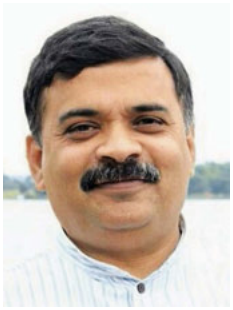
The Comparative and Analytical Study of LibraryThing Tags with *Library of Congress Subject Headings*

Praveenkumar Vaidya* and N. S. Harinarayana**

*Tolani Maritime Institute, Induri, Pune. Maharashtra, India, <vaidyapraveen@gmail.com>
** University of Mysore, Mysuru, Department of Studies in Library and Information Science, Karnataka, India, <ns.harinarayana@gmail.com>

Mr. Praveenkumar Vaidya is a research student at the Department of Studies in Library and Information Science, University of Mysore, Mysuru. Presently, working as Librarian at Tolani Maritime Institute, Pune, Maharashtra.

Dr. N. S. Harinarayana is working as an associate professor at the Department of Studies in Library and Information Science, University of Mysore, Mysuru. He teaches the courses library automation, information retrieval, and library classification and content organization. He has had the experience of working at Indira Gandhi National Open University, New Delhi as Assistant Librarian for six years.



Vaidya, Praveenkumar and Harinarayana, N. S. "The Comparative and Analytical Study of LibraryThing Tags with *Library of Congress Subject Headings*." *Knowledge Organization* 43 no. 1: 35-43. 30 references.

Abstract: The internet in its Web 2.0 version has given an opportunity among users to be participative and the chance to enhance the existing system, which makes it dynamic and collaborative. The activity of social tagging among researchers to organize the digital resources is an interesting study among information professionals. The one way of organizing the resources for future retrieval through these user-generated terms makes an interesting analysis by comparing them with professionally created controlled vocabularies. Here in this study, an attempt has been made to compare *Library of Congress Subject Headings* (*LCSH*) terms with LibraryThing social tags. In this comparative analysis, the results show that social tags can be used to enhance the metadata for information retrieval. But still, the uncontrolled nature of social tags is a concern and creates uncertainty among researchers.

Received: 3 October 2015 Revised: 25 November 2015 Accepted: 9 December 2015

Keywords: tags, terms, subject headings, LibraryThing, *Library of Congress Subject Headings*

1.0 Introduction

Recently, social tagging has emerged as a new approach for creating library metadata. Social annotation or collaborative tagging is a popular activity among researchers in this era of Web 2.0. The inception of the first social bookmarking system del.cio.us in 2003 led to one of the most common ways to organize data in sites like del.cio.us, Flickr, YouTube, CiteULike and LibraryThing. The general purpose of the social tagging system is to help users store, share, organize and retrieve the resources of their choice when in need. Hence, social tagging (Golder and Huberman 2006) can be viewed as a process by which many users add metadata in the form of keywords to shared content. Social tags reveal a personal interpretation of resources which would be sharable with others. These are free-form keywords that provide an opportunity for information institutions to expand the access points of their resources beyond professionally created metadata and index terms (Hotho 2006), thus improving the accessibility of their resources.

The analysis of these social tags reveals the opinion of users, interpretation of items and communities of practice and gender also. These provide an insight into how

users characterize resources (Mathes 2004) and thus help and guide professionals towards improved resource description and discovery tools. Information professionals (Kipp 2006; Tennis 2006; Hotho 2006; Magnuson 2009; Fox 2012; Fox and Reece 2013) have recognized the potential of tags in information retrieval and knowledge organization systems. Thus, information discovery or image discovery (Konkova et al. 2014) at this juncture of information overload justifies the research studies conducted on this trend, even though tagging is done in an uncontrolled environment with users' own verbal descriptors. Social tags are also blamed by Kipp (2011) for imprecision, semantic ambiguity and lack of hierarchy. Despite these issues, researchers such as Peters et al. (2011), Ding et al. (2009) and Kipp and Campbell (2010) insist on the coexistence of social tags and controlled vocabularies.

The controlled vocabulary which is also known as expert-created vocabularies is considered of high quality with good precision in retrieval. But the high cost to build them and to scale up in the context of immense volumes of digital resources becomes a constraint. The users are subject specialists in their narrow field (Hjørland 2006) but might be undermined by the use of advanced technology to organise the knowledge by library professionals. This may create a barrier for users to search professionally built subject catalogues, which may result in complete inaccessibility to resources. But still it depends on the nature of the user also, who may be a purposive or non-purposive reader. The purposive reader is more familiar with the library and knows how the traditional library classification works. This is not usually the case with the non-purposive reader. Traditional library classifications and catalogs are therefore considered to be more adequate for purposive readers (who are in the minority), while nonpurposive users (who are in the majority) are assumed to have access to the materials by browsing. The distinction between purposive and non-purposive readers (Martínez-Ávila and San Segundo 2013) is thus sometimes defined by the relationship between access methods and the knowledge of the system. Interestingly, in this case, the social tags may prove rich information access and discovery for both categories of users.

This research is done to examine whether there is any evidence of differences available and if so would be demonstrated by descriptive statistics. The result of this study with term comparison also could reveal the thesaural relations, related tags, unrelated tags, time and task management tags, geographic tags and specific groups, general vocabularies and also emerging terms.

2.0 Objective of the study

This research examines the differences and connections between social tags and controlled vocabularies, which are the primary source of library metadata. This study is an attempt to shed light on the following research questions (RQ):

RQ1. Do social tags comprise similar vocabulary to that of controlled vocabulary?
RQ2. Can social tags can be used for metadata enrichment?
RQ3. To what extent do user generated tags provide different words not available in controlled vocabularies?

For this research, we have collected social tags from LibraryThing and compared them with a Library of Congress controlled vocabulary, the *Library of Congress Subject Headings* (*LCSH)*.

3.0 Related literature

The research on social tagging (Bartley 2009) has motivated many information professionals to embark on interesting studies and have suggested various layers of social tagging in terms of precision, disambiguation, and quality of tag terms, meaningfulness and their usefulness. The comparative study of social tags and controlled vocabularies (Peters et al. 2011) reveals an interesting analysis of retrieval effectiveness of tags. The results show that folksonomies work best with short queries although recall values are high and precision values are low.

In the following studies, the comparison of social tags has been made with other types of controlled vocabularies. Rolla (2009), Thomas et al. (2009) and Lu et al. (2010) have studied comparative work between LibraryThing tags and *LCSH*, and the objective of this study was to examine the difference and connections between the tag systems, exploring the feasibility and obstacles of implementing social tagging in library systems. The dataset was collected from the Library of Congress bibliographic records in MARC (Machine Readable Cataloging), and only ISBN (International Standard Book Number) records were crawled and respective titles were extracted from the LibraryThing website. The tags assigned to those titles were used for the study with the subject information shown in MARC records. The comparison results suggest that while user tags can enhance subject access to library collections, they cannot replace the valuable functions of controlled vocabulary like *LCSH*.

In another work, Lu, et al. (2010), the analysis was done at two levels: the collection level and book level. At the collection level, the comparison was done with the social tags and subject terms present in the whole dataset, and at the book level, social tags are compared to subject terms applied to the same book. The researchers checked

the frequency or popularity of the overlapping terms in tags and *LCSH*s and were represented in statistical charts. The study also reveals the social tags might help to enhance the subject access to library collections by describing library resources with terms different than those used by experts, and results also indicate the same.

Similarly, in another study, Matthews et al. (2010), with the same approach, tried to investigate the semantic interoperability in the context of repositories and digital collections. In the study, user-generated catalogue records were chosen from subject areas of political science and other Science and Technology Facilities Council's repository in which authors can allocate tags for the articles they submit to the repository. These records were classified by the *DDC* (*Dewey Decimal Classification*) *IBSS* (*International Bibliography of Social Sciences Thesaurus* 2009) and *HASSET* (*Humanities and Social Science Thesaurus*). The study found that adding terms for information retrieval has the advantages of consistency of retrieval.

Kakali and Papatheodorou (2011) in their study, in which librarians and cataloguers assessed the value of the semantics of inserted tags from LibraryThing and also investigated the possibility of using them for subject indexing, examined the semantic value of the tags that are not included in the local authority file of the library. For this purpose, five systems were selected, namely the Library of Congress Authorities, Greek National Documentation Centre (NDC) Thesaurus, Thesaurus of Social Sciences Index Terms (SSIT), Wikipedia and WordNet. This selection was based on three criteria: coverage, language and relevance. The researchers admitted that social tagging could help them to approach the user's way of thinking effectively as well as to observe the community's terminology evolution.

Yi and Chan (2009) in their work, tried to link folksonomies to *LCSH* with emerging Faceted Application of Subject Terminology (FAST) methodology. The results of the study can be adopted for the development of innovative methods of mapping between folksonomies and LCSH, which contributes to effective access and enhancement in retrieval.

Lu et al. (2010), Lee and Schleyer (2012) and Wu et al. (2012) used the Jaccard Similarity Coefficient method to measure the similarity between the frequent sets of tags and controlled terms. These researchers found it was appropriate to use the Jaccard Similarity Coefficient for the type of dataset to be used in their study to compare social tags and subject terms. The results show that the folksonomies and controlled vocabularies are quite distinct lexically and semantically, reflecting the different viewpoint and processes between them. However, the lexical overlap between the two corpora is minimal. Nevertheless, despite limitations, the social tags have the potential to become a complementary source to expand and enrich the controlled vocabulary system.

In another study, Voorbji (2012) conducted a study to determine the value of LibraryThing tags, where the random sample of 600 records, evenly distributed among humanities, social sciences and natural sciences were taken from the library catalogue. The study revealed the nature and value of the tags by further categorization. This study focuses on the importance of professional subject indexing and concludes that replacing them with user-generated tag assignment would be detrimental for the recall. With the uncontrolled nature of folksonomies, tags are inherently imprecise, inexact and overly personalized. Hence, the result is chaotic and negatively affects the retrieval where a user's search term would not match with a controlled vocabulary.

Similarly, Lu and Kipp (2014) investigated the retrieval effectiveness of collaborative tags and author keywords in different environments through controlled experiments. The findings suggest that including tags and author keywords in indexes can enhance recall but may improve or worsen average precision depending on retrieval environments. The experimental design of this study follows Cranfield paradigms. To conduct a retrieval test, a test collection, a list of topics and relevant judgements are needed. Another interesting study by Choi and Syn (2015) 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.

In summary, the previous studies indicate that an evaluation of tags can provide insight into users' interpretation of the content of resources that will be significant and beneficial for other users. This study also complements the previous works of researchers, and an attempt is made to compare the social tags of library and information science books with *LCSH*.

4.0 Methodology and data collection

The subject area chosen for this work was library and information science. It was important to collect the "titles of books" published between 2000 and 2015. The Library of Congress (LOC) online catalogue, http://catalog.loc.gov/index.html, was used to search these publications. The same titles with ISBNs were also searched in LibraryThing, https://www.librarything.com, to collect social tags. LibraryThing is a cataloguing and social networking site (Lu et al. 2010) where users can annotate tags, ratings and reviews for the book. Care was taken to collect such titles from LibraryThing by assigning a minimum of two tags.

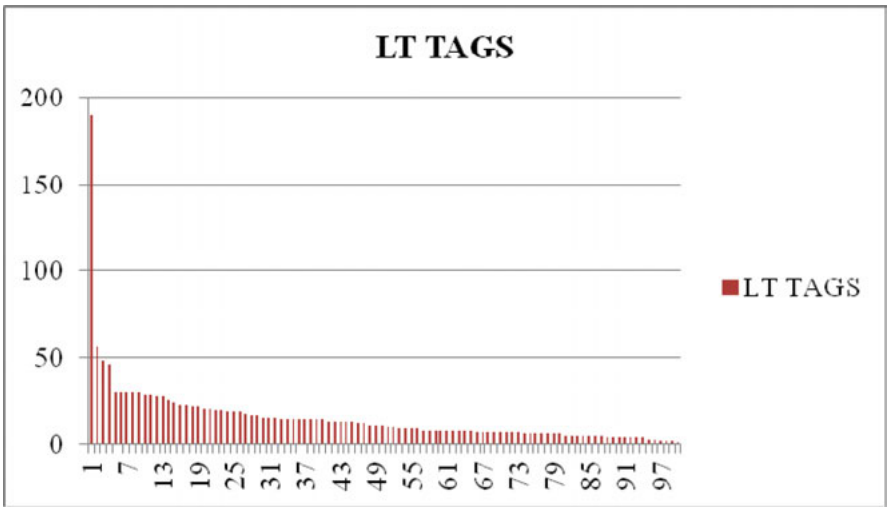


Chart 1. Tag application frequency per document for 100 documents

One hundred such titles were selected for the study. The social tags were collected from the main page of the tag cloud of selected books with numbers which usually includes only top frequency tags.

These social tags were exported to a spreadsheet software to enable us to assess the duplicate terms and to identify unique tags. All duplicate, trashy tags (Thomas et al. 2010) and repetitive entries were removed to identify unique tags. In social tags, it is very common to find few unidentifiable, inconsistent and messy tags along with jargon and tags with non-alphabetic characters because of their free character, as these tags are assigned by a large number of users in a totally uncontrolled and free-flow environment. Such tags were removed searching WordNet and Google to discover if they were meaningful words. WordNet, http://wordnetweb.princeton.edu/perl/webwn, 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. Some words were also Googled to confirm "context" of the social tag.

The Library of Congress catalogues books with "Full Record" along with "MARC Tags." For each record, we extracted the contents of MARC field 6XX where the *LCSH* headings are listed. Such expert assigned data was exported to a spreadsheet software to remove duplicate entries. In our study, we treated the keywords contained in the 6XX fields and subfields as separate subject terms instead of subject headings. Even subject heading combinations were split into several concept terms and duplicates were removed. For example, the subject heading string "Libraries-Activity programs-United States-Case studies" was split into "Libraries," "Activity programs," "United States" and "Case Studies."

The selected records with their ISBNs were crawled on LibraryThing to search the corresponding titles and user generated tags that were extracted during August 2015. In total, we could extract 341 unique LCSH keywords and 2,476 tags, out of which 744 are unique tags for these 100 titles. We removed a few trashy tags from the collection by searching them in WordNet, and a few tags were also searched in Google. For example, the term "Do-it-your Self" is listed as an *LCSH* term, but in the LT Tags, it is marked as "DIY" even though the WordNet search does not give any result to this word. In this case, the term was searched in Google to authenticate about the correctness of the term usage.

Usually, we can observe large numbers of social tags in contrast to *LCSH* keywords in such works due to the simple fact that headings are expert assigned whereas social tags are user driven. It is indicated that there are 2,476 tags associated with the selected 100 titles with an average of 24.76 tags per work. However, when duplicate entries were removed, the count came down to 744 unique tags. In the case of *LCSH* words, there were 341unique subject headings with an average of 3.41 terms for each book.

5.0 Data analysis and results:

5.1 Vocabulary overlaps:

This portion of the study considers the comparison of social tags and *LCSH* terms. The tags and *LCSH* terms associated with the same work were compared in a term-by-term manner. The overlap was identified with an exact or almost exact match in spelling even including singular or plural forms with case variation. The abbreviations or acronyms were considered the same as the full form of

the words and prepositions, punctuation marks and symbols were ignored to make them a comprehensive set of terms.

First, we compared an entire set of LT (LibraryThing) unique tags with *LCSH* terms to the entire collection of 100 books. It was found that only 160 tags are common tags assigned by users as well as experts.Therefore, it was found that only 21.5% of LT tags are in common with *LCSH* keywords. Further, 78.5% of the LT tags do not appear in *LCSH* terms, but also 46.9% of *LCSH* terms were used by taggers to assign to this collection. Please refer to Figure 1 for the data representation in a Venn diagram.

However, the rest of the terms that did not overlap with *LCSH* were also relevant but different from subject headings (Table 1, Figure 2).

5.2 Book level comparison

If we observe the LT tags closely, these tags can be analysed in many classification frames also. At the book level, we can compare them with subject headings for each book.

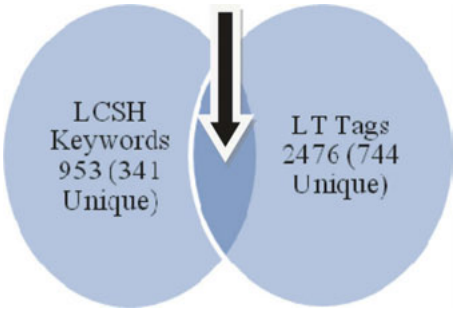


Figure 1. 160 tags are common both in *LCSH* keywords and LT tags

Tag	Frequency	Tag	Frequency
libraries	166	children literature	22
library and information science	132	information science	21
librarian	68	databases	17
Career	50	business	13
librarianship	48	information literacy	13
books	44	Collection Development	12
guide	38	information	12
information retrieval	27	information technology	11
internet	24	ebook	10
LIS 9006	24	education	9

Table 1. Twenty most frequently indexed terms representing reader perspective

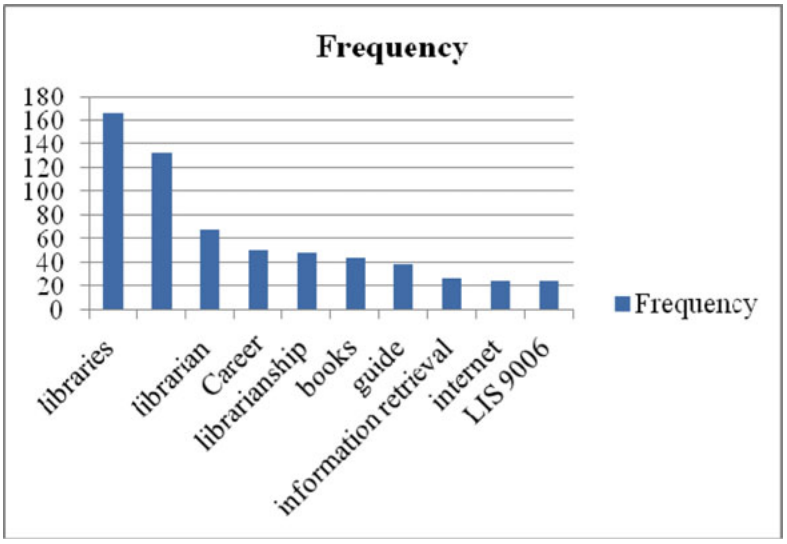


Figure 2. Ten most frequently applied tags.

Some of the tags were associated with the bibliographic description of the title. The tags may include the name of the author, subject, language, timeframe, and country of origin. Many such tags appeared in this work. For example, "The USA," "Library and Information Science," "18th Century," "English," etc. A few tags were related to personal reference. Examples include "to read," "todo," "best books," "new," "my books," "not to read," "not read yet," "borrowed," etc. It was not surprising to find opinionated tags like "moved."

With further analysis, the tags and subject headings at the individual book level are quite interesting. From the list of 100 books, 12 (12%) of them do not have a single word common between *LCSH* and LT tags. The remaining 88 (88%) books have a minimum of one tag in common from their list. This result means in most cases, at least for one word, both users and experts agree to describe the source. Among these 100 books, at least 35 (35%) books have more than five similar tags and up to nine for one particular book. We calculated the 160 common unique terms annotated by both users and subject experts, which indicates both have some similar understanding about the semantics and understanding of the term assigned.

5.3 Tag similarity: similar LT tags and Library of Congress Subject Headings

5.3.1 Tag similarity for the whole collection

To further understand whether tags share the controlled vocabulary of subject experts or represent the different set vocabulary of users, in this section we compare top n frequent *LCSH* terms. The Jaccard Similarity Coefficient (also called Jaccard index) is used to measure the similarity between the frequent sets of tags and *LCSH* terms (Lu, et al. 2010) and is calculated according to following equation.

$$J(T, L) = \frac{L \cap T}{L \cup T}$$

Where L is *LCSH* term and T is LT tags.

The Jaccard index, also known as the Jaccard similarity coefficient (originally coined coefficient de communauté by Paul Jaccard), is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard similarity coefficient measures the similarity between finite sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets. The Jaccard similarity coefficient (Niwattanakul et al. 2013) measures the share properties of both objects L and T whereas all of the objects L and T are represented by 0 and 1 respectively. This work has taken consideration of all *LCSH* terms and LT tags extracted from the 100 books. Replacing the values to find the Jaccard similarity coefficient, where L intersects T, is 160 and L union T is (341+744-160) 925, which works out to be 0.147, which should be less than one.

5.3.2 Tag similarity for top 35 tags

We can also calculate this index for the top 35 LT tags and *LCSH* terms respectively. In this context, it was calculated that the value of L was 470 and the value of T was 1,356. Replacing these values in the formula, the value of the Jaccard index is 0.44 which should be less than one, which indicates a very low overlapping terms between top frequent LT tags and top frequent *LCSH* terms. This also means the frequent terms used by subject experts and users are very different and even though the terms may be popular, subject experts may not assign them to resources.

As we calculate the Jaccard index for these sample sets, we can also determine dissimilarity of the same. The "Jaccard distance," which measures dissimilarity between sample sets, is complementary to the Jaccard similarity coefficient and is obtained by subtracting the Jaccard similarity coefficient from 1 or by dividing the difference between the sizes of the union and the intersection of the two sets of the size of the union.

$$dj(L,T) = 1 - J(L,T) = \frac{|L \cup T| - |L \cap T|}{|L \cup T|}$$

Replacing the values of the Jaccard distance, worked out to be 0.56. The dissimilarity index for the selected data indicates a very small percentage of similar words in the sets of words selected. The poor similarity proves that social tags are free in nature and have no explicitly defined relationship or hierarchy between the terms.

This set of data can also be tested for the "Cosine similarity." Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: (Wikipedia 2015) two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0,1].

5.3.3 Coverage ratio

The corpora of LibraryThing tags and *LCSH* terms in our study have different sizes, so we also calculated the coverage ratio. The coverage ratio is defined as the fraction of the common annotations for an article covered by

its *LCSH* terms and LibraryThing tags respectively. The coverage ratio is useful to determine whether to suggest existing LibraryThing tags or *LCSH* terms to users as annotations.

$$\text{Coverage Ratio of LCSH terms} = \frac{L \cap T}{L} \quad \text{i.e } 1.712$$

$$\text{Coverage ratio of LibraryThing Tags} = \frac{L \cap T}{T} \quad \text{i.e } 0.593$$

In this case, it may make sense to suggest existing tags to users, because they are most likely to contain appropriate terms for annotations as opposed to *LCSH* terms.

5.4 Tag overlap

A careful look at Tables 2 and 3 indicate the top 35 frequent tags and top 35 frequent *LCHS* terms which were extracted for the current work. Here we found 13 common words out of the 35 words from both LT tags and *LCSH* terms which is about 45.71%. However, this seems to be a higher side comparing with other similar works. The tags from LibraryThing closely resemble *LCSH*s for non-common words. For example, "library technology," "librarianship" and "children's literature;" these are well thought out and popular tags which are very close to a subject expert. However, tags like "nonfiction," "career," "profession," "read," "books about books" and "self-help" are popular descriptors which were annotated for personal recall.

Even the non-overlapped tags also reveal much information about the sources. These tags may be interpreted in a different way by representing the concepts and the semantic relationships among the terms and also differentiation in subject analysis between users and subject experts. If we just compare these tags with subject headings, these appear to be more exhaustive and describe more themes and topics covered in the source (Tables 2 and 3).

6.0 Conclusion

The Web 2.0 environment of the Internet has increased the participation and interaction of the users to the extent of providing their own terms to recall their resources. These user-generated tags might help other users to retrieve the resources and act as a bridge between the professional terms.

To answer RQ 1 in this work, despite the duplication of social tags comprising 21.5% of controlled vocabularies, it is clearly visible that social tags have the inherent limitation of uncontrolled language. The problems of homographs, synonyms and polysemy, are common in them. Therefore, skepticism and ambiguity still exist in the professional cataloguing community about the value of social tagging.

The analysis of RQ 2 definitely helps to understand the possibility of metadata enrichment in the context of social tagging. Even though the 160 tags found to be common in *LCSH* and LT tags, the rest of the tags reveal a different dimension and expansive view about the source which may be useful for source recall. If we consider the top 35 frequently used tags and terms, the popular tags may differ with expert terms but might help to enhance the subject access point for the library collection.

The RQ 3 has been answered adequately in this work about the words not available in a controlled vocabulary. The results show that, there is a 21.5% of common words, but many works also show that there is a very tiny

Books about books, business, Career**, children's literature**, **Collection Development**, databases , ebook, education, guide, information, **information literacy**, information retrieval, **information science** , **information technology**, internet, **librarian**, librarianship, **libraries**, **library and information science**, LIS 9006, literature, MLIS, non-fiction, online, searching, own, **professional** , reading, **Reference**, **research**, searching , technology, **teen**, textbook, to read, writing

Table 2. Top 35 frequent tags

Academic libraries, Activity programs, Administration & Management , Bibliography, **Books and reading**, Case studies**, Children**, **Collection Development**, Congresses, EDUCATION,Effect of technological innovations on, General, **Information literacy**, **Information science**, **Information technology**, LANGUAGE ARTS & DISCIPLINES, **Librarian**, **Libraries**, Libraries and electronic publishing, Libraries and society, Libraries and teenagers, Libraries and the Internet, **Library & Information Science**, Library education, **Professional Development**, **Reference**, Relations with faculty and curriculum, **Research**, School libraries, School Media, Special collections, Study and teaching, **Teenagers**, United States, Young adults' libraries,

Table 3. Top 35 frequent LCSHs

portion of common words available when compared. But non-overlapped words provide a richer description of the book's subject matter and is exhaustive in nature.

A major comparison of this study was limited to 100 books from the library and information science subject. The results may vary if the number of books is scaled up for the study. This work uses a statistical tool Jaccard index to determine the similarity, but Cosine similarity can also be used for similar data. But this research gives an indication that the LT tags are inconsistent at sometimes with no common tags at all and sometimes just a few. Still, the rules for comparison are to be discussed for some inconsistencies. For future work, the dataset would be more comprehensive and exhaustive which apply resources to investigate the overlapped and non-overlapped subject terms with other research methods.

References

Bartley, Peishan. 2009. "Book Tagging on Librarything: How, Why, and What are in the Tags?" *Proceedings of the American Society for Information Science and Technology* 46:1-22.

Choi, Youngok and Sue Yeon Syn. 2015. "Characteristics of Tagging Behavior in Digitized Humanities Online Collections." *Journal of the Association for Information Science and Technology* preview doi: 10.1002/asi.23472

Ding, Ying, Elin K. Jacob, Zhixiong Zhang, Schubert Foo, Erjia Yan, Nicolas L. George and Lijiang Guo. 2009. "Perspectives on Social Tagging." *Journal of the Association for Information Science and Technology* 60: 2388–401.

Fox, Melodie J. 2012. "Communities of Practice, Gender and Social Tagging". In *Categories, Contexts and Relations in Knowledge Organization: Proceedings of the Twelfth International ISKO Conference 6-9 August, 2012 Mysore, India*, edited by A. Neelameghan and K.S. Raghavan. Advances in Knowledge Organization 13. Würzburg: Ergon Verlag, 352-58.

Fox, Melodie and Austin Reece. 2013. "The Impossible Decision: Social Tagging and Derrida's Deconstructed Hospitality." *Knowledge Organization* 40: 260-5.

Golder, Scott A. and Bernardo A. Huberman. 2006. "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science* 32: 198–208.

Hjørland, Birger 2006. "Laymen as Knowledge Organizers". Accessed November 2, 2015 http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/laymen_as_knowledge_organizers.htm

Hotho, Andreas, Robert Jäschke, Christoph Schmitz and Gerd Stumme. 2006. "Information Retrieval in Folksonomies: Search and Ranking." In *The Semantic Web: Research and Applications: Proceeding of 3rd European Semantic Web Conference, 11-14 June 2006, Budva, Montenegro,* edited by York Sure and John Domingue. Berlin; New York: Springer, 411-426.

Kakali, Constantia and Christos Papatheodorou. 2011. "The Exploitation of Social Tagging in Libraries." In *Proceeding of the First Workshop on Digital Information Management, 30-31 March 2011, Corfu, Greece.* 76-88. http://eprints.rclis.org/15850/1/07.Kakali.pdf

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* 30: 1-15.

Kipp, Margaret E. I. 2011. "Comparing Controlled Vocabularies and Tags: Research Methodologies and Research Goals." In *Exploring Interactions of People, Places and Information, Proceedings of the Annual Conference of the Canadian Association for Information Science, University of New Brunswick/St. Thomas University, Fredericton, N.B. Canada, June 2 – 4, 2011, edited by* Pam McKenzie, Catherine Johnson and Sarah Stevenson http://cais-acsi.ca/cais-conferences/

Kipp, Margaret E.I. and, Grant D. Campbell. 2010. "Searching with Tags: Do Tags Help Users Find Things?" *Knowledge Organization* 37: 239-55.

Konkova, Elena, Ayşe Göker, Richard Butterworth and Andrew MacFarlane. 2014. "Social Tagging: Exploring the Image, the Tags, and the Game." *Knowledge Organization* 41: 57-65.

Lee, Danielle H. and Titus Schleyer,. 2012. "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: 1747-57.

Lu, Caimei, J. R. Park and Xiaohua Hu. 2010. "User Tags versus Expert-Assigned Subject Terms: A Comparison of LibraryThing Tags and Library of Congress Subject Headings." *Journal of Information Science* 36: 763-79.

Lu, Kun and Margaret E. I. Kipp. 2014. "Understanding the Retrieval Effectiveness of Collaborative Tags and Author Keywords in Different Retrieval Environments: An Experimental Study on Medical Collections." *Journal of the American Society for Information Science and Technology* 65: 483–500.

Magnuson, Lauren. 2009. "Folksonomies: Meaning, Discourse, and Information Retrieval". In *Mapping the 21st Century Information Landscape: Borders, Bridges and Byways: Proceedings of the Annual Conference of CAIS, 28-30 May 2009, Ottawa, Ontario*, edited by Paulette Rothbauer, Siobhan Stevenson and Nadine Wathen. http://www.cais-acsi.ca/ojs/index.php/cais/issue/view/25

Martínez-Ávila, Daniel and Segundo Rosa San. 2013. "Reader-Interest Classification: Concept and Terminology Historical Overview." *Knowledge Organization* 40: 102-14.

Mathes, Adam. 2004. "Folksonomies: Cooperative Classification and Communication through Shared Metadata." Accessed August, 15.2015 http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Matthews, Brian, Catherine Jones, Bartłomiej Puzoń, Jim Moon, Douglas Tudhope, Koraljka Golub and Marianne Lykke Nielsen. 2010. "An Evaluation of Enhancing Social Tagging with a Knowledge Organization System." *AP Aslib Proceedings* 62: 447-65.

Niwattanaku, Suphakit, Jatsada Singthongchai, Ekkachai Naenudorn and Supachanun Wanapu. .2013. "Using of Jaccard Coefficient for keywords Similarity." In *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013, 13-15 March 2013, Kowloon, Hong Kong*, edited by S. I. Ao, Oscar Castillo, Craig Douglas, David Dagan Feng and Jeong-A Lee. Hong Kong: Newswood Limited, 380-4.

Peters, Isabella, Laura Schumann, Jens Terliesner and-Wolfgang G. Stock.. 2011. "Retrieval Effectiveness of Tagging Systems." *Proceedings of the American Society for Information Science and Technology* 48, no. 1: 1–4.

Rolla, Peter J. 2009. "User Tags versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?" *Library Resources & Technical Services* 53: 174-84.

Tennis, Joseph T. 2006. "Social Tagging and the Next Steps for Indexing." *Advances in Classification Research Online* doi:10.7152/acro.v17i1.12493

Thomas, Marliese, Dana M. Caudle and Cecilia M. Schmitz. 2009. "To Tag or Not to Tag?" *Library Hi Tech* 27, no. 3: 411-34.

Thomas, Marliese, Dana M. Caudle and Cecilia Schmitz. 2010. "Trashy Tags: Problematic Tags in LibraryThing." *New Library World* 111, nos. 5/6: 223-35.

Voorbij, Henk. 2012. "The Value of LibraryThing Tags for Academic Libraries." *Online Information Review* 36, no. 2: 196-217.

Wikipedia 2015. "Cosine similarity" https://en.wikipedia.org/wiki/Cosine_similarity

Wu, Dan, He Daqing, Jin Qiu, Ruonan Lin and Yang Liu. 2012. "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: 169-87.

Yi, Kwan and Lois Mai Chan. 2009. "Linking Folksonomy to Library of Congress Subject Headings: An Exploratory Study". *Journal of Documentation* 65: 872-900.