Implications of Social Tagging for Digital Libraries: Benefiting from User Collaboration in the Creation of Digital Knowledge

디지털 도서관을 위한 소셜 태깅의 의미: 이용자 협력을 활용한 디지털 지식 생성

Yunseon Choi*

ABSTRACT

This study aims to answer whether social tagging through user collaboration could be utilized for the creation of digital knowledge of the web, and whether we could verify the quality and efficacy of social tagging to obtain benefits from it. In particular, this paper examines the inter-indexer consistency of social tagging in comparison to professional indexing. It employs two different similarity measures, both of which are based on the Vector Space Model to deal with numerous indexers. It contributes to the utilization of social tagging in the organization of the web, and encourages to adopt social knowledge in developing suitable vocabularies for resources newly generated in the digital library environment. Furthermore, the comparative analysis with two different measures produced more credible results by illustrating a similar pattern of indexing tendency in both measures.

초 록

본 연구는 이용자 협력에 의한 소셜 태깅(social tagging)이 웹 자원을 위한 디지털 지식 생성에 활용될 수 있으며, 태깅의 양질성(quality)과 효율성이 실증적으로 증명될 수 있는가를 다루었다. 이 논고는 특별히 소셜 태깅의 색인 일관성(indexing consistency)을 평가하고 전문가들의 색인 일관성과 비교하여 분석하였다. 많은 수의 색인자들 간의 색인 일관성을 측정하기 위해 벡터 공간 모델(Vector Space Model)에 기반한 두 가지의 유사성 측정 공식을 사용하였다. 본 연구는 웹자원 관리에 있어서 소셜 태깅의 활용성 증진에 공헌하며, 디지털 도서관 환경에서 새롭게 생성되는 자료들에 대한 보다 적합한 어휘를 개발하는 데에 있어 소셜 지식을 적극적으로 수용할 필요가 있다고 주장한다. 또한 두 가지 공식에 의한 비교분석은 두 공식에서의 비슷한 색인 경향을 보여주면서 보다 신뢰적인 결과를 제공하였다.

Keywords: indexing, consistency, social tagging, folksonomies, controlled vocabularies, tags
색인, 일관성, 소셜 태깅, 포크소노미, 통제 어휘, 태그

---

* Ph.D. Candidate, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign(ychoi10@illinois.edu)

▪ Received : 27 May 2010 ▪ Revised : 4 June 2010 ▪ Accepted : 12 June 2010
▪ Journal of the Korean Society for Information Management, 27(2): 225-239, 2010.
〔DOI:10.3743/KOSIM.2010.27.2.225〕

1. Introduction

Libraries have a long history in organizing and providing access to resources. As networked information resources on the web continue to grow rapidly, today's digital library environments have led librarians and information professionals to index and manage digital resources on the web. Thus, this trend has required new tools for organizing and providing more effective access to the web. Subject gateways and web directories are such tools for internet resource discovery. Yet, studies have shown that such tools based on traditional organization schemes are not sufficient for the web (Nicholson et al. 2001; Mai 2004).

Social tagging has received significant attention since it helps organize digital contents by collaborative indexing based on user-generated tags by user collaboration. Several researchers have discussed social tagging behavior and its usefulness for classification or retrieval. However, further research is needed to systematically investigate social tagging and to justify its benefit.

This study aims to answer whether social tagging through user collaboration could be utilized for the creation of digital knowledge of web resources, and we could verify the quality and efficacy of social tagging to obtain benefit from it. In this paper, particularly examines the inter-indexer consistency of social tagging in comparison to professional indexing. Traditionally, consistency in indexing is considered as its quality (Cooper 1969, Rolling 1981). Leonard (1977) asserts that indexing consistency has a positive influence on retrieval effectiveness. Most indexing consistency studies have been conducted with a small number of professional indexers, and they tended to exclude users. Additionally, the studies mainly have focused on physical library collection, for example, physical books and periodicals not web resources. Accordingly, this research intends to bridge these gaps by (1) extending the scope of resources to various web documents indexed by users and (2) employing the Vector Space Model (VSM) - based indexing consistency method since it justifies its appropriate measure when dealing with a large number of indexers. Furthermore, in order to produce a more convincing and valuable analysis and decrease possible bias by a single measure, this study compares two different measures to measure indexing consistency.

2. Theoretical Background

2.1 Subject Gateways Using Controlled Vocabulary

Subject gateways can range from "loosely collated commercial directories" such as Yahoo! subject categories, to "collections of quality assessed web resources compiled by the academic or research community" (University of Kent 2009). In this study, I will refer to the concept of the latter for further discussion. The subject gateways emerged in response to the challenge of "resource discovery" in a rapidly developing Internet environment in the early and mid-1990s.

As one of subject gateways, the BUBL Information Service is “an Internet link collection for the library and higher education communities, operated by the Centre for Digital Library Research at the University of Strathclyde, and its name was originally short for Bulletin Board for Libraries” (Wikipedia). Since 1993 the BUBL Information Service has been a structured and user-friendly gateway for web resources in order to direct librarians, information professionals, academics and researchers (Gold 1996). Intute is funded by the Joint Information Systems Committee (JISC) which supports “education and research by promoting innovation in new technologies and by the central support of ICT services” in the UK higher and further education sectors (JISC Home).

Many subject gateways provide controlled vocabularies: either “home-made” or “standard library/information tools” such as classification schemes, subject headings and thesauri (Bawden and Robinson 2002). BUBL offers broad categorization of subjects based on the Dewey Decimal Classification scheme (http://bubl.ac.uk/). Intute (http://www.intute.ac.uk/) mainly uses the Universal Decimal Classification (UDC) and DDC for classification and has adapted them for in-house use. Intute also uses several thesauri for its subject relevance and comprehensiveness.

As described, subject gateways such as BUBL and Intute use controlled vocabularies, i.e., bibliographic classification schemes, subject heading lists and thesauri. However, controlled vocabularies have been challenged in their ability to index the range of digital web resources, e.g., slowness of revision, expensive indexing, and terms limited to topics found in physical and traditional library collections (Fidel 1991; Golub 2006; Macgregor and McCulloch 2006; Nowick and Mering 2003).

2.2 Social Tagging Using Uncontrolled Vocabulary

Social tagging data is one example of natural language terms, that is, uncontrolled vocabulary created by users. Social tagging is a promising way to complement the disadvantages of professional indexing because it is low-cost since a great number of users from everywhere contribute to the creation of tags. Social tagging is described as “user-generated keywords” (Trant 2009). Since tags indicate users' perspectives and descriptions on resources, they have been suggested as a means to improve search and retrieval of resources on the web. Flickr, Delicious and LibraryThing are popular social tagging sites.

The term “social tagging” is frequently associated with the term “folksonomy” which was coined by Thomas Vander Wal from ‘folk’ and ‘taxonomy’ (Smith 2004). While Trant (2009) provides good reviews of the overall trends of research on social tagging and folksonomy, she distinguishes the two terms “social tagging” and “folksonomy” by providing short definitions:

- Tagging: “a process with a focus on user choice of terminology”
- Folksonomy: “the resulting collective vocabulary (with a focus on knowledge organization)”
- Social tagging: “a sociotechnical context within which tagging takes place (with a focus on social

computing and networks)"

On the other hand, folksonomy has been criticized with its ambiguity of terms, a large number of synonyms, a lack of hierarchy, unstable term specificity, and variations of spelling etc. (Quintarelli 2005; Spiteri 2005). Merholz (2004) also describes drawbacks of tags as synonyms and inaccuracy, and emphasizes the contribution of the traditional classification and vocabulary control (Merholz 2004). Furthermore, Peterson (2006) criticizes folksonomy in that it has an intrinsic defect caused by its inability to produce the accuracy of formal classification.

2.3 Combination of Controlled and Uncontrolled Vocabulary

As discussed, both controlled vocabulary and uncontrolled vocabulary have their own advantages and disadvantages. Several researchers suggest the combination of both approaches since both may complement each other. Knapp et al. (1998)'s study illustrates that combining both approaches produced more effective retrieval performance rather than using only one approach. They conducted an experimental study to identify whether the free-text search terms could add supplementary relevant documents which are not retrieved by the controlled vocabulary. Their study allowed humanities scholars to search using both controlled vocabulary and free-text terms. Its results showed that when controlled vocabulary and free-text terms work together, more relevant records are retrieved. Weber's report (2006) on LibaryThing demonstrates that folksonomies and controlled vocabularies can harmoniously coexist, and the combination of both would obtain benefits, and there are useful correlations between the two. Figure 1 illustrates that LibraryThing supplies links to statistically related tags and subject headings.

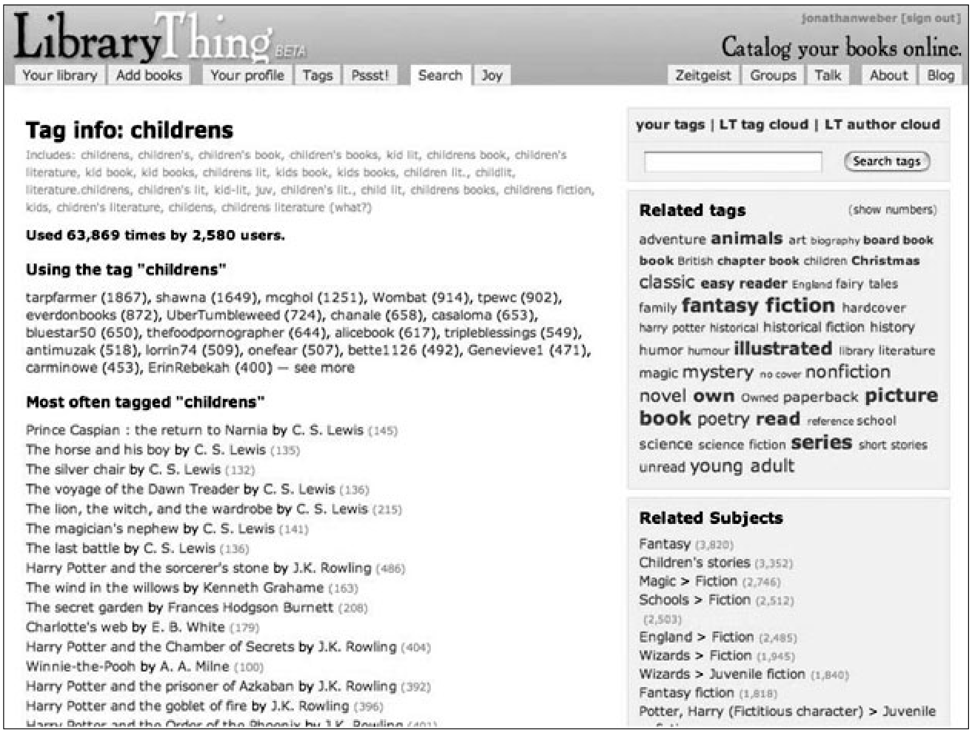
Many researchers have suggested that social tagging has potential for user-based indexing (Golder and Huberman 2006; Lin et al. 2006; Tennis 2006). It can be recognized that the participation of users in building controlled vocabulary is being realized in a social tagging environment where users created or generated search keywords based on their intuitive principles.

3. Methodology

3.1 Measures of Indexing Consistency

This study measures the consistency of tagging to verify its quality and efficacy in comparison to professional indexing. Zunde and Dexter (1969) define indexing consistency as "the degree of agreement in the representation of the essential information content of the document by certain sets of indexing terms selected individually and independently by each of the indexers in the group."

This research employs the Vector Space Model (VSM) - based indexing consistency method. The VSM was developed by Salton in 1975. In the model, documents and index terms are represented as vectors in the term space, and the documents are ranked



〈Figure 1〉 LibraryThing tagpage for tag “childrens”, showing (1) tag combinations, (2) related tags, (3) related subjects (Source: Weber 2006)

by closeness to terms. Figure 2 shows a typical three-dimensional index space where each item is identified by up to three distinct terms (Salton et al. 1975).

A typical three-dimensional space may be extended to n dimensions when n different index terms are present. A document matrix *V* for a document set consisting of *m* documents and *n* terms is shown in Figure 3.

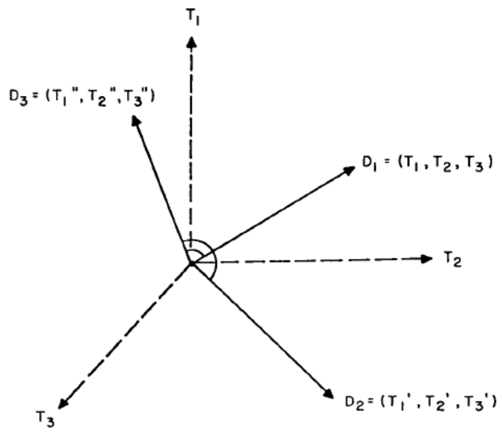
Wolfram and Olson (2007) applied the concept of document space in the VSM into the terms assigned by a group of indexers to a document, and defined an Indexer/Tagger Space. They calculated the distance between each indexer/tagger's resulting vector and the indexing centroid (or average vector across all indexers/taggers). In their measurement, high density space among indexers/taggers means more similarity and higher consistency (Figure 4).

Wolfram and Olson's Inter-Indexer (Tagger) Consistency Density (ICD) is calculated as follows:

$$ICD = \frac{\sum_{i=1}^{m} Sim\ (I_i, C)}{m}$$

where m is the number of indexers/taggers, C denotes Centroid and I_i is an indexer vector.

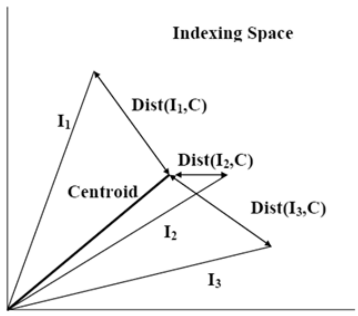
This research adapted Wolfram and Olson's formula with two different VSM based measures: Cosine



〈Figure 2〉 Vector representation of document space
(Source: Salton et al. 1975)

$$V = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix}$$

〈Figure 3〉 Document matrix V for a document
(Source: Olson and Wolfram 2007)



〈Figure 4〉 Indexer distances from the indexing centroid
(Source: Wolfram and Olson 2007)

similarity and Dot product similarity.

Cosine similarity is measured by the cosine of the angle between two vectors of same dimensions. The cosine similarity (Θ) is represented using a dot product and magnitude as:

Cos Θ =

$$Similarity\ (A, B) = \frac{A \cdot B}{|A||B|}$$

|A|: magnitude of vector A

|B|: magnitude of vector B

Θ: angle between vector A and vector B

The dot product of two vectors A (indexer 1) and B (indexer 2) is as follows:

Similarity (A, B) = A · B

(A · B is the dot product of vectors)

Thus, for measuring consistency among a large number of taggers, the adjusted Cosine similarity is:

Cos Θ =

$$Similarity\ (I_i, C) = \frac{I_i \cdot C}{|I_i||C|}$$

| I_i | : magnitude of vector I_i

|C| : magnitude of vector Centroid

Θ : angle between vector I_i and vector C

Also, the adjusted Dot product similarity is:

$$Similarity\ (I_i, C) = I_i \cdot C$$

Comparing two different similarity measures will produce a more convincing and valuable analysis and decrease possible bias by each measure when considering the characteristic of each measure. For example, in Dot product, similarity is proportional to the magnitude of indexer vector while in cosine similarity, magnitude is not considered.

This study measures two principal values of the indexing consistency (1) among social tagging users and (2) between two groups of professional indexers.

3.2 Data Collection

As a target social tagging site, this study chose Delicious. Delicious consists of a broad range of web resources, not limited to scholarly documents (e.g., journal articles on CiteUlike.org) or specific types of resources (e.g., photos and videos on flickr.com). Also, in order to examine professional indexing, two major subject gateways, BUBL and Intute were investigated.

This study collected 31,330 Delicious tags assigned to 118 web documents listed to BUBL and Intute. Web documents were randomly selected using the True Random Number Generator (www.random.org) based on 10 subject categories BUBL provides as top-level categories (Table 1). Indexing consistency was measured on tags from more than 50 up to 100 taggers who assigned most recently since Delicious feeds up to 100 most recent bookmarks.

To collect indexers' index terms from BUBL, for example, regarding a document, Amazon.com, the following paths are recognized and analyzed:

News media, journalism, publishing > Publishers and publishing > Booksellers and bookshops

The collection of indexers' index terms from Intute is the same as BUBL. Regarding a document, Amazon.com the following index strings of category paths is analyzed:

Communication and Media Studies > New Media > Interactive Games and Gaming Creative and Performing Arts > Music > MusicIndustry, Recording and Publishing

The summary of collected data is as follows:
- Web document samples (commonly indexed in three locations: Delicious, BUBL and Intute)
- Users' index terms tagged on the sampled documents (Delicious)
- Indexers' index terms (index term strings) on the sampled documents at BUBL and Intute

3.3 Data Analysis

3.3.1 Rules for vocabulary analysis

Based on discussion by Lancaster and Smith (1983), this study set five rules for exact match between two terms:
- Exact corresponding including singular/plural variations, e.g.) aurora to auroras
- Variant spellings, e.g.) organization to organisation
- Word forms (adjectival, noun, or verbal forms), e.g) medicine to medical
- Acronyms and full terms , e.g.) National Center for Biotechnology Information to NCBI
- Compound terms, e.g.) human/body to humanbody to human_body etc.

3.3.2 Term exclusion

This research developed a stoplist or a list of terms which can be excluded for processing. The stoplist included an explicit list of the terms that Sen et al. (2006) define as subjective and personal tags (Table 2), since those types of tags are not meaningful for indexing subjects of documents.

〈Table 1〉 BUBL subject categories

Top Categories	Subjects covered
000 Generalities	Computing, Internet, Libraries, Information Science
100 Philosophy and psychology	Ethics, Paranormal phenomena
200 Religion	Bibles, Religions of the world
300 Social sciences	Sociology, Politics, Economics, Law, Education
400 Language	Linguistics, Language learning, Specific languages
500 Science and mathematics	Physics, Chemistry, Earth Sciences, Biology, Zoology
600 Technology	Medicine, Engineering, Agriculture, Management
700 The arts	Art, Planning, Architecture, Music, Sport
800 Literature and rhetoric	Literature of Specific languages
900 Geography and history	Travel, Genealogy, Archaeology

〈Table 2〉 Three types of tags (Source: Sen et al. 2006)

Types of tags	Definitions	Examples of identified tags
Factual tags	"identifies facts about" a resource e.g., people, places, or concepts	government, socialsecurity, finance etc.
Subjective tags	"express user opinions" related to a resource	good, worth, recommend, toRead, informative etc.
Personal tags	having "intended audience of tag applied themselves"	myDaughter, forSon, etc.

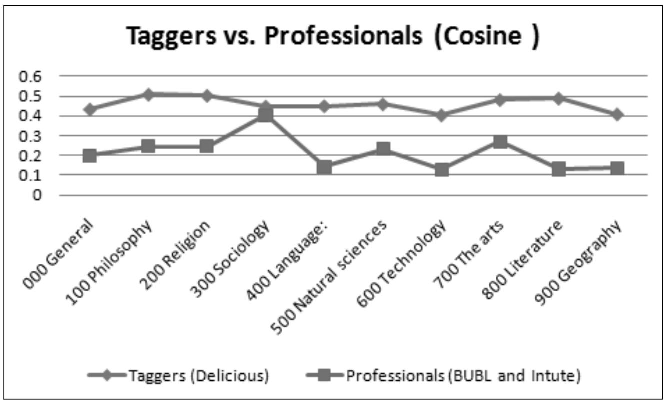
4. Findings

In both measures (cosine and dot product), it was revealed that there was consistency over all subjects among taggers in Delicious. In contrast, indexing similarity between two groups of professionals (BUBL vs. Intute) illustrated inconsistency over all subjects in both measures (Figure 5 and 6). The comparison of cosine and dot product similarity measures demonstrated an analogous pattern of consistency concerning subjects (Figure 7 and 8).

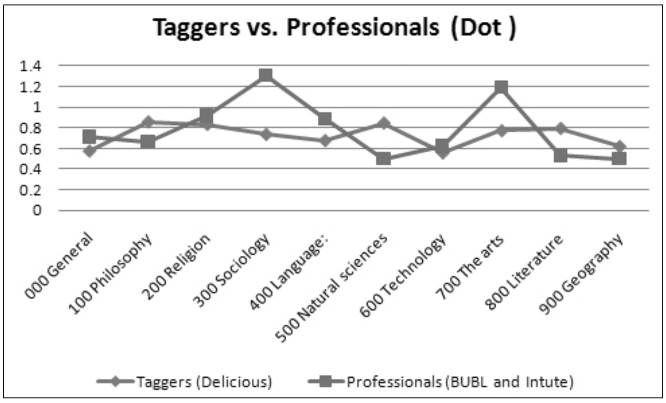
5. Discussion

5.1 Comparison of Taggers and Professionals

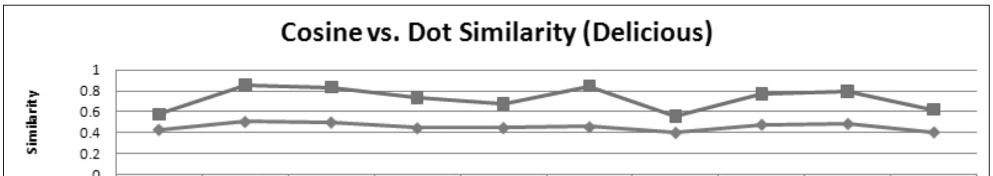
As illustrated in Figure 5 and Figure 6, the similarity on Sociology subject in two professional groups reached the highest value in both measures. It was uncovered that both BUBL and Intute located most documents in that subject into "Social science" or "Sociology" categories (Table 3). It explained that most documents on that subject were able to be simply located in the existing categories.



〈Figure 5〉 Indexing consistency using cosine similarity between taggers (Delicious) and professionals (BUBL & Intute)

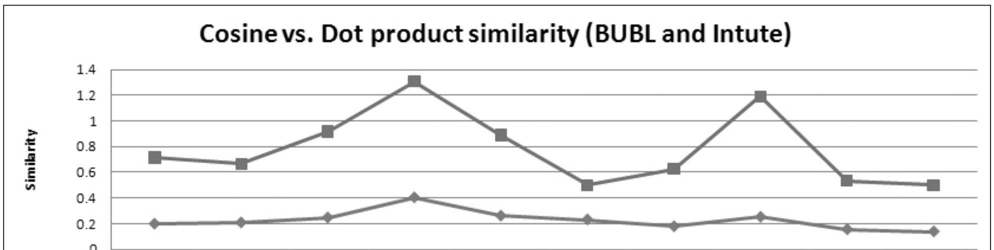


〈Figure 6〉 Indexing consistency using dot product similarity between taggers (Delicious) and professionals (BUBL & Intute)



	000 General	100 Philosophy	200 Religion	300 Sociology	400 Language:	500 Natural sciences	600 Technology	700 The arts	800 Literature	900 Geography
Cosine	0.43279526	0.50918215	0.50381515	0.44803634	0.45013366	0.45929366	0.40282815	0.48173387	0.4894299	0.40806029
Dot	0.5768708	0.85559773	0.83220772	0.73589871	0.67759503	0.84467798	0.55873665	0.77134585	0.79363593	0.62100974

〈Figure 7〉 Comparison of cosine similarity and dot product similarity of indexing among Delicious taggers



	000 General	100 Philosophy	200 Religion	300 Sociology	400 Language:	500 Natural sciences	600 Technology	700 The arts	800 Literature	900 Geography
Cosine	0.197363503	0.206930271	0.245883152	0.403838903	0.259951368	0.231200968	0.177258481	0.251171462	0.153500105	0.136637253
Dot	0.714285714	0.666666667	0.916666667	1.307692308	0.888888889	0.5	0.625	1.19047619	0.533333333	0.5

〈Figure 8〉 Comparison of cosine similarity and dot product similarity of indexing between BUBL and Intute professionals

〈Table 3〉 Indexing on Sociology between BUBL and Intute

Sociology subject	Title	BUBL	Intute
301 Sociology: general resources	Sociological Tour Through Cyberspace, www.trinity.edu/~mkearl/index.html	Social sciences, Sociology	Social sciences, Sociology
310 International statistics	IDB Population Pyramids, International Data Base (IDB) - Pyramids , http://www.census.gov/ipc/www/idb/pyramids.html	Social sciences, Statistics	Social sciences, Statistics, data, Population
330 Economics: general resources	History of Economic Thought, http://cepa.newschool.edu/het/	Social sciences, Economics	Social sciences, Economics, Sociology
355 Military science: general resources	DOD Dictionary of Military Terms, http://www.dtic.mil/doctrine/jel/doddict/	Social sciences, Military science	Social sciences, Government policy, Military science

In both measures, regarding several documents on Natural sciences, the similarity between BUBL and Intute was relatively very low (Figure 5 and Figure 6). It explains that BUBL and Intute have rather different points of view on the same documents in that subjects (Table 4). Table 4 also illustrates that Delicious taggers indexed those documents with their preferred terms (e.g., “biology” rather than “natural sciences”) and up-to-date terms (e.g., bio-informatics).

5.2 Comparison of Cosine and Dot Product Measure

Figure 7 illustrated that the similarity of cosine measure among taggers in Delicious tended to be parallel to that of dot product measure concerning over all subjects. Regarding similarity between BUBL and Intute professionals, two line graphs in Figure 8 are alike on the whole except on Natural sciences showing different positions in both graphs.

As described in Section 3.1 Measures of indexing consistency, Dot product based similarity is represented by

Similarity (A, B) = A · B
(A · B is the dot product of vectors)

and, the cosine similarity (Θ) is represented using a dot product and magnitude as:

cosΘ =

$$Similarity\ (A,B) = \frac{A \cdot B}{|A||B|}$$

|A|: magnitude of vector A
|B|: magnitude of vector B
Θ: angle between vector A and vector B

As shown above, in the cosine similarity measure, values are in inverse proportion to the magnitude

〈Table 4〉 Indexing on Natural sciences (BUBL vs. Intute vs. Delicious)

Natural Sciences	Title	BUBL	Intute	Delicious top ranked tags
500 Natural sciences: national centres	National Science Foundation, http://www.nsf.gov/	Natural sciences,	Engineering, Physical sciences	science, research, education, government, nsf, funding, reference, technology, news, grants, academic, foundation, usa, biology, national, information, resource,
570 Life sciences, biology	BBSRC: Biotechnology and Biological_sciences Research Council: http://www.bbsrc.ac.uk/	Natural sciences, Life sciences, Biology	Biological sciences	research, science, biotechnology, funding, biology, uk, education, work, bioinformatics, bioscience, development, bbsrc, research, councils, research_councils, postgraduate, news, academic biotech, biological, researchcouncil
580 Plants, general resources	Botanical Society of America Online Image Collection: http://images.botany.org/	Natural sciences, Plants	Biological sciences, Botany, Images	images, botany, plants, biology, science, research, free, photos, pictures, media, collection, horticulture, gardening, multimedia, flowers, botanica, biología, biologyguide
590 Animals, general resources	Animal diversity web: http://animaldiversity.ummz.umich.edu/site/	Natural sciences, Mathematics, Animals	Biological sciences, Zoology	animals, science, biology, reference, zoology, taxonomy, biodiversity, nature, classification, education, research, resources, diversity, database, species, wildlife, ecology, encyclopedia, environment, teaching

of vector A and vector B. The magnitude of vector is proportional to the number of terms. In the dot product measures, the value of similarity is represented by the number of common terms between two indexers (see Table 5), and the magnitude of vector is not considered.

Table 5 illustrated that there was a slight difference between cosine and dot product measure depending on the number of index terms. That is, when there was only one common term between BUBL and Intute terms, the value of dot product measure generated “1” while the value of cosine similarity turned out variant figures, e.g., 0.5, 0.707106781186547 or 0.288675134594812. This unique characteristic of each measure caused such a different tendency on similarity depicted in Figure 8.

〈Table 5〉 Cosine vs. Dot product measure of indexing on Natural Sciences between BUBL and Intute

Natural Science	Title	BUBL	Intute	Cosine	Dot
500 Natural sciences: national centres	National Science Foundation, http://www.nsf.gov/	Natural sciences	Engineering, Physical sciences	0	0
510 Mathematics, general resources	MathSciNet: http://www.ams.org/mathscinet/	Natural sciences, Mathematics	Mathematics, Computer science	0.5	1
520 Astronomy, general resources	Astronomy Picture of the Day, http://antwrp.gsfc.nasa.gov/apod/astropix.html	Natural sciences, Astronomy	Astronomy	0.707106781186547	1
550 Earth sciences	GeoGuide, http://www.geo-guide.de/	Natural sciences, Mathematics, Earth sciences	Geography, environment, Physical sciences, Earth sciences	0.288675134594812	1

6. Concluding Remarks

This study contributes to the utilization of social tagging through user collaboration in the organization of the web, and encourages to adopt social knowledge in developing suitable vocabularies for resources newly generated in the digital library environment. Consistency of tagging over all subjects verified the quality and efficacy of user indexing. Furthermore, the comparative analysis with two different measures produced more credible results by illustrating a similar pattern of indexing tendency in both measures.

This study limited the scope of sample web documents to the common document collection of BUBL and Intute, and only if a web document is listed at both locations above were tags assigned to the web document at Delicious collected and analyzed. Thus, any conclusions about properties of tags in Delicious would be limited to web documents selected for inclusion in subject gateways and indexed by professional indexers.

References

Cooper, W. S. 1969. “Is interindexer consistency a hobgoblin?” *American Documentation,* 20(3): 268-278.

Fidel, R. 1991. “Searchers’ selection of search keys: II. Controlled vocabulary or free-text searching.” *Journal of the American Society for Information Science,* 42(7): 501-514.

Gold, J. 1996. “Introducing a new service from BUBL” [Libraries of Networked Knowledge]. *The Serials Librarian,* 30(2): 21-26.

Golder, S. and B. A. Huberman. 2006. “The structure of collaborative tagging systems.” *Journal of Information Science,* 32(2): 198-208.

Golub, Koraljka. 2006. “Using controlled vocabularies in automated subject classification of textual web pages, in the context of browsing.” *IEEETCDL Bulletin,* 2(2): 1-11.

Joint Information Systems Committee. (JISC). <http://www.jisc.ac.uk/>.

Knapp, Sara D., Laura B. Cohen, and D. R. Juedes. 1998. “A natural language thesaurus for the Humanities: The need for a database search aid.” *The Library Quarterly,* 68(4): 406-430.

Leonard, L. E. 1977. *Inter-indexer consistency studies, 1954-1975: a review of the literature and summary of results.* University of Illinois, Graduate School of Library Science Cham- paign, IL. Occasional Papers. No.131, De- cember 1977. 51.

Lin, X., J. E. Beaudoin, Y. Bui, and K. Desai. 2006. “Exploring characteristics of social classification.” *17th ASIS&T SIG/CR Classification Research Workshop.*

Macgregor, G. and E. McCulloch. 2006. “Collaborative tagging as a knowledge organization and resource discovery tool." *Library Review,* 55(5): 291-300.

Mai, J.-E. 2004. "Classification of the Web: challenges and inquiries." *Knowledge Organization,* 31(2): 92-97.

Merholz, P. 2004. *Metadata for the Masses, adaptive path.* [cited 2010.6.4]. <http://www.adaptivepath.com/publications/essays/archives/000361.php>.

Nicholson, D. et al. 2001. *HILT: High Level Thesaurus Project: Final Report.* [cited 2010.6.4]. <http://hilt.cdlr.strath.ac.uk/Reports/Documents/HILTfinalreport.doc>.

Nowick, E. A. and M. Mering. 2003. "Comparisons between Internet users' free-text queries and controlled vocabularies: a case study in water quality." *Technical Services Quarterly,* 21(2): 15-32.

Peterson, E. 2006. "Beneath the Metadata: Some philosophical problems with folksonomy." *D-Lib Magazine,* 12(11).

Quintarelli, E. 2005. "Folksonomies: power to the people." *Proceedings of the 1st International Society for Knowledge Organization* (Italy) (ISKOI), [cited 2010.6.4]. <http://www.iskoi.org/doc/folksonomies.htm>.

Rolling, L. 1981. "Indexing consistency, quality and efficiency." *Information Processing & Management,* 17: 69-76.

Salton, G., A. Wong, C. S. Yang. 1975. "A vector space model for automatic indexing." *Communications of the ACM,* 18(11): 613-620.

Sen, S. et al. 2006. "Tagging, communites, vocabulary, evolution." *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work.*

Smith, G. 2004. *Folksonomy: social classification.* Atomiq/information Architecture [blog]. [cited 2010.6.4]. <http://atomiq.org/archives/2004/08/folksonomy_social_classification.html>.

Spiteri, L. 2005. *Controlled Vocabularies and Folksonomies.* Presentation at *Canadian Metadata Forum,* Ottawa, ON, September 27, 2005. 23.

Tennis, Joseph T. 2006. "Social Tagging and the Next Steps for Indexing." In Furner, Jonathan and Tennis, Joseph T, Eds. *Proceedings of the 17th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Re- search,* Austin, Texas.

Trant, J. 2009. "Studying Social Tagging and Folksonomy: A Review and Framework." *Journal of Digital Information,* 10(1).

University of Kent. 2009. *Library Services Subject Guides.* [cited 2010.6.4]. <http://www.kent.ac.uk/library/subjects/healthinfo/subjgate.html>.

Weber, J. 2006. *Folksonomy and controlled vocabulary in Library Thing.* Unpublished Final Project, University of Pittsburgh.

Wikipedia: The free encyclopedia. 2009. FL: Wikimedia Foundation, Inc. [cited 2010.6.4]. <http://www.wikipedia.org>.

Wolfram, D., and H. A. Olson. 2007. "A Method for Comparing Large Scale Inter-indexer Consistency Using IR Modeling." *Proceedings of the 35th Annual Conference of the Canadian Association for Information Science.*

Zunde, P., and M. E. Dexter. 1969. "Indexing consistency and quality." *American Documentation,* 20(3): 259-267.