

# Tagging for Health Information Organisation and Retrieval

Margaret E.I. Kipp, [margaret.kipp@gmail.com](mailto:margaret.kipp@gmail.com)

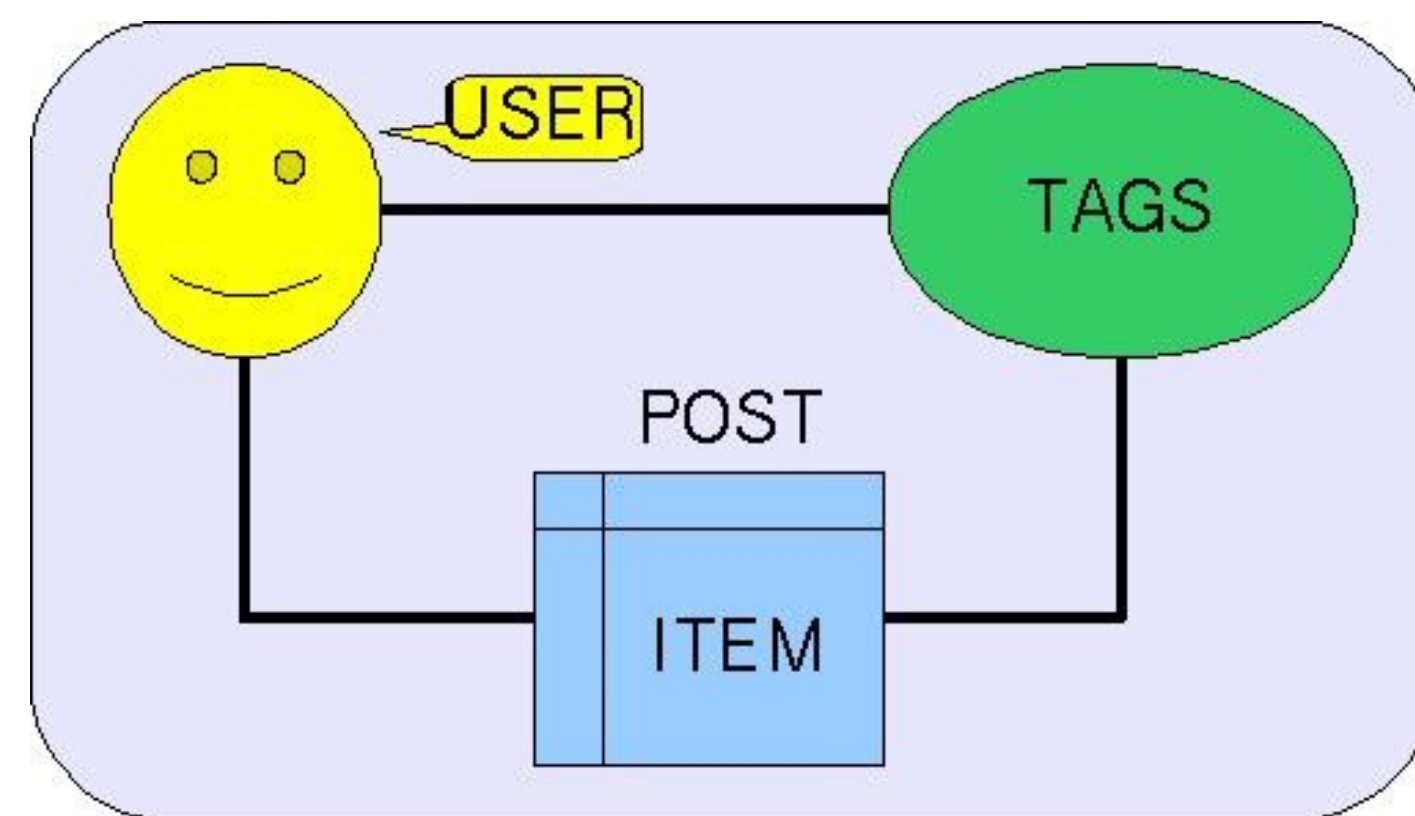
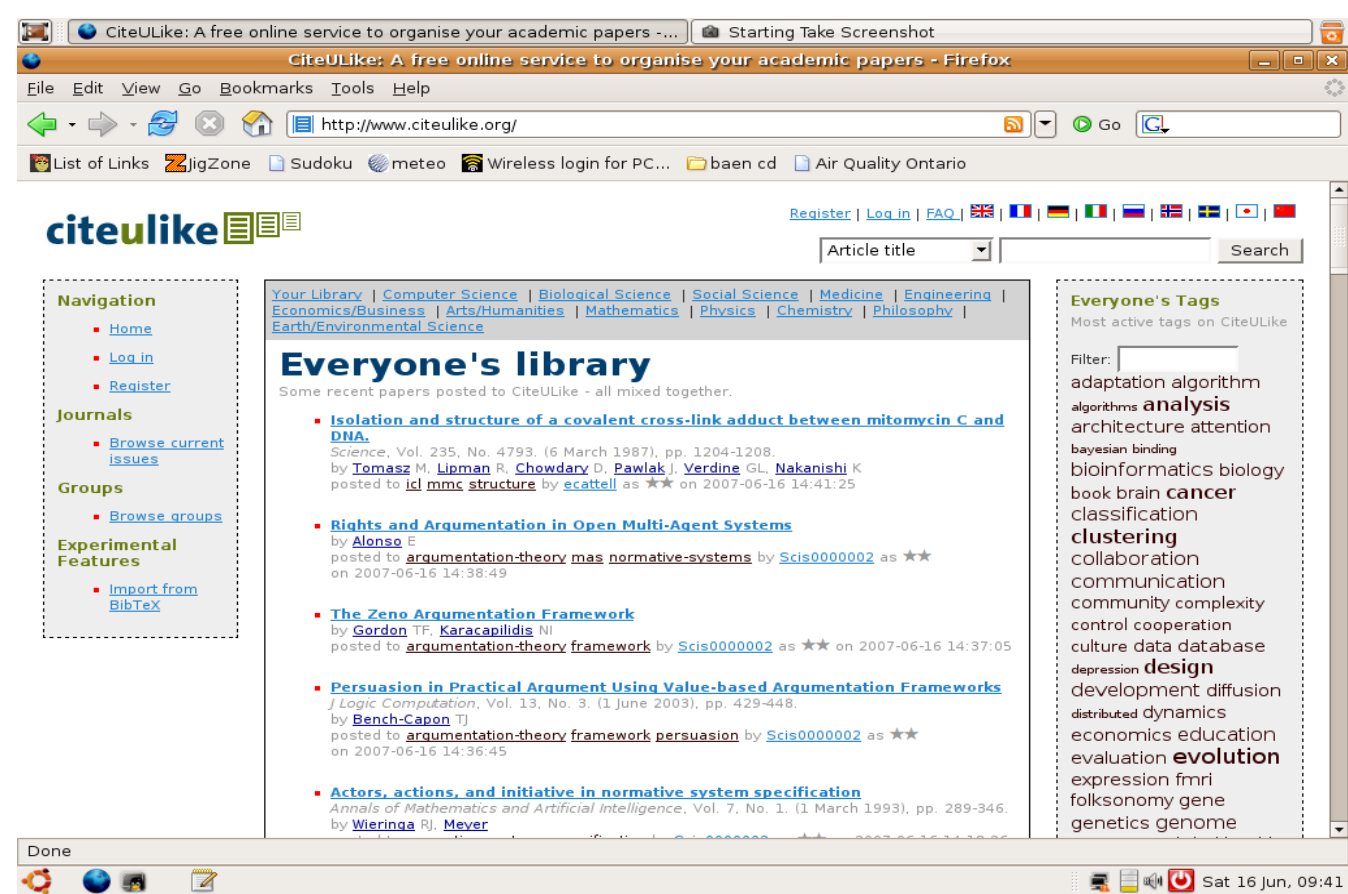
Faculty of Information and Media Studies, University of Western Ontario

## Introduction

This study examines the tagging practices evident on CiteULike, a researcher oriented social bookmarking site. Articles selected for this study were health information and biology related. Tagging practices were examined using standard informetric measures for analysis of bibliographic information and analysis of term use including comparison of tags and descriptors assigned to the same article.

## Social Bookmarking

- associating a link with a set of descriptive tags
- a post is a relationship between a user, an item and a set of tags



## Research Question

To what extent do term usage patterns of user tags and intermediary descriptors suggest a similar (or differing) context between users and intermediaries?

## Methodology

- medical or biology journals: JAMA, Proteins, J. Mol. Biol.
- 1280 articles collected from CiteULike
- associated descriptors (MeSH) from Pubmed

## Results

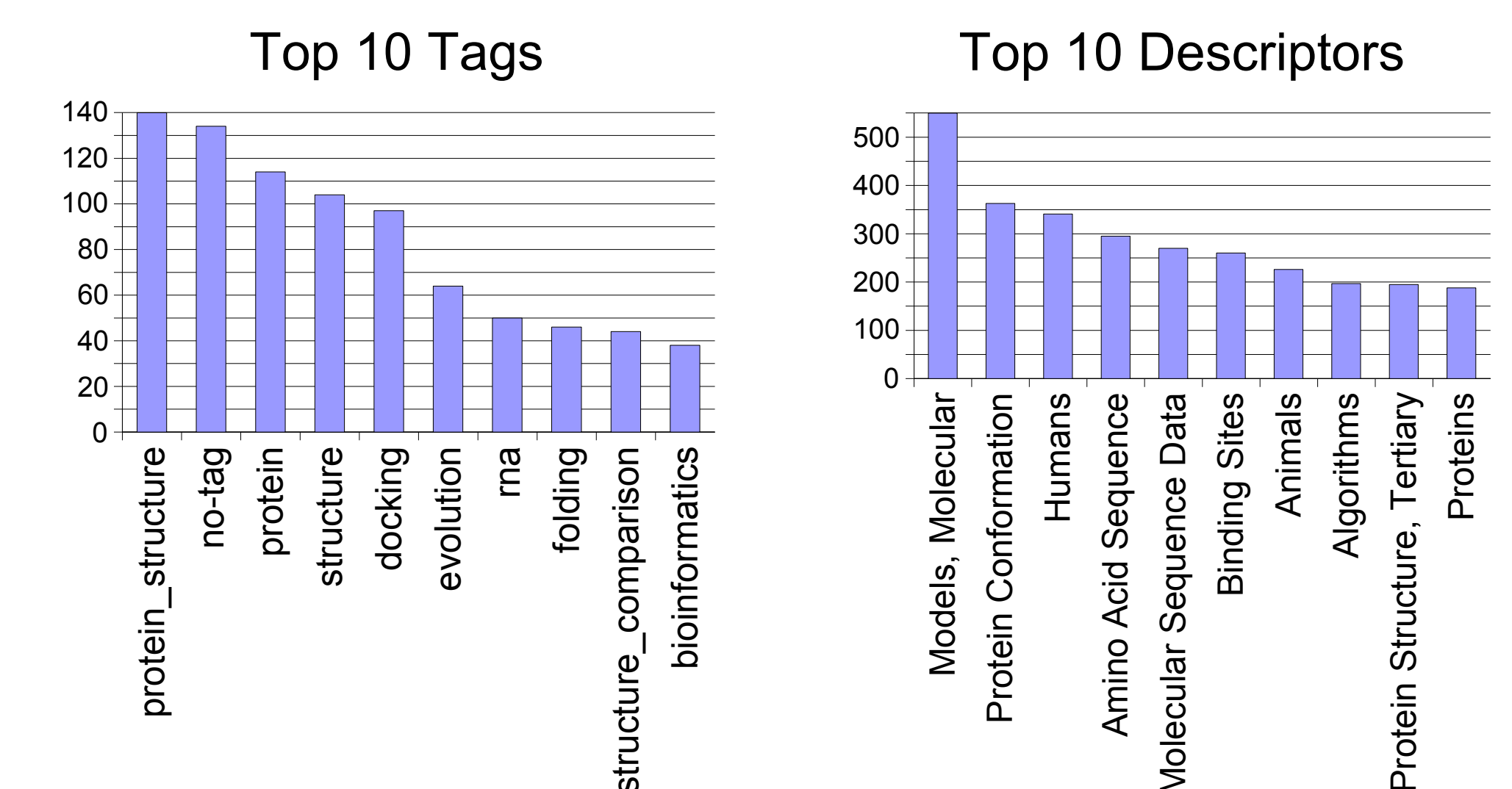
Users: 314 unique users  
articles tagged by up to 14 users (av 2)

Tags: 1449 unique (total 4289)  
average 2 per article (max 29, min 1)  
per user 1-3 per article (max 18, min 1)

Descriptors: 2746 unique (total 14507)  
average 10 per article (max 40, min 2)  
more heavily assigned to articles

## Popular Tags and Descriptors

- Proteins and J. Mol. Biol. terms were biological concepts
- JAMA terms were study or methodology related



## User Vocabulary Length

- number of unique terms used by a single user
- often but not always associated with heavy posters (more than 25 articles)

User	Total	Max/Article	Min/Article	Median/Article	Articles Posted
322	66	13	2	8	9
1143	62	8	1	3	94
1005	60	8	1	3	65

## Term Use

- many tags related to descriptors, but not formally in thesaurus
- thesaurus terms sometimes found split or joined in tags (e.g. diet and fats -> Dietary Fats)

Article: Optimal diets for prevention of coronary heart disease

Tags:

user1: chd, diet, fat, food, health, heartdisease, lipid, review

user2: coronary, diet, disease, heart

Descriptors:

Coronary Arteriosclerosis, Diet, Dietary Carbohydrates, Dietary Fats, Dietary Fiber, Folic Acid, Humans, Life Style, Lipoproteins

## Discussion and Conclusions

- metadata use matched previous study (Kipp 2006)
- previous study examined LIS journals, this study used medicine and biology
- differences were found in tag use between academic and professional journals

## Related Studies

Kipp, M.E.I. 2007. Tagging for Health Information Organisation and Retrieval. NASKO 2007. <http://dlist.sir.arizona.edu/1909/>

Kipp, M.E.I. 2006. Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator, and Intermediary Keywords. CJILS (in press). <http://dlist.sir.arizona.edu/1533/>