Annals of Library and Information Studies
Vol. 63, December 2016, pp. 289-297

The role of social tags in web resource discovery: an evaluation of user-generated keywords

Praveenkumar Vaidya[a] and N. S. Harinarayana[b]

[a]Librarian, Tolani Maritime Institute, Pune and Research Student, Department of Studies in Library and Information Science, University of Mysore, Mysuru, India, Email: vaidyapraveen@gmail.com

[b]Associate Professor, Department of Studies in Library and Information Science, University of Mysore, Mysuru, India, Email: ns.harinarayana@gmail.com

Received: 08 August 2016; revised and accepted: 22 December 2016

Social tags are user generated metadata and play vital role in Information Retrieval (IR) of web resources. This study is an attempt to determine the similarities between social tags extracted from LibraryThing and Library of Congress Subject Headings (LCSH) for the titles chosen for study by adopting Cosine similarity method. The result shows that social tags and controlled vocabularies are not quite similar due to the free nature of social tags mostly assigned by users whereas controlled vocabularies are attributed by subject experts. In the context of information retrieval and text mining, the Cosine similarity is most commonly adopted method to evaluate the similarity of vectors as it provides an important measurement in terms of degree to know how similar two documents are likely to be in relation to their subject matter. The LibraryThing tags and LCSH are represented in vectors to measure Cosine similarity between them.

Keywords: Social Tagging; Folksonomies; Taxonomies; Information Retrieval; Cosine Similarity

Introduction

The second generation Internet has really changed the provision of web services among users to participate and share the resources. The flexibility with which users can contribute knowledge and share as well, has importantly added many clusters of attractive features to web services. This change is determined predominantly by users' evolving needs and users' needs are sometimes induced by the presence of novel technologies, which make prospective, what were previously not feasible or rather impractical. Web 2.0 represents an example of such a technology[1]**.** Social tagging is also one of the features of Web 2.0 in new age web access and contents categorization. The mid-2000s have seen swift progress in levels of interest in these kinds of techniques for generating descriptions of resources for the purposes of discovery, access, and retrieval[2]. The attribution of social tags to resources in context of exceptional growth of knowledge objects has necessitated the users to new approach of resource discovery. The library professionals are familiar with different knowledge organisation tools like classification schemes and taxonomies, which provide professional perspectives in resource discovery. Folksonomies or social tags are user generated metadata for web resources mostly to describe subject contents of such objects and thereby can be used astutely for contents categorization and subsequent retrieval.

Folksonomies

A folksonomy begins with tagging. A folksonomy is a decentralized, social approach to creating metadata for digital resources[3]. It is spontaneous and Internet based information retrieval methodology consisting of collaboratively generated, open-ended labels or tags that categorise contents such as web resources, online photographs, and web links[4]. Collaborative tagging describes the process by which many users add metadata in the form of keywords to share contents. The collaborative tagging has grown

in popularity on the web, on sites that allow users to tag book marks, photographs and other contents[5]. Basically, it is a free-form tagging and user generated classification system of web contents that allows users to tag their favorite web resources with their chosen descriptors or phrases selected from natural language[6]. Hence, folksonomies are generally useful to organize information resources and support efficient retrieval of such resources. Proponents also suggest that social tagging will offer subject based indexing in areas, where indexing process is exorbitantly expensive due to collection size or completely lacking such as in many web based resources[7].

The process of assigning Folksonomies is also known as user tagging, collaborative tagging, social indexing, social bookmarking, and collaborative indexing[8-9]. Folksonomies are also known as collaborative tagging by which many users add metadata in the form of keywords to shared contents. Similarly, Macgregor and McCulloch describe 'collaborative tagging' as "a practice whereby users assign uncontrolled keywords to information resources"[10]. The tagging is done by the "users" whose involvement in the resource discovery process has generally been limited to the expression of information needs and building of search requests and recording of resource metadata.

Hence, folksonomies are characterized as user oriented, empowering, democratic, low-cost, dynamic and instructive. Therefore, such user warrant based indexing processes are considered as alternative route to supplement and complement the roles of the information professionals in subject indexing and to facilitate information retrieval and knowledge organisation over the web.

Objective of the study and Research questions

Library and information systems, all over the world, are making a quantum jump from OPAC based information retrieval systems to library discovery systems, where all kinds of library resources (locally processed and globally subscribed resources) can be retrieved seamlessly from a single-window search interface. The user interfaces of the most of such discovery systems are Web 2.0-enabled and thereby support collaboration, participation and user interaction. Social tagging or folksonomy is an essential component of library discovery systems with facility to index and search tags generated/donated by users. In this backdrop, this paper aims to discover similarities of controlled vocabulary system with user generated metadata or social tags. In the context of social tag based web retrieval, the user-generated tags play crucial role in matching user query with terms originated from literary warrant (from controlled vocabularies) as well as from user warrant (from social tags). But till date there is no obvious answer to the generic question that how and to what extent social tagging is influencing information retrieval in library discovery systems. This study attempts to answer the following specific research questions leading towards the generic issue as mentioned above:

- RQ 1. What is the relationship between social tags and controlled vocabularies?
- RQ 2. Whether these social tags and controlled vocabularies are complementary to each other?

These two specific research questions also aim to understand how social tags can enrich and update control vocabulary subject terms.

Folksonomy, Taxonomy and Information Retrieval

Folksonomies, as an uncontrolled vocabulary device lack the preciseness in information retrieval as in case of taxonomies. Taxonomies or controlled vocabularies are professionally assisted, which has strict rules and consensus for the purpose of information retrieval. Folksonomy has an advantage of inclusiveness of vocabularies of community users, and thereby ensures currency of descriptors and provides an insight to the information seeking behavior of users. Folksonomy, as a mechanism to support user warrant, is a low-cost device in implementation and in their reuse[11-12]. But, folksonomies, at the same time, are limited by factors like – no control over synonyms, lack of precision, lack of hierarchy, and lack of recall values in comparison with subject taxonomies. These are also seriously vulnerable to manipulation in an effort to make the tags more popular. Vocabulary control devices provide a systematic set of metadata for precise information retrieval, but folksonomies support user warrant and make resources more browsable and searchable. User tagging allows users to easily seek the information they need using common terms, and without having to worry about the

intricacy of the underlying mechanism of the cataloguing and indexing system[13].

Despite all its limitations there is a consensus among researchers that folksonomies can supplement, even may improve the information organisation. Furthermore, tagging is not about accuracy, authority, and not about right descriptors or wrong descriptors, but about recalling, user warrant and user acceptance based on users' needs. Hence, librarians must think of using both social tags and traditional information organisation systems like controlled vocabularies and use it simultaneously to complement and supplement information retrieval.

Information Retrieval

Information retrieval research has been conventionally concerned with the efficiency with which information systems retrieve information resources that is relevant and useful, concerning itself with matters of precision, recall and system effectiveness. Such studies contain an implicit evaluation of the categorization of the material[14]. With much emphasis on precision and recall, the information retrieval or knowledge discovery in case of social tags has attracted many researchers. The collaborative systems like delicious and CiteULike allow users to participate in the classification of journal articles by encouraging them to assign tags. The tags assigned by users in www.delicious.com and www.citeulike.org are organised and shared by all registered users. These tags play vital role in knowledge discovery assigned by other users to same information resource. With these tags linked to each user will develop a network who may play a vital role in resource discovery. This curiosity raises the questions: whether the traditional indexing system and tagging are related to each other in web information retrieval? There has been little research in this context of social tagging to provide some insight into the issue that how and to what extent tagging system can be adopted in information retrieval to enhance search process.

Review of related literature

Many researchers have examined the different aspects of the social tagging that fall into resource and information discovery. Morrison announced prime utility of folksonomy is to support successful information retrieval[15]. Information tagged by others is only suitable to the users if they understand the contents, if that practical information is retrieved that would be useful to fellow users.

As discussed above, speed, precision and recall are characters of information retrieval. It is important that websites those employ folksonomies should be able to have these characters to prove them to be useful. In order to understand the effectiveness of folksonomies at information retrieval, Hotho, in his path breaking study in collaborative tagging and retrieval, recommended that enhanced search facilities are necessary for emergent semantics within folksonomy based systems and he presented a formal model for folksonomies, the *FolkRank* ranking algorithm that takes into the account the structure of folksonomies and evaluation results on a large scale data set[11].

Morrison conducted a shootout-style study between three different kinds of web information retrieval systems; search engines, directories and folksonomies. Comparative charts were prepared to measure information retrieval effectiveness for precision and recall and also for information needs, categories, overlap and relevance and query characteristics. It is found from the study that folksonomies results were overlapped with the results from search engines and they did poorly with searches for an exact site[15].

In another study by Kipp and Campbell worked on to understand whether tagging could be sufficiently useful as index terms to be worth adding to records. The dataset used were from CiteULike and PubMed health database and it was observed that tagging does not completely replace controlled vocabularies, but provides an added dimension to subject access from the perspective of end-users and provides early access to emerging terminologies[16].

Lu and Kipp investigated the retrieval effectiveness of collaborative tags by experimental tests. The results indicated that tags improved overall retrieval performance and tags are potentially promising for retrieval[17].

Thomas et.al., has done a comparative study between user tags and controlled vocabularies with different datasets. The social tags are drawn from LibraryThing website and compared them with expert-assigned subject terms according to Library of Congress Subject Headings (LCSH) and purpose of the study is to examine the difference and connections

between these tag systems and also to explore the feasibility and obstacles of implementing social tagging in library systems[18]. Particularly, in Lu et. al., the comparison was done by using Jaccard similarity method with the social tags and subject terms present in the whole dataset at book level, social tags are compared to subject terms applied to the same book. The researcher checked the frequency or popularity of the overlapping terms in tags and LCSHs and they were represented in statistical model with formula and respective charts. The authors conclude that social taggers may help to enhance the subject access to library collections by describing library resources with terms different than those used by experts. The results also indicate that these benefits are best achieved with large number of tags[19].

In another study, Kipp and Campbell have examined how tags can enhance the experience of resource discovery. The design of this study is based on common information retrieval with an emphasis on the collection of keywords used in the search in addition to the collection of set of keywords judged relevant by the participant. It was observed and also concluded that tagging does not completely replace controlled vocabularies, but offers an added element to subject access from the viewpoint of end-users and provides early access to developing terminologies[16].

In another work, Voorbji conducted a study to determine the value of LibraryThing tags, where the random sample of 600 records were evenly distributed among humanities, social sciences and natural sciences which were taken from the library catalogue, unlike titles from LCSH. This study focuses the importance of professional subject indexing and replacing them by user generated tag assignment would be detrimental for the recall. With the uncontrolled nature of folksonomies, tags are inherently imprecise, inexact and overly personalized and the result is chaotic and negatively affect the retrieval, since user's search term would not match the controlled vocabulary[20].

Lee and Schleyer in their similar work compared MeSH terms with CiteULike social tags by determining Jaccard coefficient. The study examines the degree of difference between two categories of metadata for biomedical articles generated by professionally trained indexers and assigned social tags by readers. It was revealed that MeSH terms and tags show different understandings of two groups, the indexers and the readers[21].

Syn and Spring explored the way to obtain a set of tags representing the resource from the tags provided by users. The research selects important tags and removes meaningless ones. The results suggest that processing of users tags successfully identifies the terms that represent the topic categories and web resource content[22]. Lu and Kipp, investigate the retrieval effectiveness of social tags and author keywords in different environments through controlled experiments. The findings suggest that including tags and author keywords in indexes can enhance the recall but may improve or worsen average precision. The findings also provide useful implications for designing retrieval systems that incorporate tags and author keywords. The experimental design of this study follows Cranefield paradigms[23]. To conduct retrieval test, a test collection, a list of topics and relevant judgments are needed. In another interesting study by Choi and Syn, examines user tags that describe digitized archival collections in the field of humanities collection of Nineteenth-Century Electronic Scholarship (NINES). The study demonstrated that there is valuable potential for tags to locate related resources and to identify potential indexing terms for controlled vocabularies[24].

Yi, K, in this research work used tf-idf and Cosine based similarity with other similarity techniques including Jaccard similarity method. The analysis demonstrates to predict the semantic similarity of social tags and controlled vocabularies[25].

However, the strength of folksonomies is collaborative indexing; its weakness lies in information retrieval which lacks precision. To enhance the precision in retrieval is the resultant challenge for information architects and library scientists. To deploy various scientific methodologies and measuring their efficiency would be appropriate to understand the effectiveness of information retrieval.

In summary, these previous studies signify that an analysis of tags can offer insight into users' interpretation of the contents of resources that will be significant and beneficial for other users. This research work also complements the previous works and analysis is attempted by contrasting the LibraryThing tags with Library of Congress Subject Headings.

Research methodology and data collection

In this research work the 100 book titles in the domain of Library and Information Science (LIS) are selected which were published during 2000 and 2015. These titles and catalogue details were collected from Library of Congress (LOC) online catalogue http://catalog.loc.gov/index.html. These titles were also searched in LibraryThing https://www.librarything.com to collect the social tags assigned to these books. LibraryThing is a cataloguing and social networking site where users can contribute tags, reviews and ratings for a book and common knowledge about the book. Essentially it was noted that these selected books should have at least two tags assigned by users. These social tags were gathered from the tag cloud of selected books indicated with numbers of top frequency tags.

The duplicate terms were removed and unique tags were identified. In user generated tags, it is interesting to note that a few terms are unrelated, non-contextual and misspelled, as these tags are allocated by large number of users in uncontrolled, unrestricted and free-flow environment. Such unrelated tags were removed from the corpus through WordNet and Google search to accommodate them as meaningful words. The WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Few words were also searched in Google to confirm the context of the social tag.

The Library of Congress catalogues books published all over the world with bibliographic details. For each record we explored the Field 6XX (MARC 21 tag 650 in particular) where Library of Congress Subject Headings (LCSHs) were listed. These professionally allocated terms were stored to spread sheets and duplicate entries were removed. In the process, we have gathered the key words contained in Field and Subfields of 6XX as separate subject terms instead of Subject Headings. Even Subject Heading combinations were split into several concept terms to make them as unique terms. For example, the subject heading string was **Book industries and trade-Vocational guidance-United States** and it was split into **Book, industries, trade, vocational, guidance, United States.**

These selected book details were searched with their 'title' in LibraryThing website to identify the social tags assigned by users to these books. This experiment was undertaken and tags were extracted during March 2015. Consequently, we could extract 341 unique LCSH keywords and 2476 tags for these 100 titles.

Normally, it is noted that large numbers of social tags are assigned to these books in comparison to LCSH keywords due to the fact that LCSH are professionally assigned terms, where as social tags are user allocated. It is specified that there are 2476 frequent tags connected with these selected 100 titles with an average of 24.76 tags per book. After removing the duplicate entries the unique tags came down to 744. In case of LCSH terms this was 341 subject headings with an average of 3.41 terms per book.

This collected dataset was analysed by looking at the social tags and subject terms as two set of terms where in distribution of these two terms were organised, based on the most frequently used terms at the top and the least used terms at the bottom. Here for this work we have adopted Cosine similarity technique for determining the similarity coefficient.

Cosine similarity measure

Cosine similarity is literally the angular difference between two vectors. The similarity may be defined as the amount of how much two or more objects are alike. Similarity can also be seen as the numerical distance between multiple data objects that are typically represented as value between the range of 0 (not similar at all) and 1 (completely similar) [26]. For social tags many researchers have also used Jaccard similarity coefficient and Cosine similarity method[27-28].

Cosine based Similarity is perhaps the most popular metric and sophisticated way to measure similarity between two vectors in n-dimensional Euclidean space. It is often used when comparing two documents against each other. It measures the angle between the two vectors. If the value is zero the angle between the two vectors is 90 degrees and they share no terms. If the value is 1 the two vectors are the same except for magnitude [29]. Cosine measure is used when data is sparse, asymmetric and there is a similarity of lacking characteristics. The social tags collected and LCSH keywords are represented in vector representation to measure cosine similarity[30]. The

researchers have also adopted cosine based similarity technique between two datasets to measure Cosine value for the data extracted.

For this work, we have selected top 20 most frequently used terms appeared in both LibraryThing tags and Library of Congress words with their term frequency (Table 1). With this data we can create multidimensional points where these set of terms represent two vector points (Table 2). These vectors deal only with numbers. Hence the cosine similarity is equal to the cosine of the angle between them, **Theta**[31]**.**

The Cosine Similarity of two vectors (d1 and d2) is defined as:

$$\text{Cos}(\text{ d1,d2 }) = \frac{\text{dot (d1,d2)}}{\text{||d1|| ||d2||}}$$

where dot (d1, d2) = d1 [0]*d2 [0] + d1 [1]*d2 [1] …
and where ||d1|| = Sqrt (d1 [0] ^2 + d1 [1] ^2 …)

||d2|| = Sqrt (d2 [0] ^2 + d2 [1] ^2 …)

In this work, let d1 be LT tags and d2 be LOC words. By replacing the relevant values, following calculation is done to determine Cosine value of these two vectors.

Let d1 = 166 142 132 132 78 68 50 48 44 38 38 38 35 27 24 24 22 21 21 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Let d2 = 23 0 81 0 8 7 0 0 0 0 0 0 0 0 0 0 0 13 0 0 11 14 6 7 11 58 7 10 56 7 7 9 14 38 7

The formula used to measure Cosine similarity is as mentioned below.

$$\text{Cosine Similarity (d1, d2)} = \frac{(d1, d2)}{||d1||||d2||}$$

dot (d1, d2) = **15883**

||d1|| = **330.54651715**

||d2|| = **128.10932831**

$$\text{Cosine Similarity (d1, d2)} = \frac{15883}{(330.54651715)(128.10932831)}$$

$$= \frac{15883}{42346.0922872}$$

Cosine Similarity (LT, LOC) = **0.375075931263**

Analysis and discussion

In this paper we explored the cosine similarity measure between top 20 high frequency LibraryThing tags and Library of Congress words by representing vector format. It is observed that cosine score is 0.375 which indicates 38% of similarity in top high frequency words in both vectors. The vocabulary assigned by users in the form of tags is much less similar to controlled vocabularies. This is purely mathematical expression of words analysed by applying formula and by clustering together these set of words. It can implicate the difference between user generated tags and expert assigned words to the resources that we selected for this work. This mathematical observation is in void of semantic meaning of the words which leads to calculate the degree of similarity or dissimilarity of the selected dataset.

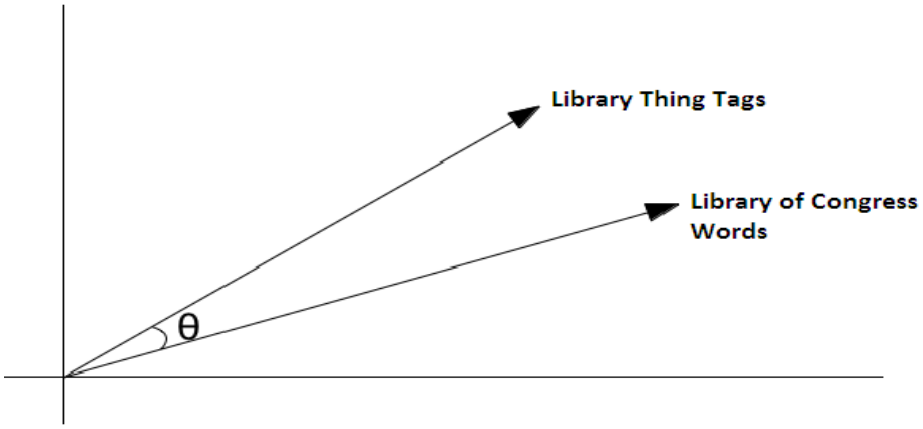

Fig 1—The Cosine angle between LibraryThing tags and Library of Congress words

Table 1—List top 20 high frequency words from LibraryThing tags and LOC words

LT	Freq	LOC	Freq
libraries	166	Library & Information Science	81
non fiction	142	General	58
library and information science	132	LANGUAGE ARTS & DISCIPLINES	56
online searching	132	United States	38
Reference	78	Libraries	23
librarian	68	Books and reading	14
Career	50	Study and teaching	14
librarianship	48	Information science	13
books	44	Bibliography	11
guide	38	EDUCATION	11
reading	38	Information technology	10
searching	38	Professional Development	9
technology	35	Reference	8
information retrieval	27	Collection Development	7
internet	24	Information literacy	7
LIS 9006	24	Librarian	7
children's literature	22	Libraries and the Internet	7
information science	21	Library education	7
textbook	21	Young adults' libraries	7
to-read	21	Children	6

Table 2—The words with their frequency represented in Vector format

WORDS from LT and LOC	LT	LC	WORDS from LT and LOC	LT	LC
libraries	166	23	textbook	21	0
non fiction	142	0	to-read	21	0
library and information science	132	81	Bibliography	0	11
online searching	132	0	Books and reading	0	14
Reference	78	8	Children	0	6
librarian	68	7	Collection Development	0	7
Career	50	0	EDUCATION	0	11
librarianship	48	0	General	0	58
books	44	0	Information literacy	0	7
guide	38	0	Information technology	0	10
reading	38	0	LANGUAGE ARTS & DISCIPLINES	0	56
searching	38	0	Libraries and the Internet	0	7
technology	35	0	Library education	0	7
information retrieval	27	0	Professional Development	0	9
internet	24	0	Study and teaching	0	14
LIS 9006	24	0	United States	0	38
children's literature	22	0	Young adults' libraries	0	7
information science	21	13			

This work also answers the RQ. 1 about the relationship between social tags and controlled vocabularies in very distinct manner by determining the cosine angle between LibraryThing tags and LOC for the selected group of words. The words allocated by users vary in large scale in comparison to professionals. Due to free nature of the tags by users make them less similar to professionally assigned vocabularies. In this context it implies that the folksonomies may not necessarily enhance the

metadata value of the resources in very big or impressive manner. The influences of metadata enrichment through social tagging are nominal in their values.

However, even if we find 38% of words are similar, the non-similar words from LibraryThing tags may also give good retrieval results for users. For example, by observing Table 1, the word 'non-fiction' is not there in LOC, but it is third most popular word in LT tags. The parallel word is 'General' in LOC which has got frequent mention but users find their own way to assign the keyword. Users tend to make distinction of the book they refer. Similarly, many words from LT tags indicate good value for the retrieval and if the popularity is also considered. 'Career' 'Librarianship' 'Information retrieval' 'Children's literature' such words find in LT tags with great frequency which are absent in LOC. These words would also enhance the retrieval effect while searching in a database. Therefore, for RQ1 the relationship between LT and LOC are complementary in nature where social tags add value to controlled vocabularies for resource discovery.

The RQ. 2 is addressed sufficiently by determining the cosine score between LT tags and LOC words. With this score of 0.375, we can notice how these social tags and controlled vocabularies are relatively different from each other and at the same time how these two sets of tags are able to complement each other. Generally, the social tags are assigned more in numbers by users for their own references which may sometime help others to access these tagged resources. But these donated descriptors in comparison to controlled vocabularies (like professionally assigned descriptors from LOC) may not be structured or networked but with 38% of similarity, LT tags may complement to LOC in retrieval (as revealed in this study). It is interesting to know that how the LT tags supplement to LOC for information retrieval of resources. The dataset from Table 2 shows that 15 LT descriptors used frequently by readers are absent in set of descriptors from LOC. It suggests the popular words may not find a place in professional vocabulary control device, but find high degree of acceptance among general users. This gap may be bridged in designing library discovery systems where social tags donated and assigned by common users will automatically move into the retrieval system to enhance the efficiencies of resource discovery. To elaborate, let us consider the words 'librarianship' and 'career' from the LT tags, which won't find place in LOC words. These popular words are assigned by users are not recommended by professionals in LOC. In this study, as we have considered books from Library and information science, it is but natural for users to tag as 'librarianship' and 'career' which is right from users' point of view and for further use in accordance to user warrant. The professionals assign keywords to these books in context of contents and thereby ensure literary warrant. Hence, library discovery systems with the facilities of social tagging and subsequent inclusion of those tags in retrieval system may be considered as an ideal mechanism to bridge the gap between user warrant and literary warrant. Therefore, it is obvious that social tags definitely complement to controlled vocabularies but may not replace them fully. This answers RQ2 sufficiently and shows the complementary nature of social tags and controlled vocabularies.

Conclusion and future work

The study of cosine similarity technique is one of the most important issues in the context of information retrieval process. This research work prominently tries to highlight the relation between social tags and controlled vocabularies by representing them in vector space to determine the cosine value for them. The cosine score reveals similarity or dissimilarity between tags and vocabularies which is expressed in mathematical value and not by semantic meaning of the words chosen. Hence meaning of the word has no role in determining the cosine value for these set of terms. This study of social tags proves the fact that they could not replace the value of controlled vocabularies in the context of information retrieval (IR). The controlled indexing has greater IR value than social tags for efficient retrieval results. The future studies need to be carried out by increasing the number of social tags and descriptors from controlled vocabularies to test if there is any variance in the cosine similarity score for better understanding of the complementary and supplementary relation between user warrant and literary warrant.

References

1. Anfinnsen S, Ghinea G and De Cesare S, Web 2.0 and folksonomies in a library context. *International Journal of Information Management*, 31(1)(2011) 63–70.

2. Furner J and Tennis J, 17th Annual ASIS&T SIG/CR Classification Research Workshop Saturday, November 4, 2006 – Austin. Available at http://ella.slis.indiana.edu/-~klabarre/SIGCR.html (Accessed on 25 April, 2016).
3. Pink D H, Folksonomy. *The New York Times* (2005) Available at http://www.nytimes.com/2005/12/11/magazine/-11ideas1-21.html (Accessed on 17 April, 2016).
4. Sterling B, What's the best way to tag, bag, and sort data? Give it to the unorganized masses (2005). Available at http://www.wired.com/wired/archive/13.04/view.html?pg=4 (Accessed on 17 April, 2016).
5. Golder S and Huberman B A, Usage patterns of collaborative tagging systems, *Journal of Information Science*, 32(2) (2006) 198–208.
6. Vander Wal T, Folksonomy (2007). Available at: http://-vanderwal.net/folksonomy.html (Accessed on 17 April, 2016).
7. Shirky C, Ontology is overrated: categories, links, and tags (2005). Available at: http://www.shirky.com/writings/ontology_overrated.html (Accessed on 23 April, 2016).
8. Furner J, User tagging of library resources: toward a framework for system evaluation. *International Cataloguing and Bibliographic Control*, 37 (2007) 47–51.
9. Peters I and Stock W G, Folksonomy and information retrieval, *Proceedings of the American Society for Information Science and Technology*, 44(1) (2007) 1–28.
10. Macgregor G and McCulloch E, Collaborative tagging as a knowledge organisation and resource discovery tool, *Library Review,* 55(5) (2013) 291-300.
11. Hotho A, Jäschke R Schmitz C and Stumme G. *Information Retrieval in Folksonomies: Search and Ranking*. In The Semantic Web: Research and Applications. Springer Berlin Heidelberg. 2006, p. 411–426. Available at: http://link.springer.com/chapter/10.1007/11762256_31 (Accessed on 23 April, 2016).
12. Kipp M E I, Comparing controlled vocabularies and tags: research methodologies and research goals (2011). Available at: http://www.cais-acsi.ca/proceedings/2011/24_Kipp.pdf (Accessed on 23 April, 2016).
13. Kroski E, The Hive Mind: Folksonomies and User-Based Tagging (2005). Available at: http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/ (Accessed on 12 April, 2016).
14. Cleverdon.pdf. Available at: https://www.ischool.utexas.-edu/~stratton/rdgs/Cleverdon.pdf (Accessed on 21 April, 2016).
15. Morrison J P, Tagging and Searching: Search Retrieval Effectiveness of Folksonomies on the Web (2008). Available at: http://etd.ohiolink.edu/sendpdf.cgi/Morrison,%20Patrick-%20Jason.pdf?acc_num=kent1177305096 (Accessed on 12 April, 2016).
16. Kipp M E and Campbell D G Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices (2006). Available at: http://arizona.openrepository.-com/arizona/bitstream/10150/105181/1/KippCampbellASIST.pdf (Accessed on 20 April, 2016).
17. Lu K and Kipp M E I, An experimental study on the retrieval effectiveness of collaborative tags (2010). Available at: https://mail.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/379_Final_Submission.pdf (Accessed on 20 April, 2016).
18. Thomas M, Caudle D M and Schmitz C, Trashy tags: problematic tags in LibraryThing, *New Library World*, *111*(5/6) (2010) 223–235. Available at: http://www.emerald-insight.com/journals.htm?articleid=1858820&show=abstract (Accessed on 23 April, 2016).
19. Lu Caimei, Park J and Hu X, User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings (2010). Available at: http://jis.sagepub.com/content/36/6/763.full.pdf (Accessed on 26 March, 2016).
20. Voorbij H, The value of LibraryThing tags for academic libraries, *Online Information Review*, *36*(2) (2012) 196–217. Available at: http://doi.org/10.1108/14684521211229039 (Accessed on 26 March, 2016).
21. Lee D H and Schleyer T, Social tagging is no substitute for controlled indexing: A comparison of Medical Subject Headings and CiteULike tags assigned to 231,388 papers *Journal of the American Society for Information Science and Technology,* 63 (9) (2012) 1747-1757.
22. Syn S Y and Spring M B, Finding subject terms for classificatory metadata from user-generated social tags, *Journal of the American Society for Information Science and Technology*, 64(5) (2013) 964–980.
23. Lu K and Kipp M E I, Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: An experimental study on medical collections, *Journal of the Association for Information Science and Technology*, 65(3) (2014) 483–500.
24. Choi Y and Syn S Y, Characteristics of tagging behavior in digitized humanities online collections, *Journal of the Association for Information Science and Technology*, 67(5) (2016) 1089–1104.
25. Yi K, A semantic similarity approach to predicting Library of Congress subject headings for social tags, *Journal of the American Society for Information Science and Technology*, 61(8) (2010) 1658–1672.
26. Bakos, Yong Joseph, Data mining portfolio: Similarity techniques (2009). Available at: http://mines.humanorient-ed.com/classes/2010/fall/csci568/portfolio_exports/bfindley/similarity.html (Accessed on 26 March, 2016).
27. Wu D, He D, Qiu J, Lin R and Liu Y, Comparing social tags with subject headings on annotating books: A study comparing the information science domain in English and Chinese, *Journal of Information Science*, 39(2) (2013) 169–187.
28. Yi K and Chan L M, Linking folksonomy to Library of Congress subject headings: an exploratory study, *Journal of Documentation*, *65*(6) (2009) 872–900.
29. Salton G and McGill M J, *Introduction to Modern Information Retrieval.* (McGraw-Hill, Inc.; New York), 1986. Available at: http://dl.acm.org/citation.cfm?id=576628 (Accessed on 26 March, 2016).
30. Baeza-Yates R A and Ribeiro-Neto B, *Modern Information Retrieval*. (Addison-Wesley Longman: Boston), 1999. Available at: http://dl.acm.org/citation.cfm?id=553876 (Accessed on 26 March, 2016).
31. Richard, (2010), "Getting Cirrius: Calculating Similarity (Part 1): Cosine Similarity", available at: http://www.gettingcirrius.com/2010/12/calculating-similarity-part-1-cosine.html (Accessed on 26 May, 2016).