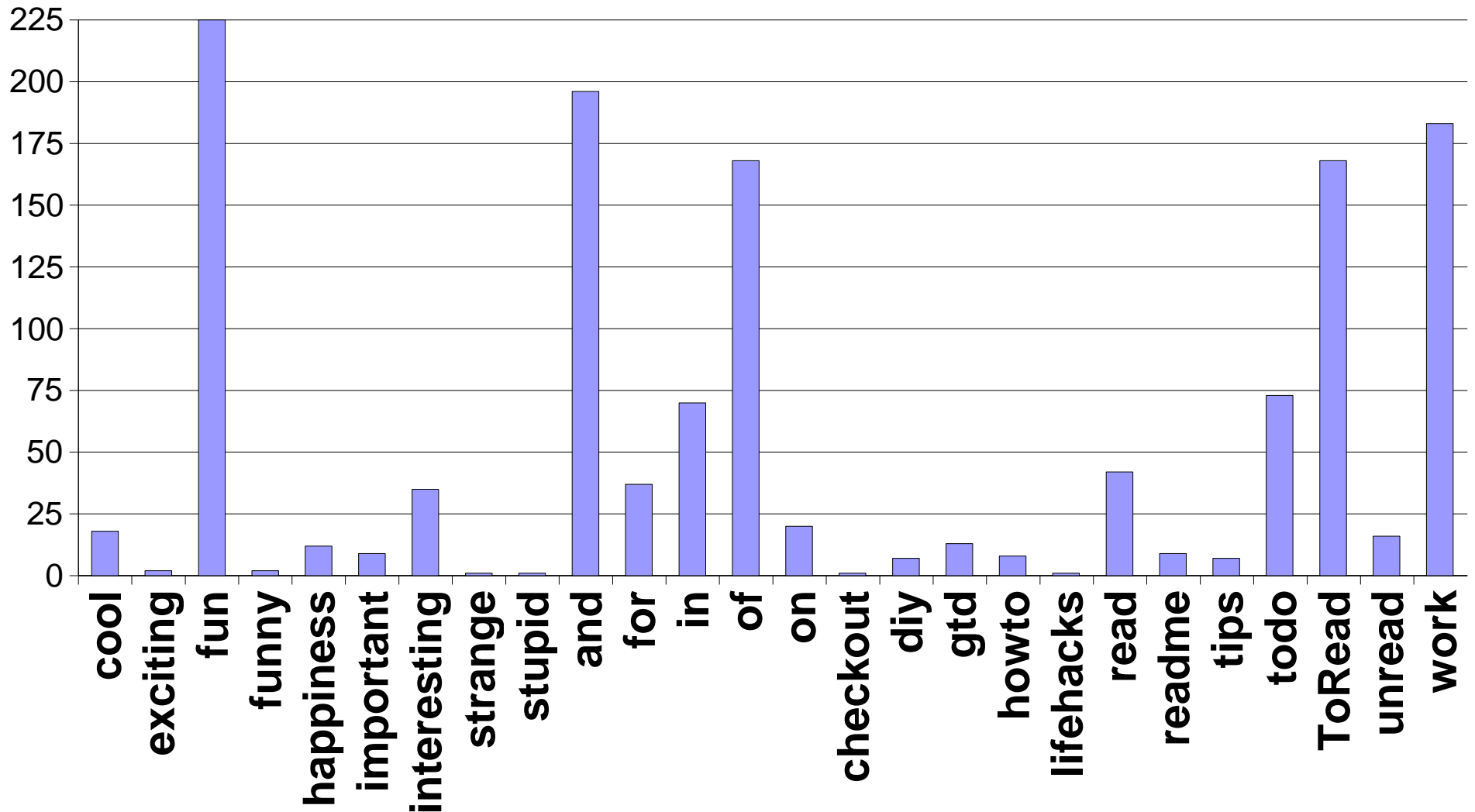
- some time and task or affective tags are very popular
 - cool, fun, funny, toread appeared in main del.icio.us tag cloud
- affective terms appear more frequent on Citeulike and Connotea than Del.icio.us when normalised for the size of the respective populations

General Results 2

- ToRead and fun are popular tags on all three sites
- prepositions and conjunctions were surprisingly popular (and appears in the citeulike tag cloud and of in the connotea tag cloud)
 - examination of the tag lists from which these tags occur suggests that the use of these prepositions is by people who are using phrases to tag an item instead of individual words or meaningful compound words

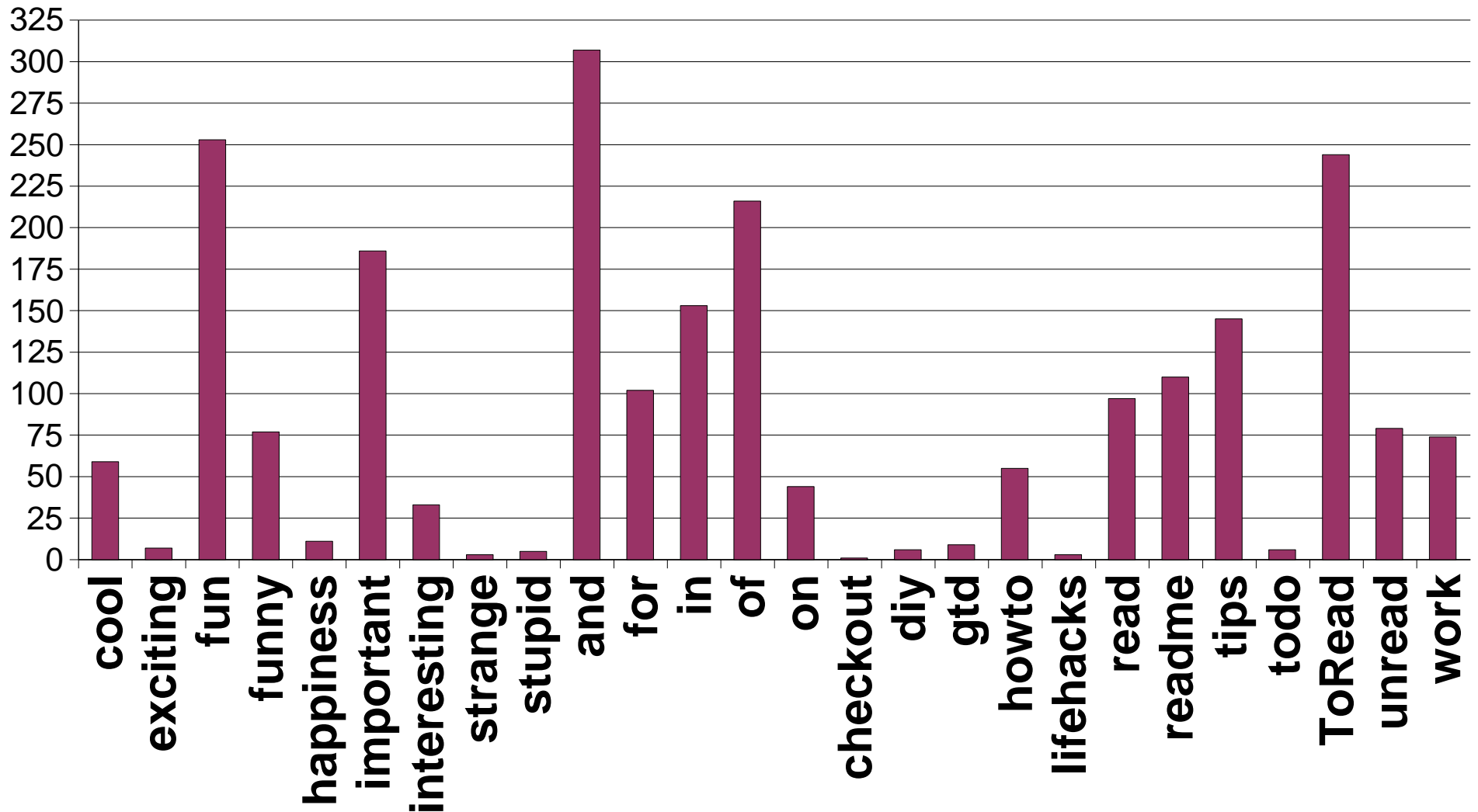
Popular Non-Subject Tags

Non Subject Tags - Citeulike



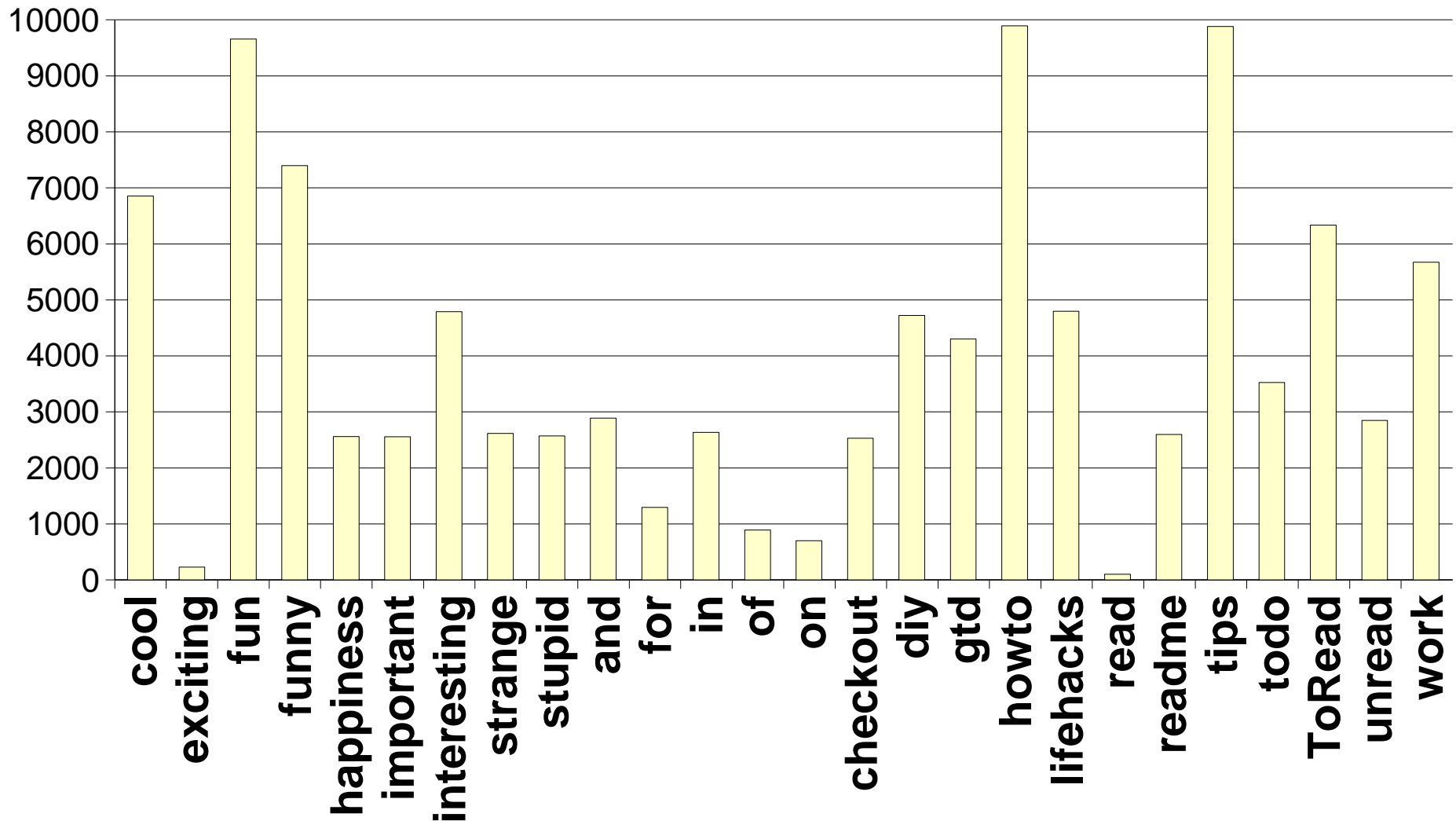
Popular Non-Subject Tags

Non Subject Tags - Connotea

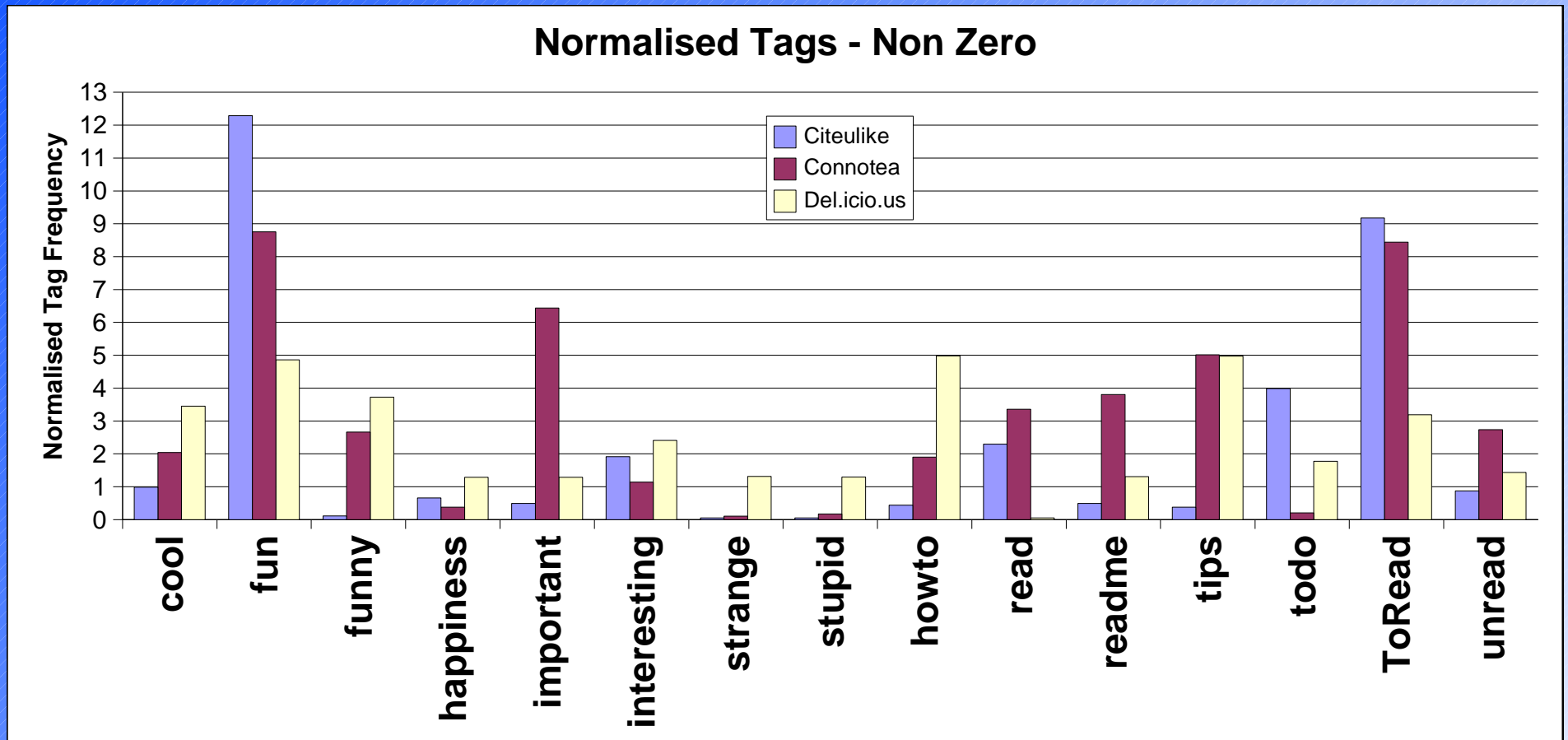


Popular Non-Subject Tags

Non Subject Tags - Del.icio.us



Comparing Tag Frequency



Popular non subject related tags compared by site. Tag counts have been normalised to allow comparison.

Time and Task Tags

- many time and task related tags are variations on toread
 - @toread
 - @read
 - readlater
 - unread
- fewer variations for toread are found on citeulike and connotea
- citeulike offers a way to mark interest in reading an article

Time and Task 2

- is the toread tag useful to other users?
- Amazon's recommendation system relies on purchase data as an indicator of interest
- could a toread tag have a similar function?
- could this function like a colleague's e-mail pointing to an article or book?

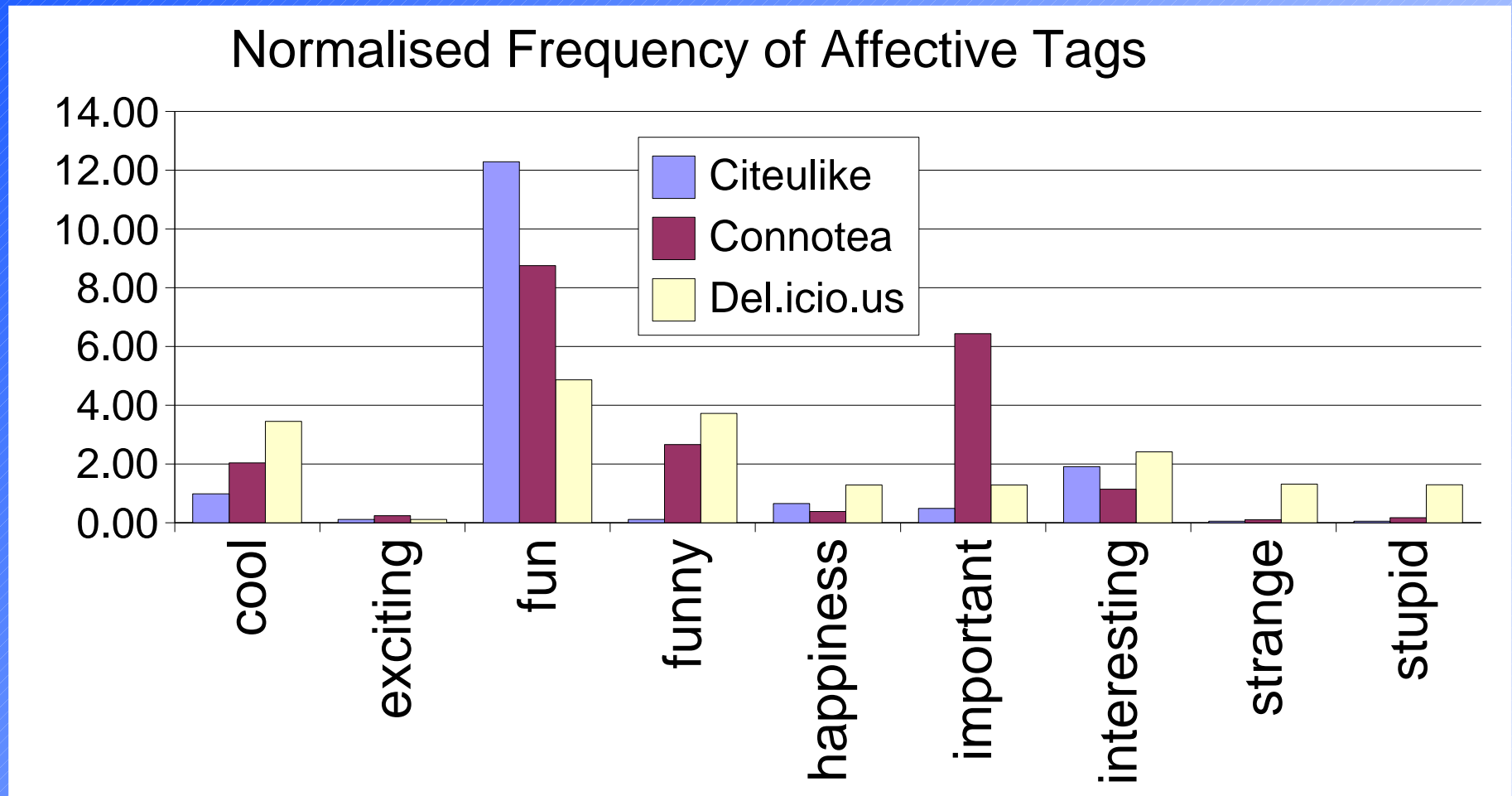
Personal Information Management

- time and task related tags suggest a connection to research into personal information management
- research into how people organise their documents shows the use of contextual project information for classification
- project and task related tags were found in previous studies
 - e.g. tags which appeared to be course codes (lis501)

Affective Tags

- affective tags represent an emotional reaction to an item
 - cool
 - fun
 - strange
- tags such as cool or fun do not appear to add anything to the subject classification of an item
- seem to be poor candidates for search terms for information retrieval

Affective Tags 2



fun: Citeulike - 225, Connotea - 253, Del.icio.us - 9659

Non Subject Tags with Subject Tags

- non subject tags were frequently used with subject related tags
- academic articles on citeulike and connotea were tagged with terms such as fun and cool
- what do affective tags add to a tag list?

Non Subject Tags: Citeulike

- Title: Symmetry and Self-Organization in Complex Systems
- URL: <http://arxiv.org/pdf/cond-mat/0609274>
- taglist: automata, **fun**, graphs, mathematics, networks, statistical-mechanics, symmetry

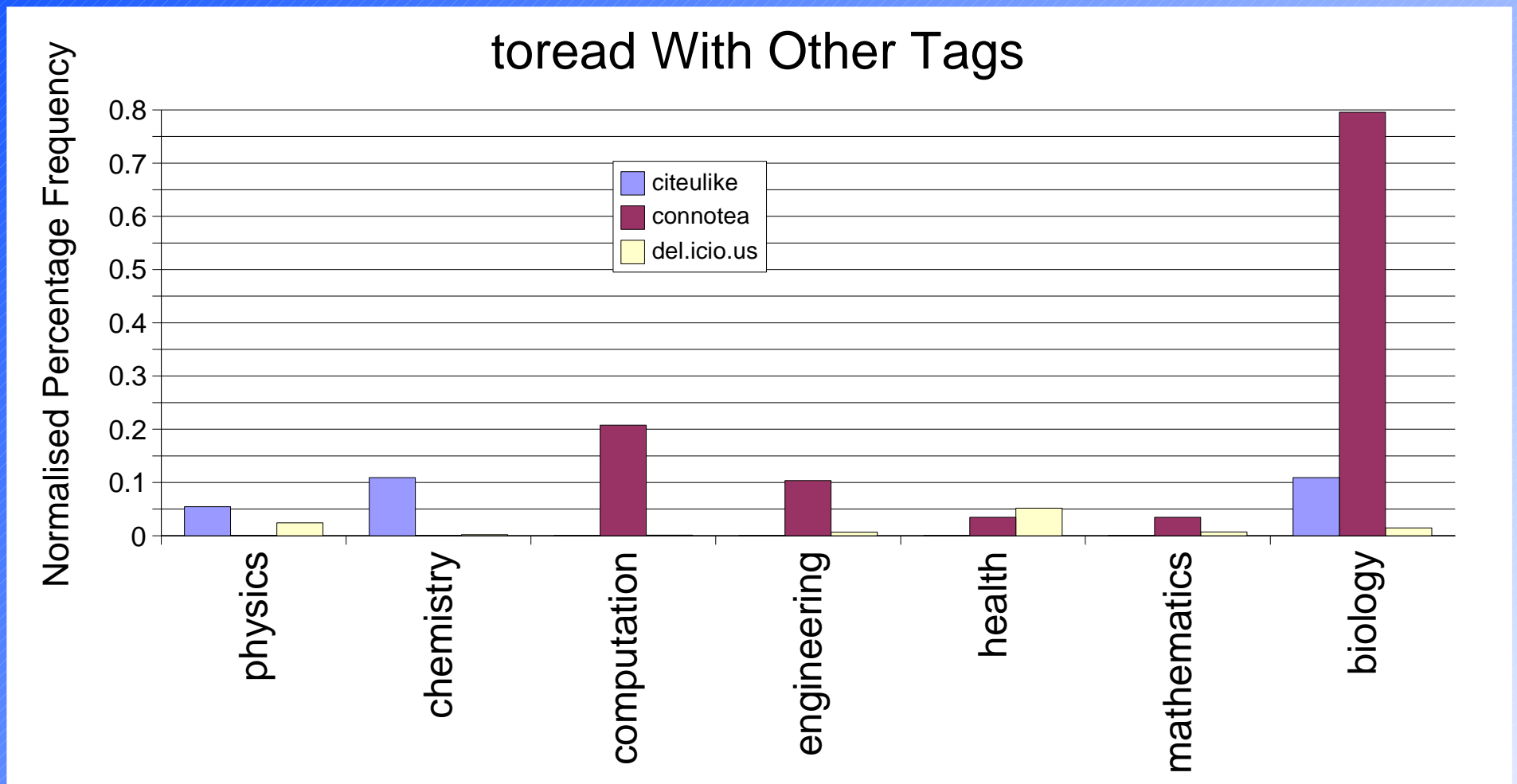
Non Subject Tags: Connotea

- Title: Foundations for engineering biology
- URL:
<http://www.nature.com/nature/journal/v438/n7067/full/nature04342.html>
- Journal: Nature 438(7067): 449-453
(doi:10.1038/nature04342)
- taglist: complex systems, network, systems biology, synthetic biology, comics, **fun**

Non Subject Tags: Del.icio.us

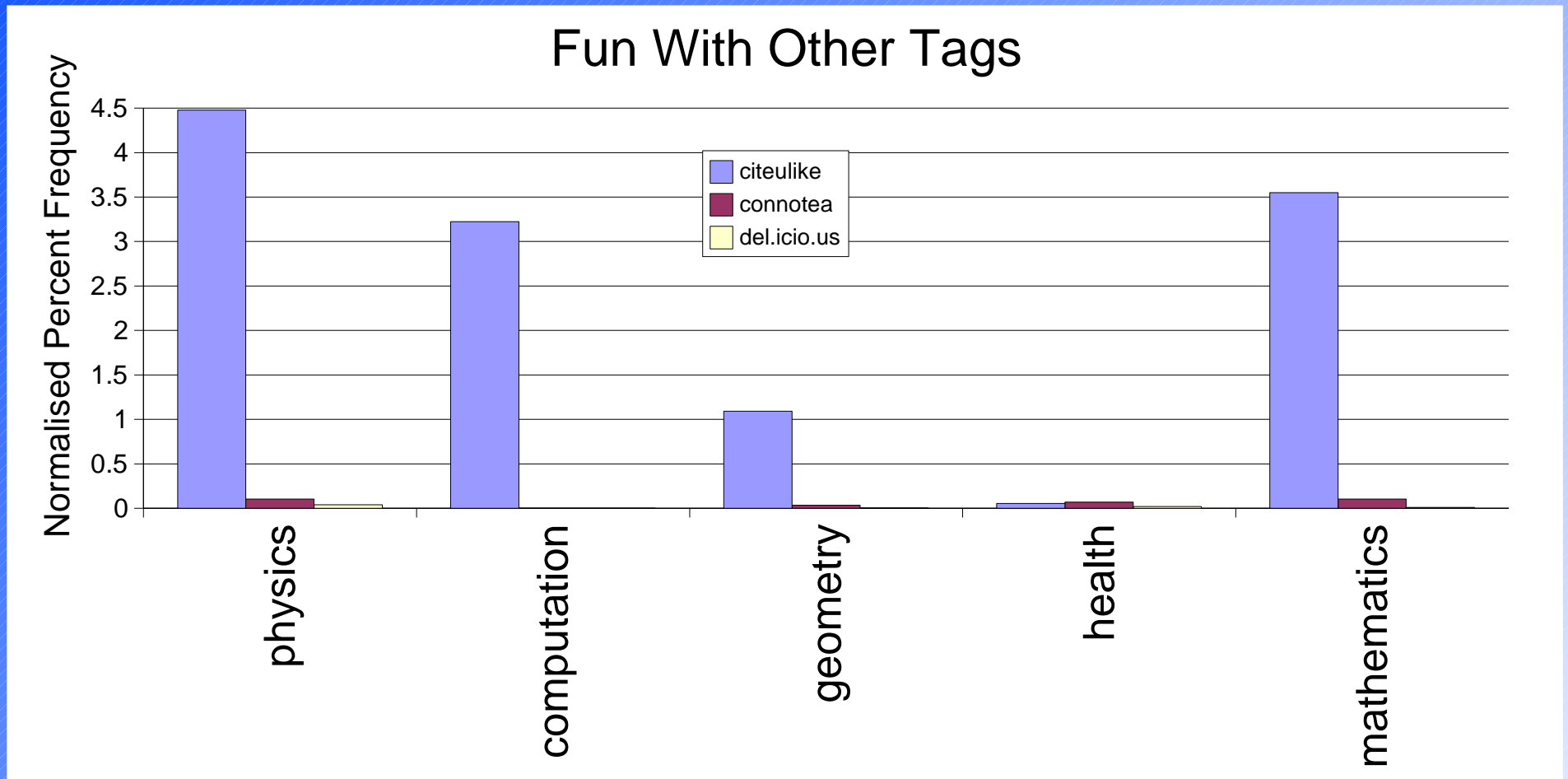
- Del.icio.us:
- Title: 36 Humorous Proof Methods
- URL:
<http://www.themathlab.com/geometry/funnyproofs.htm>
- taglist: **fun**, humor, math, proof

toread with Subject Tags



biology: citeulike - 2, connotea - 23, del.icio.us - 29;
mathematics: citeulike - 0, connotea - 1,
del.icio.us - 14

fun with Subject Tags



physics: citeulike - 82, connotea - 3, del.icio.us - 82;
mathematics: citeulike - 65, connotea - 3,
del.icio.us - 20

Discussion

- non subject tags are intrinsically time-sensitive
- express response from user not subject of document
- suggest active engagement with the text
- show that user links perceived subject matter to:
 - specific task
 - specific set of interests
 - specific emotional reactions

Discussion 2

- use of time and task or affective tags shows that tagging expresses a dynamic relationship between users and documents, suggesting possible new ways of modelling information access
- research into personal information management systems show users classify by task and project as well as by subject

Final Thoughts

- 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
Faculty of Information and Media Studies
University of Western Ontario
margaret.kipp@gmail.com
<http://publish.uwo.ca/~mkipp/>

Thank you!

Questions?