

# Tagging and Findability: Do Tags Help Users Find Things?

Margaret E. I. Kipp

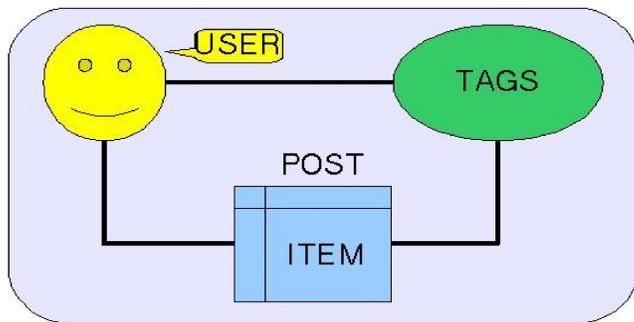
Palmer School of Library and Information Science

College of Information and Computer Science

Long Island University

## Introduction

Information retrieval research has been traditionally concerned with the efficiency with which information systems retrieve information that is relevant and useful, concerning itself with matters of precision, recall, and system effectiveness. This study aims to explore those questions in a new context. Proponents of tagging and social bookmarking often suggest that tags could provide at worst an adjunct to traditional classification systems and at best a complete replacement for such systems. (Shirky 2005) A reasonable method for testing the usefulness of a classification system for enabling retrieval is to perform an information retrieval study on a social bookmarking system to study the usefulness of tags in the support of information retrieval.



## Social Bookmarking Tools

Social bookmarking tools allow users to store their favourite bookmarks in a publicly accessible manner on the web. Users are encouraged to add descriptive terms or tags to each bookmark. Tagging is the process of assigning a label (whether classificatory or otherwise) to an item and is often combined with social bookmarking or the organisation of other information on the web, for example organising pictures on flickr.com.

## Research Questions

- Do tags appear to enhance findability? Do users feel that they have found what they are looking for?
- How do users find searching social bookmarking sites compared to searching more classically organised sites? Do users think that tags assigned by other users are more intuitive?
- Do tagging structures facilitate information retrieval? How does this compare to traditional structures of supporting information retrieval?

## Study Design

- IR (information retrieval) study focusing on relevance judgements of users (Cosijn and Ingwersen 2000; Tang and Sun 2003; Oppenheim et al. 2000)
- Collect a) judgement of users of tagging systems on the effectiveness of tags in finding relevant materials b) responses of users to using other people's tags
- Compare to users responses to using a controlled vocabulary thesaurus (Medical Subject Headings or MeSH)
- Searchers asked to search Pubmed and CiteULike
- CamStudio (screen capture software) used to record user input and mouse motions
- Think aloud protocol to capture user comments during and after search
- Preliminary study used library and information science students for their experience in searching various sites

## Timeline

Activity	Description	Length
Welcome	Initial greeting and welcome	2-3 minutes
Introduction to session	Introduction to the study discussing the session itself and the tasks they will be asked to perform.	5-7 minutes
First search task (citeulike or pubmed)	The first of two tasks consisting of: 1) the user's generation of keywords for search, 2) collection of articles, 3) analysis of retrieved articles for relevance, and 4) assignment of relevance judgements to the articles, 5) assignment of new set of keywords for search	15 minutes
Second search task (pubmed or citeulike)	same as first task	15 minutes
Post search discussion	A semi-structured interview involving a discussion of the participant's results and their own thoughts as to the usefulness of the terms they used to search and the terms used to describe the documents they retrieved.	15 minutes
Conclusion	Final comments and a thank you for participating.	3-5 minutes

## Study Process

- Participants will search for information using a traditional on-line database with assigned descriptors and a social bookmarking site
- Participants will be given a written description of a research need and asked to produce an initial set of keywords they would use to search for relevant articles

## Search Task

"You are a reference librarian in a science library. A patron approaches the reference desk and asks for information about the application of knowledge management or information organisation techniques in the realm of health information. The patron is looking for 5 articles discussing health information management and is especially interested in case studies, but will accept more theoretical articles as well."

## Data Collected

- Participants asked to select top 5 articles and assign relevance score to article based on an examination of available metadata
- Keywords collected: initial search terms, search terms used during search process, final set of search terms developed during search process

Keyword	Freq.
knowledge management	2
information organisation	2
case studies/"case stud"	2
health information	2
health information management	1
consumer health	1
knowledge management	1
health case studies	1
health theory	1

Initial Keywords

## Preliminary Results

- Users tended to prefer the search experience on the system used first
- All users used multi word keywords initially
- Most users separated final keywords by tool
- Users used between 3-5 keywords initially and suggested 4-5 keywords for CiteULike use and 1-4 for Pubmed
- Visual analysis of the articles showed little overlap.

## References

- Cosijn, Erica; Ingwersen, Peter. 2000. Dimensions of relevance. Information Processing and Management 36: 533-550.  
Morville, Peter. 2005. Ambient Findability. Sebastopol, CA: O'Reilly.  
Oppenheim, Charles; Morris, Anne; McKnight, Cliff. 2000. The Evaluation of WWW Search Engines. Journal of Documentation 56(2): 190-211.  
Shirky, Clay. 2005. Ontology is Overrated: Categories, Links, and Tags. [http://shirky.com/writings/ontology\\_oversrated.html](http://shirky.com/writings/ontology_oversrated.html)  
Tang, Muh-Chyun; Sun, Ying. 2003. Evaluation of Web-Based Search Engines Using User-Effort Measures. Libres 13(2): 1-11.