



Measuring the applicability of user-generated social tags along with expert-generated LCSH descriptors in Sociology: a heuristic study

Kalyan Sundar Samanta^a and Durga Sankar Rath^b

^aLibrarian, Prabhu Jagatbandhu College, Howrah-711302, West Bengal, India &

Research Scholar, Department of Library and Information Science, Vidyasagar University, Midnapore-721 102, West Bengal, India,
Email: kalyansundarsamanta@gmail.com

^bProfessor, Department of Library & Information Science, Vidyasagar University, Midnapore-721102, West Bengal, India,
Email: dsrath@mail.vidyasagar.ac.in

Received: 14 June 2020, revised: 06 February 2021, accepted: 03 March 2021

The study attempts to compare user-generated social tags with expert-generated LCSH descriptors of one thousand sociology books. The objective is to examine if social tags can be used to enhance the accessibility of library collections. The study found that both datasets do not follow the same vocabulary. Though, the Spearman's rank correlation (0.89) indicates a good association between common terms in both vocabularies. The Jaccard similarity coefficient ($J = 0.13, 0.14, 0.17, 0.15$ and 0.16) in different word clusters proves that top frequent social tags and top frequent LCSH descriptors used by users and experts are different. The comparison with each book also reveals that 555 books (55.5%) have 50 to 100 percent matching between both vocabularies. LCSH descriptor vocabulary contains more subject terms (24) than social tag vocabulary (12) out of the top thirty frequent terms. The comparison of social tags with MARC subfields (\$a, \$x, \$y, \$z, \$v) reveals that users use more or less all the subfield terms as tags but either they do not use chronological terms (\$y) for tags or use different terms other than experts for chronological information. Further, comparison with each book title reveals that social tags alongside LCSH descriptors can enhance the title-based search of libraries. Moreover, the study suggests that usage of social tags will not only enhance the accessibilities of library resources under sociology but also complement to controlled vocabularies by supplementing a variety of terms other than experts.

Keywords: Social tags; Social tagging; User-generated social tags; Expert-generated LCSH descriptors; LibraryThing

Introduction

The subject metadata provides access to information resources in both traditional as well as digital library systems. Subject metadata defines the subject matter of any information resources. Subject metadata is derived from the concept of subject cataloguing which is carried out to keep all the library materials on a given subject under one uniform word or group of words.¹⁻² To achieve the best retrieval, adequate subject metadata should be assigned unless the inappropriate assignment of subject metadata yields to retrieval failure which determines the system is unable to retrieve relevant documents.³ Since past, there is centralization in the metadata generation in traditional library systems such as libraries, archives and museums.

Generally, librarians or cataloguers or trained library professionals assign subject metadata for the library materials in the form of subject headings. They assign subject headings using vocabulary control devices like Library of Congress Subject Headings

(LCSH), Sears List of Subject Headings (SLSH), different subject glossaries and thesaurus after analysing the content of library materials. The subject headings consisting of word or group of words were assigned with the tune that can act as an access point to ensure the searching and retrieving of resources by its subject.⁴

Social tagging and folksonomy

With the rise of Web 2.0 applications, the concept of folksonomy or social tagging has emerged. Social tagging is also known as 'social bookmarking', 'user tagging' and 'collaborative tagging' etc.⁵⁻⁷ Rolla defines 'social tagging' as a process where users can provide their keywords for different web resources like website, images and other contents.⁸ Social tagging is derived from the concept of 'folksonomy' which is coined by Thomas Vander Wal in 2005. Vander Wal describes folksonomy, a combination of two terms (a) folk which means people (b) taxonomy means a system of classification. He also defines that

folksonomy is a personal free tagging of digital resources on social networking sites for the retrieval of one's own information.⁹

Social tagging allows users to describe any web resources using keywords in the form of tags and without following any standard vocabulary. But in libraries, librarians or cataloguers assign subject terms following vocabulary control devices. Moreover, the assigned terms in the folksonomy can be updated at any time when the vocabulary changes and the user need changes. Not only resource description, folksonomies can navigate and retrieve those resources on the web in future also. Folksonomies can be treated as effectively as other information retrieval systems like search engines and directories. Folksonomies can improve web search capacity.¹⁰ Even the social tags help users to navigate the search direction by suggesting more search terms. It helps users to pay attention to the metadata or subject headings which might be fit with the search or might be familiar to the users. Folksonomies provide more user-oriented documents compared to the traditional bag-of-words model.¹¹

Social tagging applications

Due to the many advantages of social tagging, many libraries started to implement social tagging application in their libraries like the University of Pennsylvania designed *PennTags*, the University of Michigan also applied a tool *Mtagger*. Besides, Danbury Public Library and San Francisco State University also implemented LibraryThing for Libraries (LTFL).¹² Besides, many libraries redefined their library catalogue from OPAC to SOPAC (social OPACs)⁶ like Ann Arbor District Library also applied SOPAC, social online public access catalogue. Darien Library, Connecticut also used SOPAC 2.0 (<http://www.darienlibrary.org/catalog>) which allows users to make comments, reviews, ratings and tags to catalogue records.¹³ Apart from implementation in libraries, many social cataloguing sites like Goodreads (www.goodreads.com), LibraryThing (www.librarything.com), Anobii (www.anobii.com), Litsy (www.litsy.com), Readgeek (www.readgeek.com), etc., are becoming popular among common users worldwide day by day. As a result, millions of users are getting engaged (2.5 million in LibraryThing and 90 million in Goodreads) with these social platforms and catalogue socially billions of books (145 million in LibraryThing and 2.6 billion in Goodreads).¹⁴⁻¹⁵

Barriers of social tagging

Despite many advantages, social tags suffer from many inherent issues. Those issues make a hindrance for its incorporation into libraries. The first issue is that like other uncontrolled vocabularies social tags suffer from the problem of ambiguity and polysemy. Second, social tags also suffer from synonymous issues like 'Mac', 'apple' or 'macintosh' to represent the one object, that is the computer. Third, folksonomy provides a flat namespace which creates basic level terminological variations such as one user may assign tags to a document using 'perl' and 'javascript' and another user may assign tags to the same document as 'programming'. Fourth, social tags suffer from lexical anomalies in the choice of tags, such as singular vs. plural, spelling variants, verb tenses etc.¹⁶ Fifth, being flat systems in nature, folksonomy provides no hierarchy of terms. Sixth, due to having uncontrolled terms, folksonomies face lack of recall and precision while retrieving information.^{2,7,17-18} Further, social tags suffer from personal tags such as 'wishlist', 'kindle', 'own', 'hardcover' etc. which are not useful for retrieving resources. Users only assign personal tags just to meet their own needs.⁷

Still having many barriers, many researchers and information scientists also suggest that implementation of social tags can enhance the library catalogues by supplementing the controlled vocabularies.¹⁹

Review of literature

Several studies on social tagging have been carried out by many information scientists over the years to measure its relevancy in libraries. Lops et al (2013) suggest that application of folksonomy tags provides a democracy among library users where they can share their ideas, concepts, thinkings with tags regardless of any biases and background.²⁰ West (2007) suggests that folksonomy provides users with opportunities where users can search common vocabularies rather than using standard terms defined by libraries.²¹ Spiteri (2007) and Bianco (2009) identifies that folksonomy has massive impact enabling users to describe library resources through another way of resource description using their own tags, though some tags are used for personal benefit.²²⁻²³ Bartley (2009) also suggest that libraries can adopt sizeable folksonomy and can apply it in their library OPACs that will benefit users.²⁴

Besides, some information scientists also compared user-generated social tags with expert-assigned

subject headings to measure its effectiveness in libraries. In a study, Adler (2009) investigated WorldCat subject headings with tags in LibraryThing database on transgender themes. The results indicated that frequently used tags and WorldCat subject headings are different.²⁵

Lawson (2009) compared LC subject headings with user-generated tags from Amazon and LibraryThing database for 31 different subject divisions. The study indicated that both Amazon and LibraryThing database contains subject tags that the librarians can use to enhance subject access.²⁶ Lu et al (2010) also conducted a comparative study to measure the similarities and dissimilarities between user-generated social tags from LibraryThing and expert-assigned subject headings from Library of Congress to measure the effectiveness of social tags in libraries. The study focuses though both datasets contain different vocabularies; social tags can be used to enhance the accessibility of library collections.²⁷

Heymann and Garcia-Molina (2009) also conducted a comparison between LibraryThing and Library of Congress subject headings based on 309,071 records. The study explored that 50% of LC subject headings are used as tags but the usage of those keywords by annotators is quite different.¹⁹

In another study, Spiteri (2006) wanted to measure the effectiveness of folksonomies in public library catalogues. The study explored that combination of folksonomies and controlled vocabularies will be more effective for developing client-based customizable features in library catalogues.²⁸

Lee and Schleyer (2012) also compared CiteULike tags with Medical Subject Heading (MeSH) terms for 231388 citations index in the MEDLINE database. The study found that the tags contained by CiteULike tags and MeSH terms contained by MEDLINE are different. The study also indicated that both social tagging and controlled indexing were done for different viewpoints.²⁹

Wu et al (2013) also compared social tagging with controlled vocabulary terms in the field of information science from English and Chinese sources. The study found that more terms were matched between both vocabularies. The study also found more similarity (Jaccard's coefficient) among frequently used keywords in both datasets.³⁰

Research questions

The present study was carried out based on some research questions. These are as follows:

RQ 1. Do social tags and LCSH descriptors follow the same vocabulary?

RQ 2. Do both social tags and LCSH descriptors are the same based on usage?

RQ 3. Can social tags enhance the subject access like LCSH descriptors?

RQ 4. Do social tags and LCSH descriptors enhance title-based search?

RQ 5. Can social tags complement the controlled vocabularies?

Methodology

The present study randomly selected one thousand books in the field of sociology. We collected 30292 user-generated social tags from the LibraryThing database (www.librarything.com) and 4508 LCSH descriptors from the Library of Congress Online Catalogue (<https://catalog.loc.gov/vwebv/searchBrowse>).

After removing duplicates, unique social tags and unique LCSH descriptors have become 4655 (average 4.66 per book) and 1132 (average 1.13 per book) respectively under the study. The present study was carried out based on those unique social tags and LCSH descriptors. The study used mainly seven parameters to carry out the comparison between both vocabularies. Though there are several social cataloguing websites like Anobii (www.anobii.com), Litsy (www.litsy.com), Readgeek (www.readgeek.com), and Goodreads (www.goodreads.com), we chose LibraryThing.

This is because the LibraryThing visualizes the accumulated tags assigned by users under a given title which is technically called as 'Tag cloud' (Fig. 1). In the Tag cloud, assigned tags are arranged in alphabetical order and contain different frequency which reveals the popularity of tags. Besides, the LibraryThing database has a vast collection of books from Amazon.com, British Library, Library of Congress and 4967 other sources also.

We collected a thousand books (1000) in the field of Sociology those had at least three tags (≥ 3) assigned by users which appear on Tag Cloud portion of each book and also catalogued by at least ten members (≥ 10) in the LibraryThing database. We collected those tags that have tag frequency at least two (≥ 2) or more than that. Each tag holds a frequency under Tag Cloud portion of each book in the LibraryThing database. The tag frequency indicates the number of times the tag is being used. The more tag frequency means the tag is used for more times to describe different resources.³¹ On the

other hand, the study ensures that those titles were collected which had at least one subject heading (≥ 1) present in the Library of Congress Online Catalogue under MARC field 650 (topical term). The LCSH descriptors were collected from the Library of Congress Online Catalogue (<https://catalog.loc.gov/vwebv/searchBrowse>). Bibliographic records are described with many fields and subfields in MARC format under a given title in the Library of Congress database (Fig. 2).

The study selects field 020 for ISBN and field 245 for Title Statement for basic information. In the MARC database, field 6XX contains many fields for representing subject-related information. The present study selects field 650 (Subject Added Entry - Topical Term) and field 651 (Subject Added Entry - Geographic Name) under the field 6XX (Subject Access Fields-General Information). Both fields (650 & 651) contain two indicators, the first indicator reflects the level of subject whereas the second indicator reflects thesaurus. The study ensures that for field 650 first indicator (level of subject) would be 0 to 2 and the second indicator (Thesaurus) will be 0 as Library of Congress Subject Headings. In the case of field 651, the first indicator is undefined (#) and the second indicator will be 0 as Library of Congress Subject Headings.

In traditional cataloguing system, subject headings generated by the experts are like a string of words (Slavery--United States—History) which is concatenated by a hyphen (-). But in MARC format, subject headings for each bibliographic record are divided by different subfields. Each subfield represents different subject facets which altogether form subject headings under a given title. Both field 650 and field 651 contain many subfields but the study selects only five subfields, these are as follows: \$a- Topical term or geographic name entry element; \$x- General subdivision; \$y- Chronological subdivision; \$z- Geographic subdivision; \$v- Form subdivision. Each subject heading appeared on that subfields were collected separately. While collecting subject headings from subfields, it was found that some headings appeared more than one subfield e.g., the heading 'united states' appeared in subfield \$a and subfield \$z. But the study selects the heading 'united states' for once. Though headings were collected separately from subfields but were accumulated and book-wise unique headings were selected for the study.²⁷

Analysis

Terminological matching

The study compared total unique social tags with total unique LCSH descriptors for the entire collection



Fig. 1 — LibraryThing Tag Cloud (<https://www.librarything.com/work/154064>)

650	_0	a Blacks z Brazil x History.
650	_0	a African Americans x Social conditions.
650	_0	a Slavery z Brazil x History.
650	_0	a Slavery z United States x History.
651	_0	a Brazil x Race relations.
651	_0	a United States x Race relations.

Fig. 2 — LCSH descriptor record in MARC format (<https://catalog.loc.gov/vwebv/staffView?searchId=3668&recPointer=0&recCount=25&bibId=3144193>)

to know how many terms are used by both social taggers and experts. Fig. 3 & Fig. 4 indicate 30292 social tags and 4508 LCSH descriptors used by users and experts overall for different books. Table 1 also reveals that total 4655 unique social tags and 1132 total unique LCSH descriptors were compared and 689 unique terms are common. That means 689 unique terms are used by both users and experts for defining books. Besides, the common terms comprise the least portion (14.80%) of total unique social tags and a major portion (60.87%) of LCSH descriptors. That means a major portion of social tags (85.20%) are not found in LCSH term vocabulary. Further, the common terms also comprise about 60% (60.87%) of total unique LCSH descriptors. That means there is a 60% chance of LCSH descriptors to used as social tags by users.

Rank correlation of common terms

We used Spearman’s rank correlation to measure the association between common social tags and LCSH descriptors. The study determined common social tags as X and common LCSH descriptors as Y. The Spearman’s rank correlation is based on the assumptions of different ranks to different individuals.

So, Correction Factor (CF) is essential in case of tied ranks where m = number of observations tied to a particular rank.

The following equations are used for the Spearman’s rank correlation where n = 689. As multiple observations were found tied to particular ranks, so the value of m became multiple for social tag vocabulary as well as for LCSH vocabulary under the study. The study calculates Correction Factor for social tags (CF_X) = 933905 and Correction Factor for LCSH descriptors (CF_Y) = 2880378 and also calculates corrected $\sum d^2 = 5821683$. Spearman’s rank correlation becomes 0.89 for the study. That denotes there is a strong association between common terms in both vocabularies.

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \dots (1)$$

	Total terminologies	Unique terminologies	Common terminologies	%
Social tags	30292	4655	689	14.8
LCSH descriptors	4508	1132		60.87

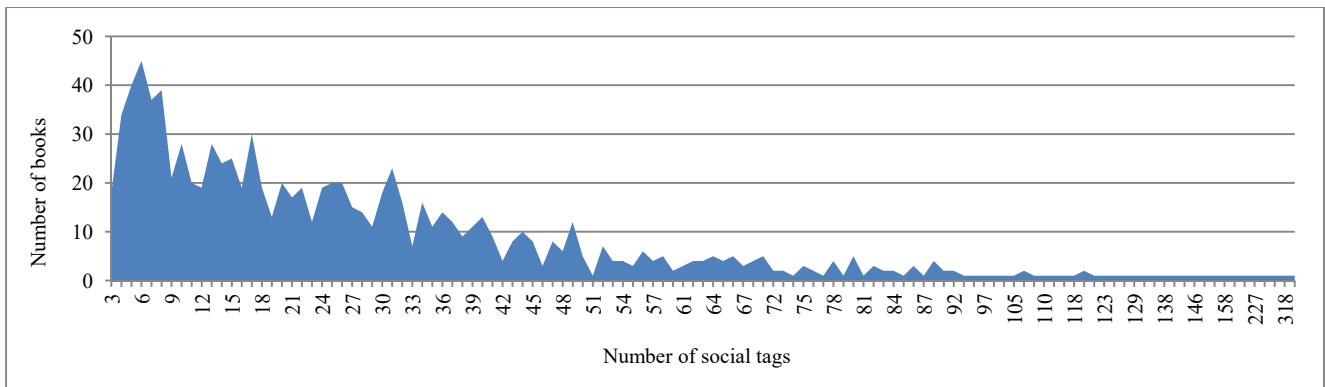


Fig. 3 — Number of social tags used by users for different sociology books

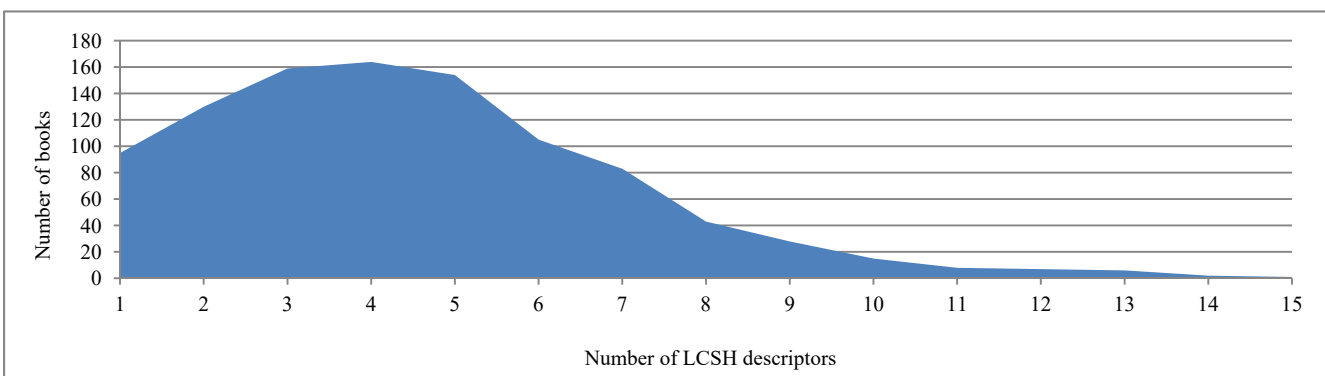


Fig. 4—Number of LCSH descriptors used by experts for different sociology books

Correction Factor for Social tags (CF_X)

$$CF_X = \frac{m(m^2-1)}{12} \quad \dots (2)$$

Correction Factor for LCSH descriptors (CF_Y)

$$CF_Y = \frac{m(m^2-1)}{12} \quad \dots (3)$$

Jaccard Similarity and Jaccard Distance based on different word clusters

We compared the top frequent social tags and the top frequent LCSH descriptors to measure the similarity and dissimilarity between them. The similarity and dissimilarity were measured based on the top five-word clusters like top 100, 200, 300, 400 and 500 terms. Jaccard similarity index and Jaccard distance were used to measure the similarity and dissimilarity. The following equation was used for the Jaccard index.³²

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad \dots (4)$$

The following equation was used for Jaccard distance.

$$1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \quad \dots (5)$$

[Where A= social tags & B = LCSH descriptors]

In the equation, n = number of top frequent terms, A = set of n frequent social tags and B = set of top n frequent LCSH descriptors. The study indicates that Jaccard index becomes 0.13, 0.14, 0.17, 0.15 and 0.16 respectively when n varies from 100 to 500. Besides, Jaccard distance becomes 0.87, 0.86, 0.83, 0.85 and 0.84 respectively. The study reveals that very low matching between top frequent social tags and top frequent LCSH descriptors which also suggests that top frequent terms used by both users and experts are different. Fig. 5 reveals Jaccard index and Jaccard distance comparatively when n varies from 100 to 500.

Top thirty frequently used social tags and LCSH descriptors

We also compared the top thirty frequent social tags with the top thirty frequent LCSH descriptors in both datasets. Table 2 elucidates the top thirty frequent terms and their corresponding frequencies in both datasets. In Table 2, it is found that only twelve social tags (12) appeared on LCSH descriptor vocabulary whereas twenty-six LCSH descriptors (26) appeared on social tag vocabulary out of the top thirty frequent terms. That means, the study reveals that social taggers use major LCSH descriptors as tag than experts use social tags as LCSH descriptors.

Further, we wanted to examine which dataset contains more subject-based terms out of the top thirty terms in both datasets. Subject-based terms mean the terms belong to particular this subject (Sociology) and non-subject terms belong to not particular this subject but allied subject. Table 2

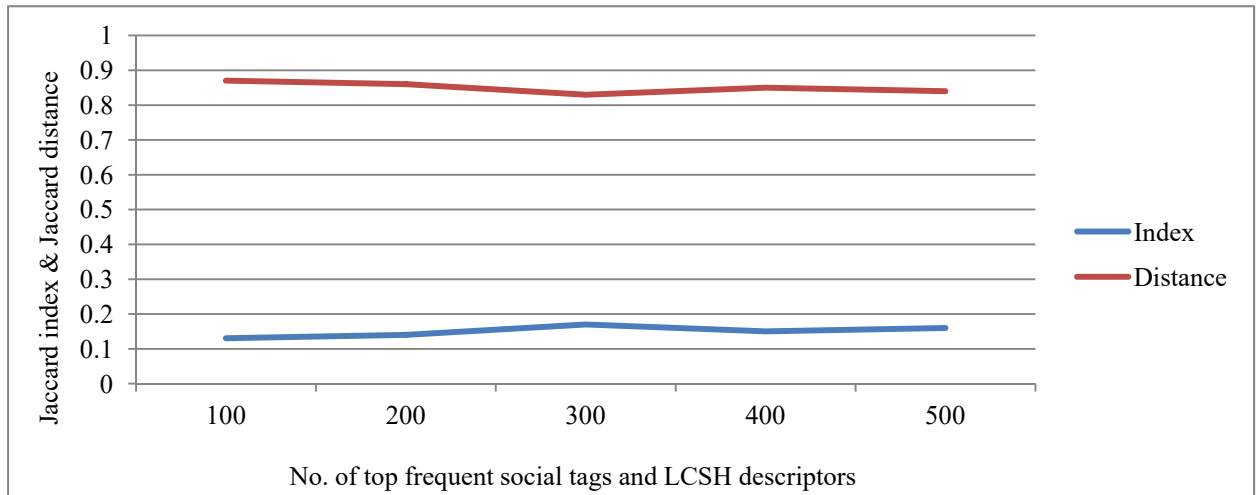


Fig. 5 — Jaccard Index and Jaccard Distance

Table 2 — Top thirty frequent Social tags and LCSH descriptors with reference to Sociology and its allied subject

Social tags	Freq. in tag vocabulary	Freq. in LCSH vocabulary	LCSH descriptors	Freq. in LCSH vocabulary	Freq. in Social tag vocabulary
non-fiction	863	-	united states	328	-
sociology	787	105	history	173	504
to-read	622	-	social conditions	164	73
history	504	173	sociology	105	787
politics	467	-	race relations	97	82
culture	441	23	social aspects	97	13
social science	432	3	african americans	56	52
unread	342	-	women	52	127
society	337	-	philosophy	41	334
philosophy	334	41	20th century	36	230
read	332	-	ethnic relations	36	9
usa	302	-	politics and government	36	9
cultural studies	292	-	economic conditions	33	8
anthropology	271	6	emigration and immigration	32	1
psychology	265	31	social change	32	46
economics	245	11	feminism	31	134
theory	232	-	psychology	31	265
20th century	230	36	racism	30	93
wishlist	214	-	social classes	28	19
political science	210	6	social history	28	132
social theory	207	-	ethnology	27	36
american history	177	-	methodology	27	35
race	164	7	social psychology	27	71
science	161	6	sex role	26	5
current affairs	158	-	social sciences	26	1
america	144	-	1980	24	-
gender	144	-	social life and customs	24	5
american	139	-	civilization	23	-
essays	138	-	culture	23	441
ebook	135	-	case studies	22	-

indicates that social tag vocabulary contains twelve subject-based terms (12), thirteen non-subject terms (13) and five personal terms (05) e.g., ‘to-read’, ‘read’, ‘unread’, ‘wishlist’, ‘ebook’, whereas LCSH descriptor vocabulary contains twenty-four subject-based terms (24) and six non-subject terms (06).

Individual book-wise comparison of social tags with LCSH descriptors

We compared social tags and LCSH descriptors for each book to measure the different levels of similarities between both vocabularies. The similarity levels were identified by comparing unique social tags with unique LCSH descriptors assigned by users and experts respectively for each book. It is found that in 881 out of thousand books where users and experts use at least one common term to annotate books. In 119 books, both users and experts use different

terminologies to annotate books. Besides, the study reveals different similarity levels between both vocabularies for each book.

Fig. 6 reveals that 202 books (20.2%) have 100 per cent similarity. That means in 202 books (20.2%), all expert-generated LCSH descriptors are used as tags by users. Further, the study reveals another level of similarities (0 to 90% similarities) between both vocabularies. It is also found that more than 555 out of thousand books that have fifty to hundred percent similarities between both vocabularies.

Social tags compared with LCSH subdivisions

We also compared social tags with LCSH subdivisions for each book. The study intended to measure which subfield terms the experts use mostly as LCSH descriptors and which subfield terms the

users use mostly as tags. Table 3 reveals that the experts use mostly \$a terms (958) and others are \$x (118), \$z (67), \$y (51) and \$v (17) respectively. On the other hand, Table 3 reveals that users mostly adapts \$z terms (71.64%) as tags, then use \$v terms (58.82%), \$a terms (56.68%), \$x terms (45.76%) and \$y terms (9.8%) respectively. That means the experts mostly prefer subfield \$a terms whereas the users mostly adapt subfield \$z terms but the users do not adapt \$y terms for tagging.

Further, the study assessed which types of terms the users use to tag major books. Table 4 reveals that subfield \$a and \$x contain major books (80% and 55.91% books respectively). That means subfield \$a and \$x are the highly used subfields. On the other hand, subfield \$z (22.39%), \$y (25.63%) and \$v (39.66%) contain least books. That means users prefer to use topical or geographic name entry elements terms (\$a) and topical subdivisions (\$x) for tagging but do not prefer or different terms other than experts to use geographic

subdivisions (\$z), chronological subdivisions (\$y) and form subdivisions (\$v). Fig. 7 also reveals subfield wise terms used by experts and users comparatively.²⁷

Table 3 — Total terms & unique terms appeared on LCSH subfields

	\$a	\$x	\$y	\$z	\$v
Total no. of terms appeared	2850	1146	214	438	65
No. of unique terms appeared	958	118	51	67	17
No. of terms used as tags	543	54	5	48	10
% of terms used as tags	56.68	45.76	9.8	71.64	58.82

Table 4 — No. of books which have at least one tag appeared in LCSH subfields

LCSH subfields	\$a	\$x	\$y	\$z	\$v
No. of books with this field	1000	626	160	393	58
%	100	62.6	16	39.3	5.8
No. of books which have at least one tag appeared in this field	800	350	41	88	23
%	80	55.91	25.63	22.39	39.66

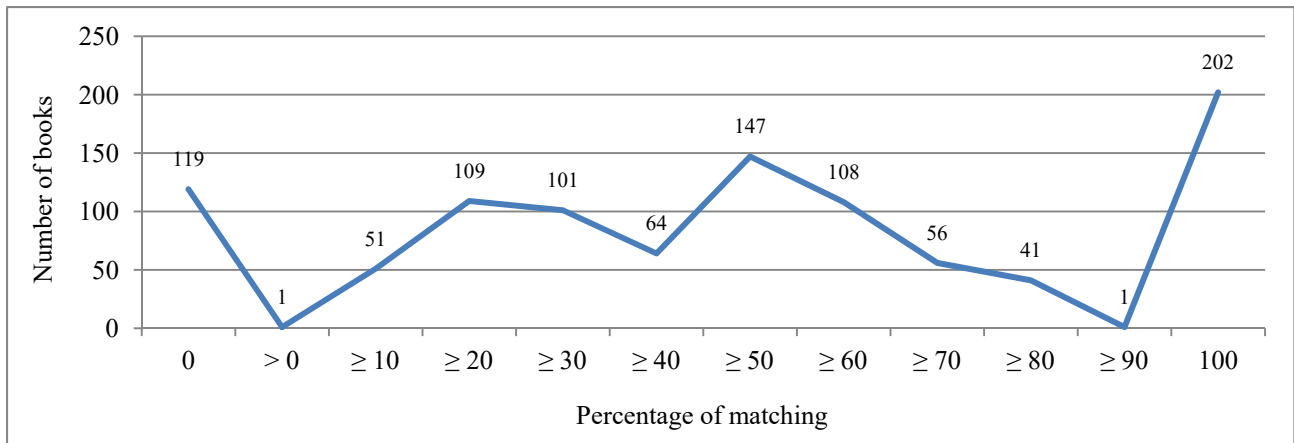


Fig. 6 — Individual book wise matching of social tags with LCSH descriptors

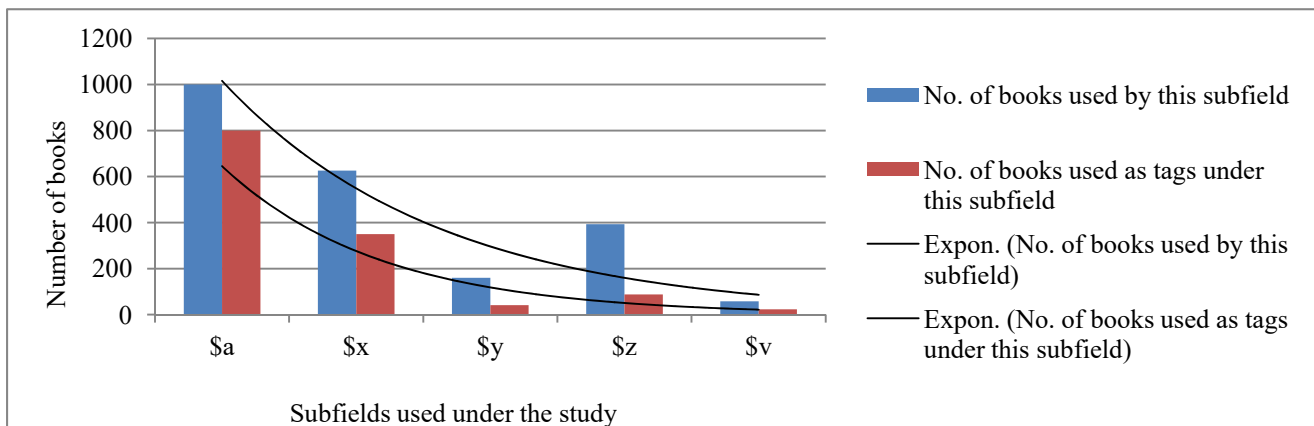


Fig. 7 — Usage of subfields by experts and users comparatively

Table 5 — Total terms, unique terms appeared on titles

	Total terms	Unique terms	%
Social tags	1284	564	12.12 of total unique social tags
LCSH descriptors	425	195	17.23 of total unique LCSH descriptors

Table 6 — Social tags & LCSH descriptors appeared on book titles

	No. of books where at least one tag appeared	No. of books where at least one LCSH descriptor appeared	No. of books where both appeared
Social tags & LCSH descriptors	779 (77.9%)	363 (36.3%)	318

Social tags and LCSH descriptors compared with each book title

Here, we tried to measure whether social tag and LCSH vocabulary contain terms beyond title terms or not. Table 5 reveals that 564 unique social tags (12.12% of total unique social tags) and 195 unique LCSH descriptors (17.23% of total unique LCSH descriptors) appeared on book titles. The study reveals that in both cases, the similarity level with title words is below 20%. That means both social tag and LCSH term vocabulary contains more than 80% terms beyond title words.

Further, the study indicates which vocabulary contains major books which have at least one tag that have appeared on book titles. Table 6 reveals that social tag vocabulary contains 779 books (77.9%) and LCSH descriptor vocabulary contains 363 books (36.3%). That means, for most books, users have taken terms from book titles. For a 363 books, experts use terms from book titles and the remaining 637 books (1000-363), experts use other terms than book titles.²⁷

Findings of the study

The vocabulary comparison reveals that very few terms matched with each other (60.87% of LCSH descriptors and 14.80% of social tags). Further, the study found that social tag vocabulary contains 85.2% terms (100-14.80%) other than expert-assigned LCSH descriptors. Though vocabulary difference exists, Spearman's rank correlation suggests that there is a good association between common terms in both vocabularies.

The Jaccard Index reveals a least matching (0.13, 0.14, 0.17, 0.15 and 0.16) between the top frequent

social tags and LCSH descriptors. It is revealed that users and experts do not use the same words to describe the same subject. The study found that more than fifty percent books (55.5%) books have 50 to 100 percent vocabulary matching and in major books (881) where users and experts use at least one common term.

The analysis of top thirty frequent terms reveals that users mostly use LCSH descriptors as tags but the experts use very fewer tags as LCSH descriptors. Further, the study found that LCSH descriptor vocabulary contains more subject-based terms (24) whereas social tag vocabulary contains less subject-based terms (12).

The comparison of social tags with LCSH subfields reveals that users mostly use geographic subdivision terms (\$z), and least use chronological terms (\$y) as tags. Further, the study found that users at least use one topical terms (\$a) for major books (800) and also use at least one general subdivision terms (\$x) for major books (350). Users seem to prefer topical terms, geographical terms and general subdivision terms for tagging but they do not prefer to use chronological terms for chronological information.

The comparison of social tags with each book title reveals that on one side, users use different terms (87.88%) beyond title-based terms for describing resources. Experts are found to use subject-based terms for describing resources in most books (603) out of 1000 books.

Conclusion

The present study reveals that social tag vocabulary is mostly different from LCSH vocabulary despite the worldwide popularity of social tags. This is because the intentions of users and experts are not the same to annotate the same books. Experts use standard vocabulary terms like LCSH descriptors to enhance the subject access of library collections and users use social tags for self-information retrieval. This indicates that users think social tags as the representation of the user's point of view and not as access points.

Despite the difference based on usage between both vocabularies, libraries can create a more user-centric environment that could lead to reaching more users by incorporating social tags. The Library of Congress Working Group on the Future of Bibliographic Control also recommends that implementation of social tagging will make the library

catalogue more user-centric and will enhance the access of library collections.³³ Further, if the libraries allow social tags, then users will be motivated and feel more engaged with their libraries. Fichter also believes that social tags are more popular and help users to connect to the library's website.³⁴ Spiteri also recommends that user-generated tags can supplement the controlled vocabularies.²⁸

To make the library catalogue more effective, libraries have to go through proper guidelines which tags should be included and which should not. In this context, the present study suggests that libraries are required to set up written guidelines before the incorporation of social tags. Spiteri recommends that if libraries want to incorporate folksonomies, there should be written recommendations for the choice and form of tags that can be used for libraries. Following those guidelines, libraries can properly implement folksonomies that can enhance the user-friendliness and interactivity of library catalogues.²²

Rolla also recommends that if libraries allow the use of folksonomy tags then they must have to develop a proper guideline regarding how to deal with the inherent problems of folksonomies like problems of polysemy, synonyms and lack of hierarchy etc. There also must be guidelines about how to deal with unhelpful tags.⁸ Further, the present study also suggests that the problem of synonyms and polysemy can be resolved if libraries use automatic indexing software for filtering tags. Now, the point is that libraries have to think about when and how the operations will be performed.

The present study reveals that user-generated tags alone cannot provide the subject access of library collection but a combination of both social tags and LCSH descriptors can provide users with the best subject access to library materials. That means using social tags alongside LCSH descriptors, libraries can improve their catalogues by supplementing terms other than experts.²⁶ Usage of both social tags and LCSH descriptors will enable libraries to describe the library materials from different perspectives and strengthen library catalogues to meet diversified search requests. In this way, using social tags libraries can enhance the accessibility of library collections and make libraries more reliable to users.

References

- 1 Safari M, Metadata and the Web, *Webology*, 1(2) (2004). http://eprints.rclis.org/7318/1/Metadata_and_the_Web.pdf (Accessed on 25 May 2020)
- 2 Macgregor G and McCulloch E, Collaborative tagging as a knowledge organisation and resource discovery tool, *Library Review*, 55(5) (2006) 291-300.
- 3 Sauperl A and Saye J, Subject determination during the cataloging process: An intensive study of five catalogers, *In Proceedings of the 9th ASIS SIG/CR Classification Research Workshop*, Pennsylvania, 25-29 October 1998, P. 119-138. <https://journals.lib.washington.edu/index.php/acro/article/viewFile/12748/11246> (Accessed on 23 May 2020)
- 4 Smith-Yoshimura, K, Social Metadata for Libraries, Archives, and Museums: Executive Summary, *A publication of OCLC Research*, Dublin, Ohio, 2012, <http://www.oclc.org/research/publications/library/2012/2012-02.pdf> (Accessed on 28 May 2020)
- 5 Hammond Tony et al., Social bookmarking tools (I): A general review, *D-Lib Magazine*, 11(4) (2005). <http://www.dlib.org/dlib/april05/hammond/04hammond.html> (Accessed on 24 May 2020)
- 6 Furner J, User tagging of library resources: toward a framework for system evaluation, *in the proceedings of the World Library and Information Congress: 73rd IFLA General Conference and Council*, Durban, South Africa, 19-23 August 2007. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.5202&rep=rep1&type=pdf> (Accessed on 22 May 2020)
- 7 Golder S A and Bernardo A H, Usage patterns of collaborative tagging systems, *Journal of Information Science*, 32(2) (2006) 198-208.
- 8 Rolla P J, User tags versus subject headings, *Library Resources & Technical Services*, 53(3) (2009) 174-184.
- 9 Wall, T V, Folksonomy. <http://vanderwal.net/> (Accessed on 20 May 2020)
- 10 Morrison P J, Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web, *Information Processing & Management*, 44(4) (2008) 1562-1579.
- 11 Kipp M E I and Campbell D G, Searching with tags: Do tags help users find things?, *Knowledge Organization*, 37(4) (2010) 239-255.
- 12 Wenzler J, LibraryThing and the library catalog: adding collective intelligence to the OPAC, *In Proceedings of the paper presented at the Workshop on Next Generation Libraries*, California Academic Research Libraries, North Information Technology Interest Group, San Francisco, CA, 7 September 2007. <http://www.carl-acrl.org/ig/carlitn/9.07.2007/LTFL.pdf> (Accessed on 24 May 2020)
- 13 Mendes L H, Quiñonez-Skinner J and Skaggs D, Subjecting the catalog to tagging, *Library Hi Tech*, 27(1) (2009) 30-41.
- 14 Goodreads. <https://www.goodreads.com/about/us> (Accessed on 26 May 2020)
- 15 LibraryThing. <https://www.librarything.com/zeitgeist> (Accessed on 28 May 2020)
- 16 Spiteri L F, The structure and form of folksonomy tags: the road to the public library catalog, *Information Technology and Libraries*, 26(3) (2007) 13-25.
- 17 Mathes A, Folksonomies – cooperative classification and communication through shared metadata, *Computer Mediated Communication-LIS590CMC*, Urbana, FL, 2004. <http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf> (Accessed on 10 October 2020)

- 18 Steele T D, The new cooperative cataloging, *Library Hi Tech*, 27(1) (2009) 68-77.
- 19 Heymann P and Garcia-Molina H, Contrasting Controlled Vocabulary and Tagging: Do Experts Choose the Right Names to Label the Wrong Things?, *In Proceedings of the paper presented at the Second ACM International Conference on Web Search and Data Mining (WSDM 2009)*, Barcelona, Spain, 9-13 February 2009. <http://ilpubs.stanford.edu:8090/955/1/cvuv-lbrp.pdf> (Accessed on 29 May 2020)
- 20 Lops P, Gemmis M, Semeraro G, Musto C and Narducci F, Content-based and collaborative techniques for tag recommendation: an empirical evaluation, *Journal of Intelligent Information Systems*, 40(1) (2013) 41-61.
- 21 West J, Subject Headings 2.0: Folksonomies and Tags, *Library Media Connection*, 25(7) (2007) 58-59.
- 22 Spiteri L, Structure and form of folksonomy tags: The road to the public library catalogue, *Webology*, 4(2) (2007). Available: <http://www.webology.org/data-cms/articles/20200515042609pma41.pdf> (Accessed on 15 May 2020)
- 23 Bianco C E, Medical librarians' uses and perceptions of social tagging, *Journal of the Medical Library Association: JMLA*, 97(2) (2009) 136.
- 24 Bartley P, Book tagging on LibraryThing: how, why, and what are in the tags? *Proceedings of the American Society for Information Science and Technology*, 46(1) (2009) 1-22.
- 25 Adler M, Transcending library catalogs: A comparative study of controlled terms in Library of Congress Subject Headings and user-generated tags in LibraryThing for transgender books, *Journal of Web Librarianship*, 3(4) (2009) 309-331.
- 26 Lawson K G, Mining social tagging data for enhanced subject access for readers and researchers, *The Journal of Academic Librarianship*, 35(6) (2009) 574-582.
- 27 Lu C, Park J R and Hu X, User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings, *Journal of information science*, 36(6) (2010) 763-779.
- 28 Spiteri L F, The use of folksonomies in public library catalogues, *The Serials Librarian*, 51(2) (2006) 75-89.
- 29 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.
- 30 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.
- 31 Samanta K S and Rath D S, Social Tags Versus LCSH Descriptors: A Comparative Metadata Analysis in the Field of Economics, *DESIDOC Journal of Library & Information Technology*, 39(4) (2019) 145-151.
- 32 Thomas M, Caudle D M and Schmitz C M, To tag or not to tag?, *Library Hi Tech*, 27(3) (2009) 411-434.
- 33 Amelung R and Latham J, Report of the Library of Congress Working Group on the Future of Bibliographic Control, USA, 2008. www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf (Accessed on 23 May 2020)
- 34 Fichter D, Intranet Applications for Tagging and Folksonomies, *Online*, 30(3) (2006) 43-46.