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

Margaret E. I. Kipp <mkipp@uwo.ca>

Faculty of Information and Media Studies, University of Western Ontario, Canada

Introduction

Social Classification or tagging is increasingly a subject of interest in library and information science (and related fields) as social bookmarking tools such as del.icio.us and flickr have become more popular. Some argue that using simple visualisations of tags such as sorting tags by frequency or displaying tag clouds in which tag size denotes popularity illuminates the formation of folksonomies (taxonomies of related tags). Others argue that the ample evidence of such symptoms of mob indexing as spelling variations and lack of synonym or vocabulary control show that such systems will never replace conventional indexing systems. Previous studies of Del.icio.us (Kipp and Campbell 2006) and Citeulike (Kipp 2006) determined that while many common tags are subject related and may form a reasonable set of "good enough" indexing terms, many other common tags are not directly subject related but are in fact affective tags dwelling on a user's emotional response to a document or are time and task related tags. These non subject related tags are interesting because they are traditionally excluded from classification systems.

Methodology

This study examines the use of non subject related tags in social bookmarking tools. These tags fall into two major categories:

Affective tags

Time and Task Related tags

Data was collected from del.icio.us, citeulike and connotea via python scripts designed to gather information on all posts related to specified tags. Posts in a social bookmarking tool consist of, at minimum, a title, URL and associated user name. A majority of posts (94% in Kipp and Campbell 2006) have associated tags. A minority of posts contain a written description or note. For each tag in the list, the python scripts collect all posts which have been tagged with this tag from each of the three tools. Examples of affective tags include interesting, fun and cool. Examples of time and task related tags include @toread, todo, and tobuy.

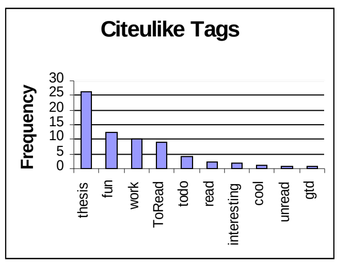
Results

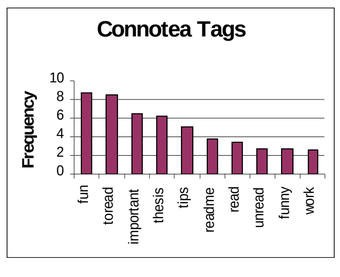
A total of 83 tags were examined in this study: 48 fell into the category of time and task, 30 were affective and the remaining 5 consisted of the prepositions for, on, in and of and the conjunction and. A majority (78) of the tags were in English; 5 tags were in French.

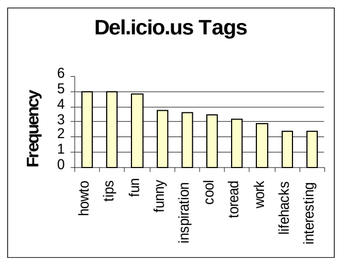
A total of 1831 posts were collected from Citeulike, 2891 from Connotea and 198630 from Del.icio.us. This gives a total of 203352 posts in all from all three sites. Since the number of posts obtained from Del.icio.us is several orders of magnitude larger than the other two sites, data was normalised by total posts per site for comparisons.

A number of the tags in this study are very popular and appear on the respective popular or frequently used tag cloud pages for their sites. As of October 31st, 2006, the tags 'cool', 'daily', 'fun', 'funny', 'toread', 'work', 'and' and 'of' appear in the popular tag clouds.

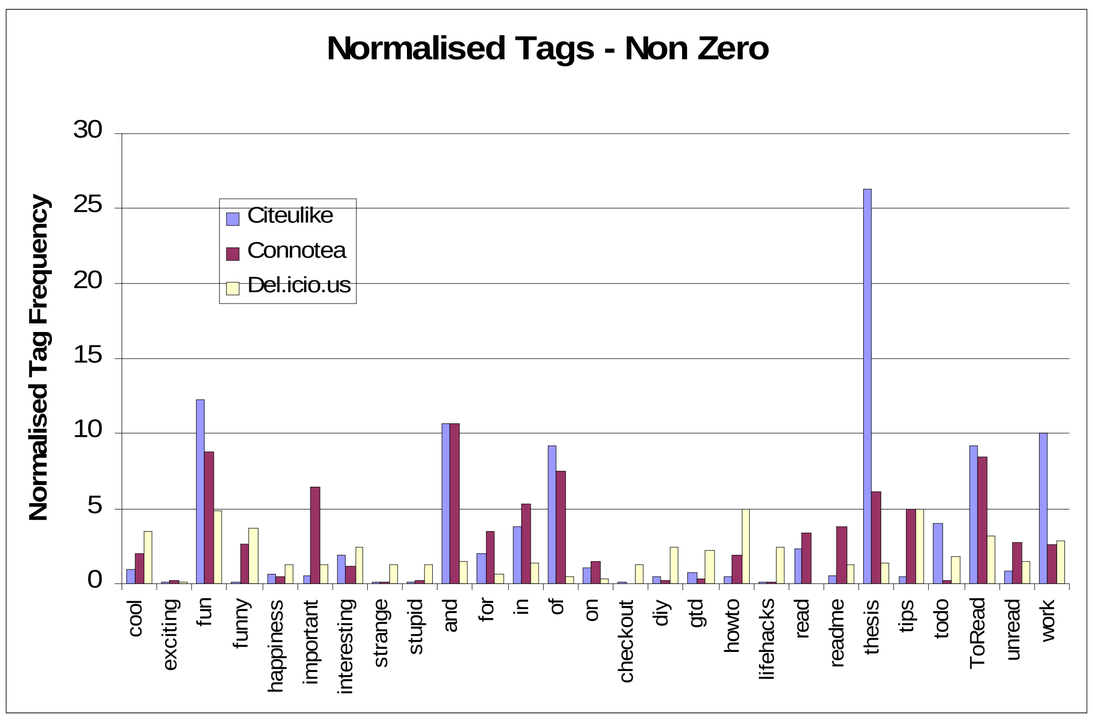
Top 10 Tags in each Tool (Affective or Time and Task)







Comparison of Normalised Tag Frequencies



Discussion

Many users of del.icio.us, citeulike and connotea appear to want to store more than just the subject of the documents they are bookmarking. Tags such as @toread, tobuy, todo, fun and cool suggest that users see their relationship to these documents in different ways. While the latter tags express an emotional connection to the document, the former show evidence of a desire to attach personal information management information to documents. This desire to combine personal information management and document classification echoes findings in document use research at Xerox in which users categorised items in order to better understand their relationship to other items and to tasks the users wished to perform. (Malone 1983; Sellen and Harper 2002)

At first glance, 'toread' seems to be a tag with very little value outside of a single person's personal organisational system, but Amazon's recommendation system has shown that collective information about buying patterns can be very useful for users who are interested in finding material that is like the material they are currently reading or watching. This suggests that the toread tag could function like a colleague's e-mail suggesting that the article is interesting and worthy of a little of your time. As a tag, it functions as an indicator of interest.

The prevalence of prepositions and conjunctions such as 'of', 'in' and 'and' in the tag lists was a surprise. An examination of the tag lists from which these tags occur suggests that sentences and phrases have been used as tags.

A large part of library science research is involved in the examination of how users seek and use information. Another important aspect of this is how they relate to information. Findings from this study suggest that users relate information to time related tasks, activities and their own emotional reactions.

References

Kipp, Margaret E.I. 2006. Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator, and Intermediary Keywords. *Canadian Journal of Information and Library Science*. (forthcoming) Preprint available from http://dlist.sir.arizona.edu/1533/

Kipp, Margaret E.I.; Campbell, D. Grant. 2006. Patterns and Inconsistencies in Collaborative Tagging Practices: An Examination of Tagging Practices. Proceedings of the Annual General Meeting of the American Society for Information Science and Technology, Austin, TX, November 3-8, 2006. (forthcoming)

Malone, Thomas W. 1983. How Do People Organize Their Desks? Implications for the Design of Office Information Systems. ACM Transactions on Office Information Systems 1(1): 99-112.

Sellen, Abigail J.; Harper, Richard H.R. 2002. The Myth of the Paperless Office. Cambridge, MA: MIT Press.