Qualitative and Quantitative Methods in Libraries (QQML) 15,1:57-91, 2026

Social Tags vs LCSH Descriptors: A Comparative Metadata Analysis of Tagore's Literature

Paromita Paul[1] & Ziaur Rahman[2]

Rabindra Bharati University[1]
Asian International School[2]

Abstract: Technological development has empowered civilization from all aspects. With the prevalence of web 2.0 applications, a new way of resource description has become popular called 'social tagging'. Social tagging has revolutionized how we describe and access web resources, offering numerous benefits over traditional cataloguing systems. This flexible and user-driven approach allows anyone to assign keywords to any resource, enabling quick adaptation to changing user needs. With social tagging, users can search for resources through tag recommendations, making it a cost-effective alternative that eliminates the need for professional librarians. The popularity of standalone social tagging applications, such as Library Thing, BibSonomy, Goodreads, Anobii, Flickr, and Readgeek, has made it accessible to a wide range of users, offering a new way to discover and share information. This study aims to evaluate the effectiveness and applicability of social tags compared to Library of Congress Subject Headings (LCSH) descriptors in Literature, specifically focusing on Rabindranath Tagore's works. By comparing the two approaches, this research seeks to determine whether social tags can be a viable alternative or complement to traditional LCSH descriptors, which are widely used in libraries for resource description.

Received: 25.8.2025 Accepted: 29.1.2026 ISSN 2241-1925

The study highlighted the advantage of social tags over LCSH (Library of Congress Subject Headings) in classifying Tagore's literature, showing that social tags achieved greater precision and recall in capturing the nuances and complexities of his works.

Keywords: Social Tags, LCSH Descriptors, Metadata, Rabindranath Tagore, Library Thing

Introduction

Web resources have become indispensable for education, research, and the sharing of cultural knowledge, including literary works. With the rapid growth of digital platforms and social networking sites, user-generated content has surged, reshaping how information is organized and accessed. In this context, metadata is crucial for maintaining the quality, reliability, and discoverability of digital content. By offering structured details about web resources, metadata enables users to evaluate the relevance and credibility of information—an essential function in today's age of misinformation and information overload.

Metadata is crucial for organizing, managing, and retrieving resources across digital platforms. Traditionally, libraries and information centers have relied on controlled vocabularies, such as the Library of Congress Subject Headings (LCSH), to create authoritative metadata for categorizing content. However, with the explosion of digital information, these expert-assigned metadata systems have shown limitations in handling the diverse and rapidly expanding range of digital content. Additionally, LCSH descriptors, while precise and reliable, can be time-consuming and costly to create, requiring specialized expertise. The rise of Web 2.0 has introduced a new paradigm: social tagging or "collaborative tagging." Social tagging enables users to create and assign descriptive metadata to digital content using free-form keywords. This bottom-up approach has gained traction on platforms like Goodreads, LibraryThing, and other community-driven sites, where users categorize content based on their perspectives and interests. Unlike LCSH descriptors, social tags are generated in an uncontrolled environment, reflecting the spontaneous and personal language

of users. While this can improve discoverability for specific user groups, it also introduces challenges such as inconsistency, ambiguity, and redundancy.

In the realm of literature, these differing approaches to metadata generation are especially relevant when applied to the works of renowned authors like Rabindranath Tagore. Tagore's literature, deeply rooted in Indian culture and philosophy, invites rich interpretation and varied perspectives, making it an interesting subject for analyzing how social tags and LCSH descriptors represent his works. LCSH, with its authoritative framework, might emphasize standardized, historically significant themes in Tagore's works, while social tags could highlight more personal and contemporary interpretations from readers worldwide.

This article presents a comparative analysis of how Rabindranath Tagore's literature is categorized using social tags versus LCSH descriptors. By examining the strengths and limitations of each approach, we aim to explore how these metadata systems shape the accessibility and interpretation of literary works in the digital space. Furthermore, this analysis highlights how the coexistence of expert-assigned metadata and user-generated tags can offer a more dynamic, inclusive framework for organizing and accessing literature, ultimately enhancing the reader's experience.

Significance of the study

The rapid development of web and internet technologies has significantly impacted various sectors, including academia, leading to increased use of e-resources by teachers, scholars, and students. As a result, academic libraries have shifted from print-based to e-resource collections, necessitating new management strategies. Traditional methods of cataloguing, relying on expert-assigned metadata, are becoming insufficient due to their rigidity, authority, and time-consuming nature. With the advent of Web 2.0, social tagging has emerged as a flexible, user-driven alternative, offering potential benefits such as improved accessibility and user engagement. This study examines the

effectiveness of social tags compared to the traditional Library of Congress Subject Headings (LCSH), particularly in the context of Tagore's literature, to explore their viability as a supplement or replacement in academic libraries.

Statement of the problem

Traditional libraries have long relied on controlled vocabulary tools like classification schemes (e.g., DDC, LC), subject headings (e.g., Library of Congress Subject Headings, Sears List), thesauri (UNESCO thesaurus), and glossaries to organize resources and facilitate user access. Librarians and catalogers use these tools to assign subject-based metadata to describe resources.

However, these metadata systems are inadequate for describing web resources. Expert-assigned metadata are rigid, slow to update, and centralized, limiting their adaptability to new terms and contexts. Moreover, users need special knowledge to navigate these systems, and the metadata often fails to represent the users' perspectives.

With the growing popularity of web-based resources, there is a demand for a new, flexible, user-friendly metadata system. Current library metadata is insufficient for web resources, so an alternative approach is needed—one that is dynamic, affordable, less time-consuming, and better suited to web content. This study advocates for the use of user-generated social tags as a way to rethink how libraries catalogue and describe their resources.

Scope of the study

The main purpose of the study is to compare user-generated social tags with expert-generated LCSH descriptors to measure whether any similarities and dissimilarities exist among them. The comparison was conducted in a single phase, focusing on word-matching between the two types of metadata for Tagore’s literature.

In this phase, the study selects two distinct databases to collect two types of metadata: the Library Thing database (https://www.librarything.com)foruser-generatedsocialtags, and the Library of Congress Online Catalog (https://catalog.loc.gov) for expert-generated LCSH descriptors. Furthermore, the study includes Tagore's literature and seventy-eight books from each dataset for comparison.

Objectives of the study

This study aims to evaluate the effectiveness of user-generated social tags versus expert-generated LCSH descriptors in describing Tagore's literature in libraries. The study has established the following objectives:

1. To compare the vocabulary used by social tags and LCSH descriptors
2. To explore the percentage of overlapping terms in Library Thing and Library of Congress Subject Heading
3. To explore the top 20 most frequent social tags and LCSH descriptors
4. To compare social tags with LCSH subdivisions
5. To explore user preferences for tagging major books based on MARC subfields
6. To compare social tags with LCSH descriptors about the titles of each book
7. To investigate whether social tags and LCSH descriptors utilize keywords from book titles.

Review of the literature

The exploration of metadata systems has gained significant attention as digital content continues to proliferate. Recent studies have compared user-generated social tags with expert-assigned Library of Congress Subject Headings (LCSH) to assess their effectiveness and applicability in organizing and retrieving library resources.

Lu et al. conducted a study comparing social tags from the Library Thing website with LCSH terms from the Library of Congress Online Catalogue. Their analysis of 176,105 unique tags and 7,628 LCSH terms across 8,562 book records revealed that only 2.2% of social tags matched LCSH terms. The study highlighted a considerable divergence between user-generated and expert-assigned metadata, noting that while social taggers utilized about half of the LCSH terms, the most frequent tags and terms differed significantly. This suggests that integrating social tags into library databases could enhance subject access, although they cannot fully replace LCSH terms (Lu, Park, and Hu 763-779).

Thomas, Caudle, and Schmitz also conducted a quantitative analysis to determine whether social tags could complement expert-generated LCSH terms. By comparing data from Library Thing and the Library of Congress for ten books, the study found that social tags provided additional subject access beyond what was covered by LCSH terms, suggesting that social tags can enrich library databases (Thomas et al. 411-434).

Heymann and Garcia-Molina's research focused on comparing LCSH subject headings with social tags from Library Thing. Their study, which included 8,783 LCSH headings and 47,957 social tags, found that approximately 50% of LCSH terms matched Library Thing tags. The study noted differences in how experts and users apply terms, underscoring the varied use of vocabulary between controlled and uncontrolled environments (Heymann and Garcia-Molina 1-4).

Lawson's study compared LCSH terms with user-generated tags from Amazon.com and Library Thing for the same titles. The results indicated that social tagging serves as a viable alternative metadata system that could enhance library resource accessibility (Lawson 574-582).

Rolla's research also explored the relationship between social tags and LCSH terms, finding some overlap in the vocabularies used by users and experts. Although the study concluded that social tags could enhance subject access, it

also emphasized that social tags alone cannot replace LCSH terms in libraries (Rolla 174-184).

Yi and Chan examined the correlation between folksonomy tags and LCSH terms, discovering that two-thirds of the tags matched with LCSH terms. This study aimed to determine if standard terms could improve the retrieval effectiveness of collaborative tagging systems, finding significant overlap (Yi and Chan 872-900).

Bogers and Petras performed a comprehensive analysis of metadata elements, including social tags and controlled vocabulary terms, across over 2 million book records. Their findings indicated that social tags and controlled vocabularies provided comparable retrieval effectiveness, with certain contexts favoring one system over the other for describing information (Bogers and Petras).

Vaidya and Harinarayana compared LCSH terms with LibraryThing tags for library and information science books published between 2000 and 2015. They found that social tags offered effective retrieval, though they also acknowledged the ambiguity and limitations inherent in uncontrolled tagging systems (Vaidya and Harinarayana 35-43).

Adler's study of subject headings from the WorldCat database and user-assigned tags from LibraryThing for transgender materials revealed that while social tags captured diverse representations and terms of gender expressions absent in WorldCat records, both systems had limitations. Adler recommended both types of metadata to enhance catalogue efficiency (Adler 309-331).

Additional comparative studies by Bogers and Petras, Wu et al., Carman, and Samanta and Rath further explored the effectiveness of social tags versus controlled vocabularies in various contexts. These studies consistently highlighted that while social tags and expert-generated terms overlap to some extent, each system has unique strengths and limitations. Integrating both can potentially improve library resource accessibility without entirely replacing

LCSH terms (Bogers and Petras; Wu et al. 169-187; Carman; Samanta and Rath).

This literature review demonstrates that while both social tags and LCSH descriptors offer distinct advantages, no study has specifically explored the effectiveness of social tags in the context of Tagore's literature This research gap underscores the uniqueness of the current study, which aims to provide a comparative analysis of social tags and LCSH descriptors in describing Tagore's works.

Methodology

The present study aims to evaluate the application of social tagging by comparing user-generated social tags with Library of Congress Subject Headings (LCSH) terms in the context of Tagore's literature. The objective is to determine whether social tags can complement controlled vocabulary terms, such as LCSH descriptors, for bibliographic descriptions.

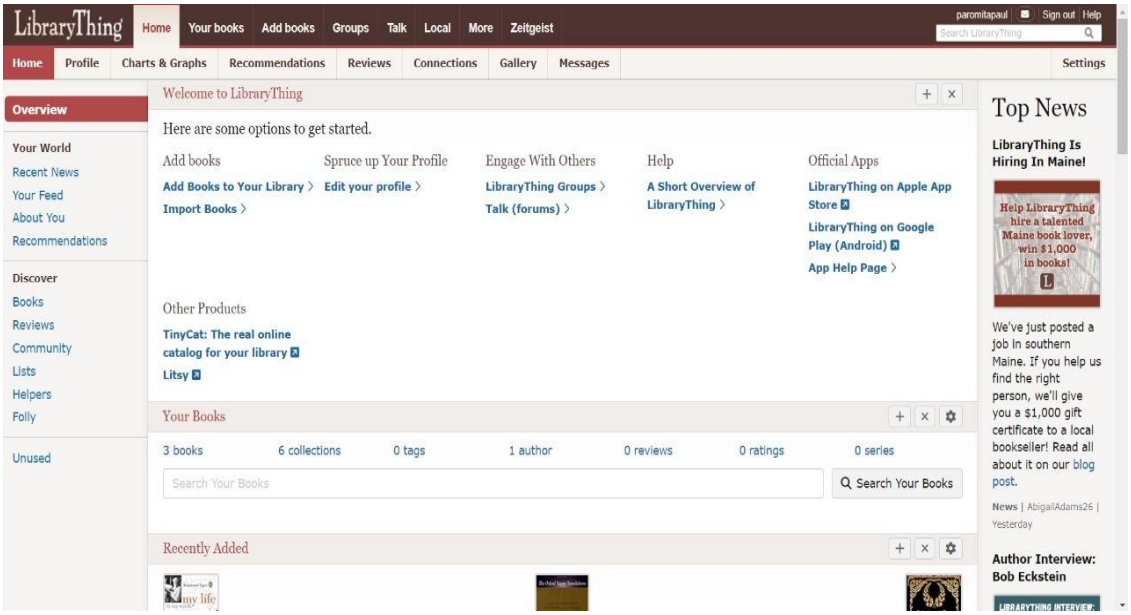
This study focuses on examining the relationship between user-generated social tags and expert-or librarian-generated standard vocabularies. Social tags are the core element of this analysis. The study begins by selecting literature as a broader discipline and then narrows the focus to Rabindranath Tagore's works. Tagore, a renowned Bengali polymath, made significant contributions as a poet, novelist, short story writer, philosopher, musician, and painter. His literary works are considered classic so Bengali literature, making them an ideal subject for this investigation.

The study utilizes two types of metadata: (1) social tags and (2) LCSH descriptors, to compare

Tagore's literature within the discipline of literature. To collect social tags, the study selects Library Thing (https://www.librarything.com/), a social cataloguing platform. For LCSH descriptors, the study uses the Library of

Congress Online Catalog (https://catalog.loc.gov). A total of 78 books from Tagore's literature are selected as the sample for this analysis.

Library Thing is a social cataloguing application designed to help individuals, libraries, and publishers organize knowledge. It allows users to quickly catalogue books and offers several core features: (a) a tag cloud (user-generated subjects), (b) links to other editions, and (c) recommendations for similar books. The tag cloud feature facilitates resource discovery by establishing relationships



between related works and enhancing the existing metadata. Library Thing was developed by Tim Spalding and launched in 2005.

Fig.01:Library Thing search interface

The Library of Congress Online Catalogue allows users (a) to browse bibliographic records by subject, authors, call numbers, titles etc., (b) to perform advanced search and (c) to perform a keyword search. Besides, the Library of

Congress Catalogue provides a Z39.50 interface through which any library can access and downloadbibliographicrecordsinMARC21format.(LibraryofCongress,Aboutthe LCCatalog)

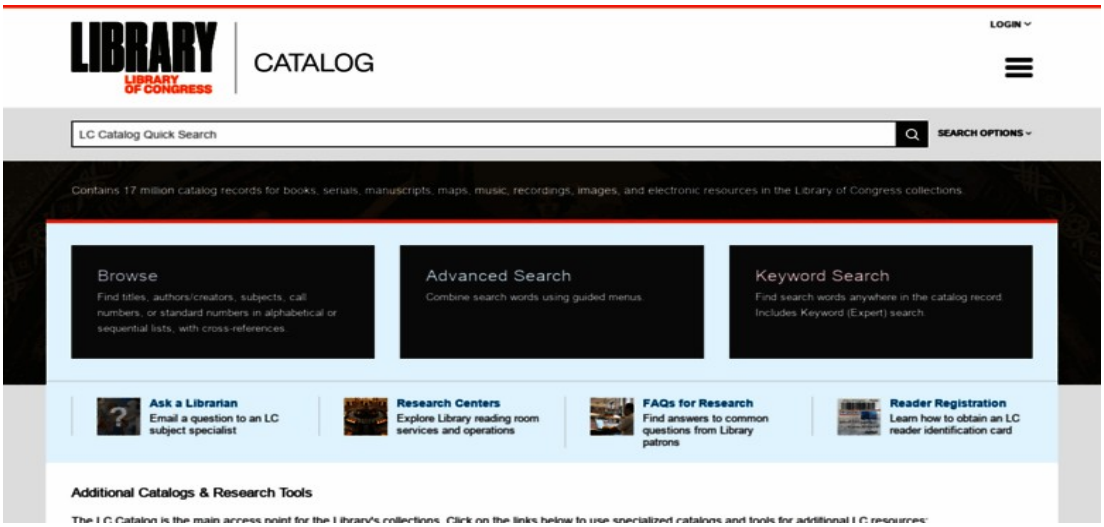
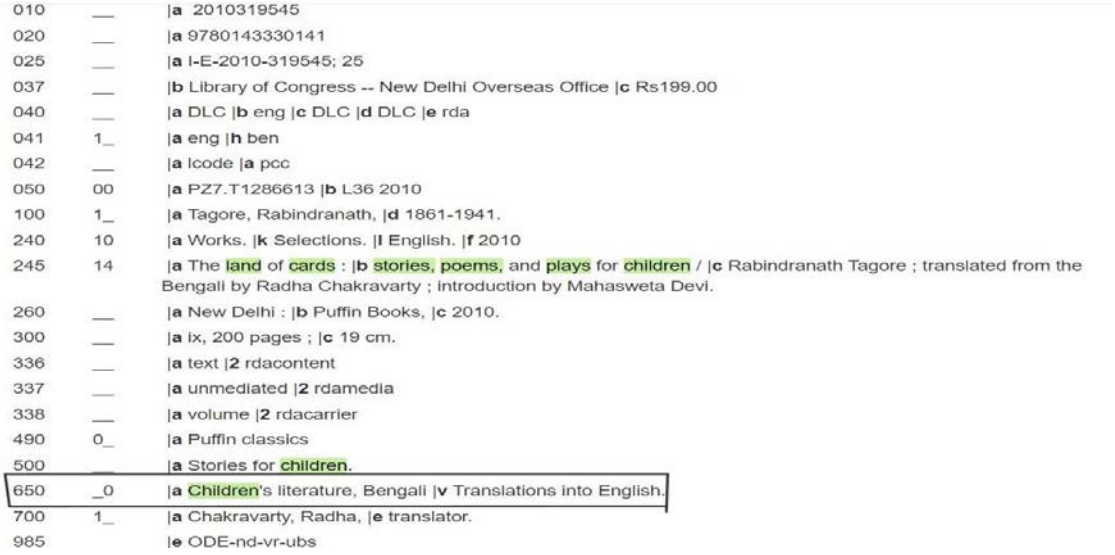


Fig.02:Library of Congress Online Catalogue Search Interface



010	___	|a 2010319545
020	___	|a 9780143330141
025	___	|a I-E-2010-319545; 25
037	___	|b Library of Congress -- New Delhi Overseas Office |c Rs199.00
040	___	|a DLC |b eng |c DLC |d DLC |e rda
041	1_	|a eng |h ben
042	___	|a lcode |a pcc
050	00	|a PZ7.T1286613 |b L36 2010
100	1_	|a Tagore, Rabindranath, |d 1861-1941.
240	10	|a Works. |k Selections. |l English. |f 2010
245	14	|a The land of cards : |b stories, poems, and plays for children / |c Rabindranath Tagore ; translated from the Bengali by Radha Chakravarty ; introduction by Mahasweta Devi.
260	___	|a New Delhi : |b Puffin Books, |c 2010.
300	___	|a ix, 200 pages ; |c 19 cm.
336	___	|a text |2 rdacontent
337	___	|a unmediated |2 rdamedia
338	___	|a volume |2 rdacarrier
490	0_	|a Puffin classics
500	___	|a Stories for children.
650	_0	|a Children's literature, Bengali |v Translations into English.
700	1_	|a Chakravarty, Radha, |e translator.
985		|e ODE-nd-vr-ubs

Fig.03: Library of Congress Subject headings appeared under subfield 650 under the title 'The Landof Cards' by Rabindranath Tagore

Parameters Used for Data Collection and Comparison

A set of parameters was developed for collecting books, social tags, and LCSH descriptors from the selected databases. Initially, the study focused on 142 books that were common between Library Thing and LCSH (Library of Congress Subject Headings) databases. During the analysis it was discovered that 17 books contained descriptors but no social tags,41books contained social tags but no descriptors and 18 books had collaborative or joint authors. As a result, the study selected books from the Library Thing data base that had at least one user-generated tag (≥1)visible in the tag display section of the books. Additionally, these books were required to be catalogued by at least one member (≥ 1) in the Library Thing database. The study also focused on collecting tags with a frequency of at least one (≥ 1) or higher. The frequency of tags is visualized in

the tag display section, reflecting how often a tag issued for a particular book (Samanta and Rath, 145-146).

For the collection of subject descriptors, the study accessed the Library of Congress Online Catalogue (https://catalog.loc.gov/vwebv/searchBrowse)for 78 books of Tagore's literary works. The LC MARC database has several fields under 6XX that represent subject-based information. The study selected books with at least one subject descriptor (≥ 1) found under MARC fields 600 (Subject Added Entry – Personal Name), 610 (Subject Added Entry - Corporate Name), 650 (Subject Added Entry – Topical Term), 651 (Subject Added Entry - Geographic Name), and 655 (Index Term - Genre/Form).The study began by selecting basic information such asMARCfield020forISBNand field245 for the title statement.

Each MARC field has specific subfields representing different subject facets. For instance:

MARC field 600: first indicator (Type of personal name entry element) and second indicator (Thesaurus)

MARC field 610: First indicator (Type of corporate name entry element) and second indicator (Thesaurus)

MARC field 650: first indicator (Level of subject) and second indicator (Thesaurus)

MARC field 651: first indicator (Undefined) and second indicator (Thesaurus)

MARC field 655: first indicator (Type of heading) and second indicator (Thesaurus)

The study specifies that underfield600, the first indicator will range from 0(zero) to1 (one), with the second indicator being 0 (zero).

For field 610, the first indicator will range from 0(zero) to2(two), with these cond indicator being 0 (zero).

For 650 field contains first indicator (Level of subject) will be 0 (zero) to 2 (two) and the second indicator (Thesaurus) will be 0 (zero).

For 651 field contains first indicator will be # (Undefined) and the second indicator (Thesaurus) will be 0 (zero).

For 655 field contains first indicator (Type of heading) will be # (Basic) to 0 (zero) and the second indicator (Thesaurus) will be 0 (zero) to represent the Library of Congress Subject Headings (LCSH).

The subject descriptors are identified by various sub fields such as:

$a –Topical term or geographic name entry element

$x – General subdivision

$y –Chronological subdivision

$z - Geographic subdivision

$v - Form subdivision

The study collected subject descriptors appearing in these subfields individually. During data collection, it was noted that some descriptors appeared in multiple fields or subfields. For example, the book *"The Weird: A Compendium of Strange and Dark Stories"(ISBN: 9781848876873, LCCN: 2012289436)* had the descriptor "short stories" appearing in both field 650 and field 655.

In other example, the *book "Leading from Within: Poetry that Sustains the Courage to Lead"(ISBN: 9780787988692, LCCN: 2007021352) had the descriptor "Poetry" appearing in both field650*and field 655.

To ensure consistency with social tags, the study retained scope notes that appeared with certain terms in the MARC data. Scope notes are used by experts to clarify the context in which a term is applied, and social taggers often include such context, e.g., "Craft Lessons(Fletcher)" or "Guided Reading J and K (White Tape)."

In the process of selecting terms from both datasets (Library of Congress and Library Thing), certain pairs of terms were treated as similar. For example, terms such as "poet" and "poets," "book" and "books," as well as "translation" and "translations," were considered equivalent for the study's analysis.

The data collection process took place from April 2024 to July 2024. After collecting the terms from both databases, the study removedduplicates and calculated the unique terms for each vocabulary. The unique terms were then organized by their frequency in both datasets for further evaluation.

008	970904s1928 nyu a 000 0 eng	
035	___	**|9** (DLC) 96911768
906	___	**|a** 7 **|b** cbc **|c** origode **|d** 5 **|e** ncip **|f** 19 **|g** y-genmicro
955	___	**|a** wd30;01/17/97
010	___	**|a** 96911768
025	___	**|a** NBIL-09572
040	___	**|a** DLC **|c** DLC
041	1_	**|a** eng **|h** ben
042	___	**|a** lcode
050	00	**|a** Microfilm PIM-ENG-508 (P)
100	1_	**|a** Tagore, Rabindranath, **|d** 1861-1941.
245	10	**|a** Fireflies **|h** [microform] / **|c** by Rabindranath Tagore ; decorations by Boris Artzybasheff.
260	___	**|a** New York : **|b** Macmillan Co., **|c** 1928.
300	___	**|a** 274 p. ; **|c** 19 cm.
500	___	**|a** Poem.
533	___	**|a** Microfilm. **|b** New Delhi : **|c** Library of Congress Office ; **|b** Chicago : **|c** Available from Center for Research Libraries, **|d** 1996. **|e** On 1 microfilm reel with other items ; 35 mm. **|f** (SAMP early 20th-century Indian books project ; item 09572) **|n** Master microform held by: ICRL.
650	_0	**|a** Aphorisms and apothegms.
650	_0	**|a** Proverbs, Indic.
650	_0	**|a** Maxims.
830	_0	**|a** SAMP early 20th-century Indian books project ; **|v** item 09572.

Fig. 04: A sample record entitled "Fireflies" by Rabindranath Tagore in MARC Format in Library of Congress Online Catalogue.

Fig. 05: A sample record entitled "Fireflies" by Rabindranath Tagore in the Library Thing database

Result and discussions

This study provides a comprehensive comparison between social tags and LCSH descriptors in the context of Tagore's literature. The analysis involved a detailed examination of seventy-eight literary titles, with data collected from Library Thing for social tags and the Library of Congress Online Catalog for LCSH descriptors. The comparison was conducted from various perspectives to achieve the studies objectives and gain valuable insights.

Overlap of Terminology in Tagore's Literary Works:

The study explored the overlap between the total number of unique social tags and LCSH descriptors to assess any similarities between the terms used by users (social tags) and experts (LCSH descriptors). As shown in Table 1, the collection includes 4,994 unique social tags and 387uniqueLCSH descriptors. Figures1 and 2 illustrate the distribution of social tags and LCSH descriptors assigned by users and experts, respectively, for each book.

Out of 3,658 unique social tags compared to 98 unique LCSH descriptors, 30 terms overlap. This overlap represents just 0.08% of the total unique social tags indicating that 99.92% of the social tags are not present in the LCSH vocabulary. On the other hand these overlapping terms make up 30.61% of the LCSH descriptors, suggesting assign if I can't potential for using LCSH descriptors asocial tags by users.

Table1: Total terms, Uniqueterms, and Overlapping Terms

	Total terms	Unique terms	Overlapping terms	Percentage
Social tags	4994	3658		0.08
			30	
LCSH descriptors	387	98		30.61

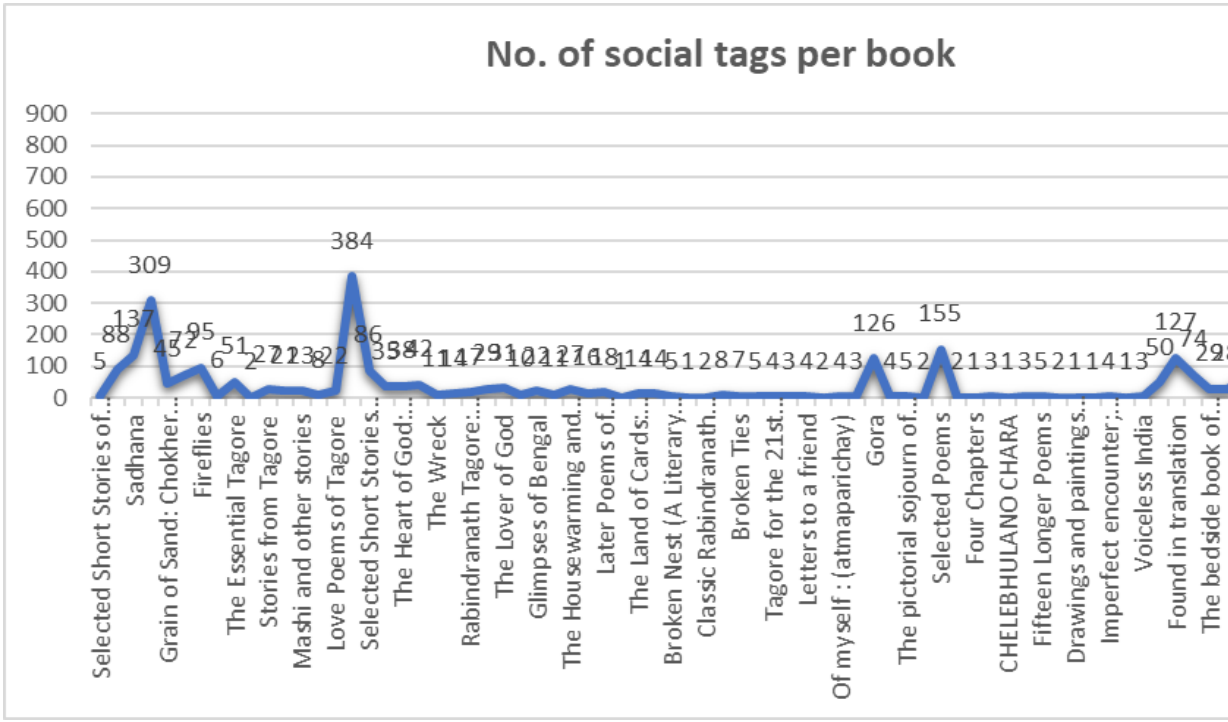


Fig.06: Social tags assigned by users per book inTagore's literature

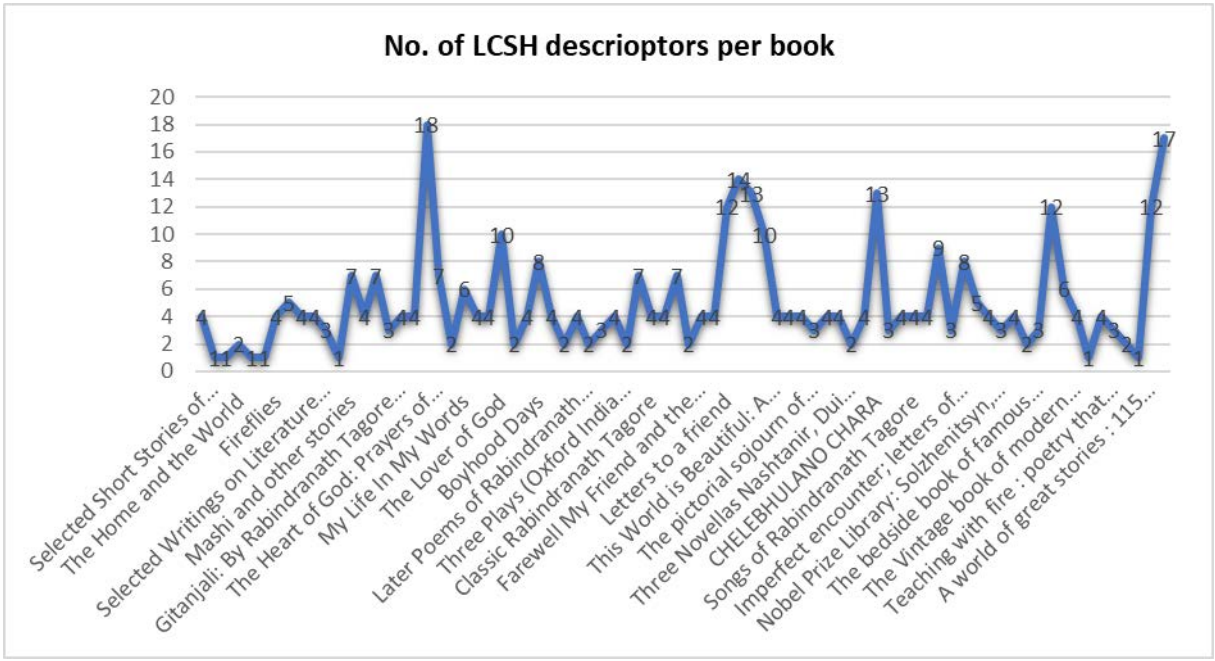


Fig: 07: LCSH descriptors assigned by experts per book in Tagore's literature

Top Twenty Most Frequent Social Tags and LCSH Descriptors in Tagore's Literature:

The study compares the top twenty most frequent social tags with the top twenty most frequent LCSH (Library of Congress Subject Headings) descriptors from both vocabularies. Table 2 presents these top twenty social tags and LCSH descriptors along with their corresponding frequencies across the two datasets.

The analysis reveals that out of the top twenty social tags, fifteen (15) are also found in the LCSH term vocabulary, while thirteen (13) of the LCSH descriptors overlap with the social tag vocabulary. This indicates a significant level of similarity in the terms predominantly used by users and experts.

Additionally, the study explores which vocabulary contains a higher proportion of subject-based terms. Subject-based terms refer to specific subjects like

'Literature,' while non-subject terms relate to broader or allied concepts such as 'Literary collections.' In the top twenty terms, the social tag vocabulary includes twelve subject-based terms, five non-subject terms, and three personal tags (e.g., 'to-read,' 'Kindle,' 'classics'). In contrast, the LCSH vocabulary consists of eleven subject-based terms and nine non-subject terms.

Moreover, the study identifies twelve common terms used in both datasets. These shared terms include 'Tagore,''Rabindranath,''India,''Bengali,''Fiction,''Poetry,''ShortStories,' '20th Century,' 'Biography,' 'Religion,' 'Collection,' and 'Nobel Prizes.' Out of these twelve terms, nine are categorized as subject-based terms.

The frequency analysis shows that the term 'Literature' appears 45 times in the social tag vocabulary but only once in the LCSH term vocabulary ,indicating that users use 'Literature' for 45 books while experts use it for just one. Similarly, 'Religion' appears for 15 books in social tags and 3 books in LCSH, while' Bengali'appears for 14 books in social tags and 13 books in LCSH. This data suggests that users are more likely to use terms related to common books compared to experts. (Rolla179-180).

This analysis specifically focuses on the top twenty terms with the highest frequencies from each dataset to draw comparisons.

Table2: Top twenty frequent social tags and LCSH descriptors with special reference to Literature and its allied subject

Social tags	Frequency in social vocabulary	Frequency in LCSH vocabulary	LCSH terms	Frequency in LCSH vocabulary	Frequency in social vocabulary
India	45	16	Tagore	50	21
Literature	45	1	Rabindranath	49	24

Poetry	34	9	1861-1941	49	0
to-read	32	0	Translation into English	32	0
Indianliterature	31	0	India	16	45
Fiction	29	10	Bengali	13	14
RabindranathTagore	24	49	Correspondence	13	0
Tagore	21	50	Fiction	10	29
Indian	20	0	Poetry	9	34
Nobel Prize	20	2	Short stories	7	14
anthology	17	0	Short stories	7	1
20th century	17	7	20th century	7	16
religion	15	3	Biography	4	13
Short stories	14	0	Literary collection	4	2
Bengali	14	7	Social life and customs	3	0
Kindle	14	13	Vedanta	2	1
philosophy	13	0	fantasy fiction	3	1
Collection	14	1	Collection	3	14
Biography	13	3	Nobel prizes	2	20
Classics	12	1	19th century	2	9

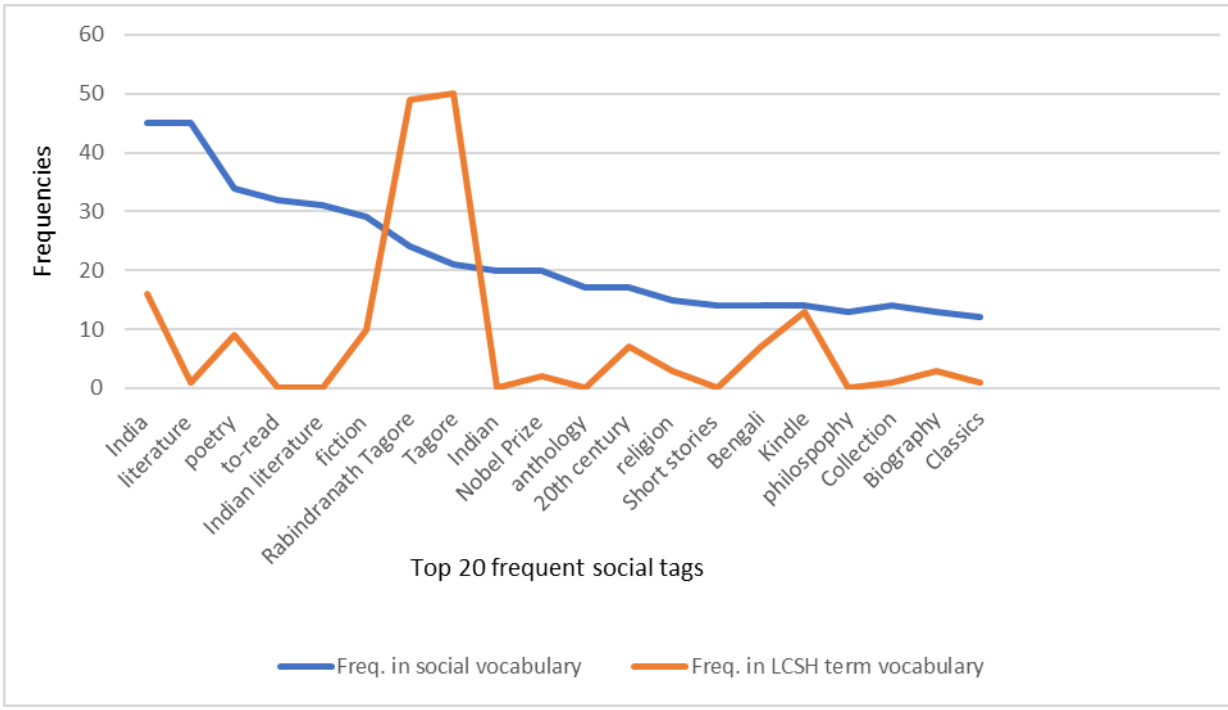


Fig. 08: Top 20 frequent social tags with their frequency in Tag and LCSH vocabulary

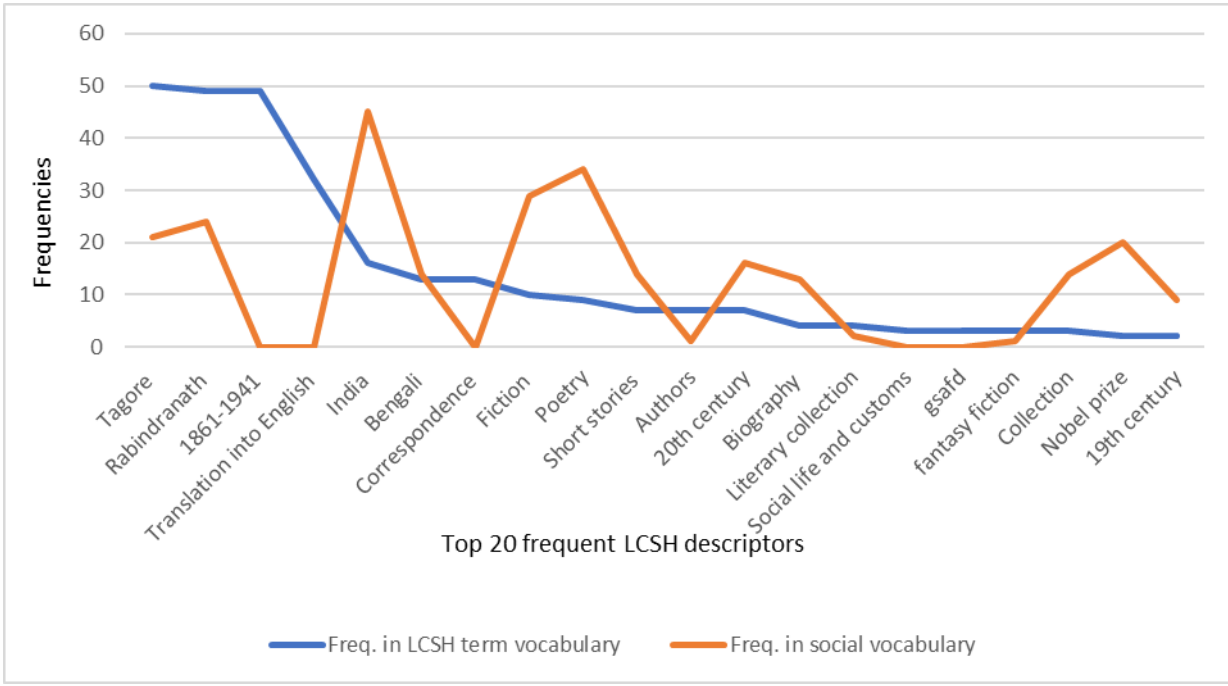


Fig. 09: Top 20 frequent LCSH descriptors with their frequency in LCSH and Tag vocabulary

Social Tags Versus LCSH Subdivision sin Tagore's Literature:

This study compares social tags with MARC (Machine-Readable Cataloging) subfields to analyze the use of terms such as LCSH (Library of Congress Subject Headings) descriptors by experts versus the tags created by users. The focus is on determining which MARC subfield terms are most frequently used by experts and which are most commonly used by users tags. The analysis considers five MARC subfields across five fields (600, 610, 650, 651, 655): $a(topicalorgeographicnameentryelement),$x(topicalsubdivision),$y(chronologicalsubdivision),$z(geographic subdivision), and $v (form division).

Table 3 illustrates that the $a subfield is the most frequently used by experts, containing 61unique terms. The distribution of unique terms across the other subfields is as follows: $x has 23 unique terms, $yhas3unique terms, $z has 2

unique terms, and $v has 12 unique terms. In total, experts use 101 unique terms across these subfields, though only 67 unique terms are used for evaluation, indicating that some terms appear in multiple subfields. For example, the term 'India' appears in $a as a topical or geographic name, in $z as a geographic subdivision, and $v as a form division.

The study also reveals that $a terms are most frequently used as tags, accounting for 29.50%of the tags. The usage of terms from other subfields as tags is as follows: $x (34.78%), $z(100%),$y(66.60%), and $v(41.60%). However, percentage-based comparisons are somewhat misleading due to the varying number of total terms across subfields. For instance, although $a has the lowest percentage (29.50%), it encompasses the highest number of totalterms(202),uniqueterms(61),andtags(18).Incontrast,$z,whichholdsthehighestpercentage(100%),has the lowest number of total terms (4),unique terms(2),and tags(2).

The detailed breakdown is as follows:

$x: 57 total terms, 23unique terms, 8 terms used as tags, 34.78%of tags.

$y: 13 total terms, 3 uniqueterms,2 terms used as tags,66.60%oftags.

$v: 30 total terms, 12unique terms, 5 terms used as tags, 41.60%of tags.

Rather than relying solely on percentage calculations, it is more accurate to assess tag usage based on the absolute number of terms used as tags from each subfield. This approach shows that $a (topical or geographic name entry element) and $x (topical subdivision) are the mostfrequentlyusedsubfieldsfortags,while$v(formdivision),$z(geographicsubdivision),and

$y(chronological subdivision)is less commonly used.

Table 4 highlights the top 10 LCSH subfield terms used as tags by social taggers, revealing that all of these terms come from the $a subfield. Other subfields contain fewer than 10 terms used as tags ($x– 8,$y–2,$z–5,$v–2). This table is each sub field term that was use data tag.

Table3: Total terms, unique terms, and terms used as tags appear e don LCSH subfields:

	$a	$x	$y	$z	$v
Total no.of terms appeared	202	57	13	4	30
Several unique terms appeared	61	23	3	2	12
Number of terms used as tags	18	8	2	2	5
%ofter ms used as tags	29.50%	34.78%	66.60%	100%	41.60%

Table4: Top 10 LCSH subfield terms used as tags:

$a	Freq	$x	Freq	$y	Freq	$z	Freq	$v	Freq
India	45	Poetry	34	20th century	17	India	45	India	45
Poetry	34	Fiction	29	19th century	8	Bengal	36	Bengal	36
Rabindranath	24	Religion	15			Poetry	34		
Tagore	21	Biography	13			Fiction	29		
Literature	20	Philosophy	13			Literary collection	1		
Nobel prizes	20	History	12						

Religion	15	Literary collections	1						
Bengali	14	Politics and Government	1						
Short stories	14								
Collections	12								

User Preferences for Tagging Major Books: An Analysis of MARC Subfields

The study further explores which types of terms users prefer for tag gingham jor books. Table 5 shows that subfield $a is used for tagging 76 out of 78 books, making it the most frequently used subfield. The distribution of subfield usage for tagging is as follows: $x is used for 28books, $v for 33 books, $y for 8 books, and $z for 2 books. This indicates that $a is the most popular subfield for users, while $z is the least utilized.

Additionally, the study investigates which subfields contain at least one tag. Table 5 reveals that subfield$a has the highest number of books (41) with at least one tag assigned ,whereas $xhas only 1 book with a tag, and $v has 2 books with a tag. Subfields$y and $z did not have any tags assigned to books.

These findings suggest that users prefer to use terms from the topic a lor geographic name entry element ($a), topical subdivision($x), and form division($v) when tagging books. Conversely, users do not tend to use terms from the geographic subdivision($z) or chronological subdivision($y).

Table5: No. Of book titles that has at least one tag appeared in LCSH subfields

LCSH Subfields	$a	$x	$y	$z	$v
No. of records with this field	76	28	2	8	33
No. of records that have at least one tag appearing in this field	41	1	0	0	2
%	53.94	3.57	0	0	6

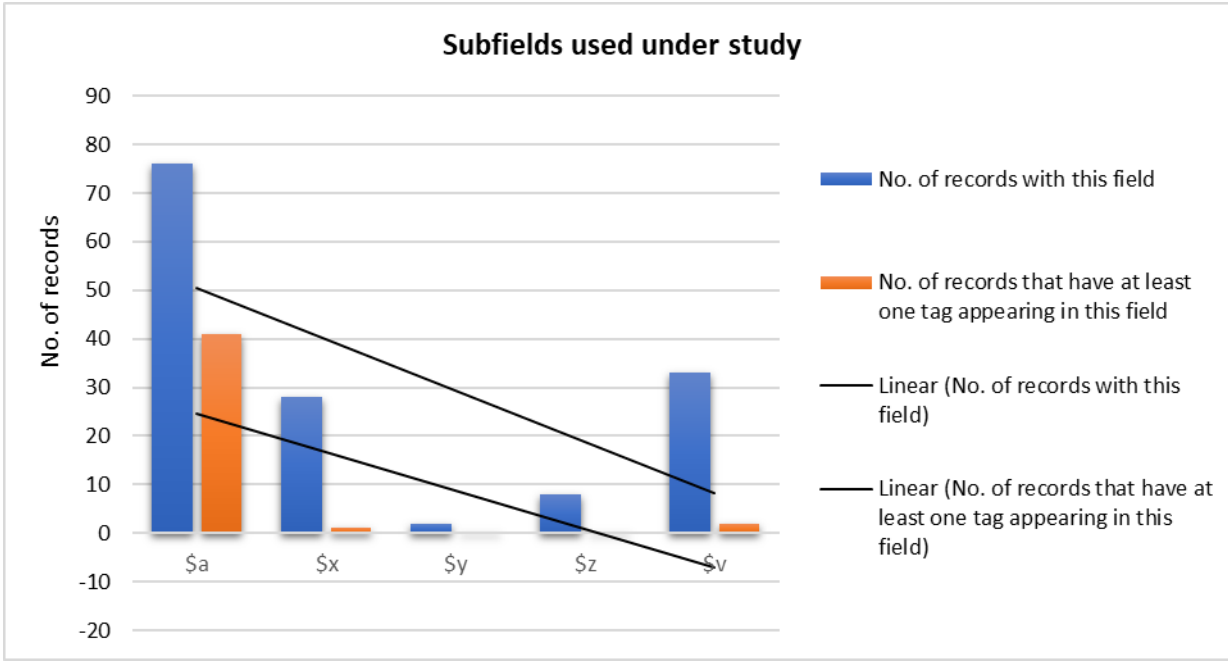


Fig.10:Usage of subfields by expert sand users comparatively

Analyzing Title-Based Terms in Social Tags and LCSH Descriptors:

Historically, document titles have served as the primary access point for searching documents, with users initially seeking information through titles before exploring subjects or authors. However, remembering the exact title of a document can be challenging, leading users to rely on topic or subject-based terms for their searches. This study examines whether social tag vocabulary and LCSH (Library of Congress Subject Headings) descriptor vocabulary include terms beyond those found in document titles.

The study analyzed a total of 155 title-based social tags and 77 title-based LCSH descriptors. Table 6 presents the findings, showing that out of 3,658 unique social tags, 69(1.88%) are title-based, while out of 98 unique LCSH descriptors, 27 (27.55%) are title-based.

These results indicate that social tag vocabulary contains a substantial 98.12% of terms beyond title-based terms, whereas LCSH descriptor vocabulary includes 62.45% of terms beyond title-based terms. This demonstrates that social tags are primarily focused on topics and subjects beyond just the titles, whereas LCSH descriptors still emphasize a significant proportion of title-based terms.

In summary, social tag vocabularies encompass a broader range of topic-based terms compared to LCSH descriptors, which retain a higher concentration of title-based terms.

Table 6: Total terms & unique terms that appeared on book titles

	Total terms	Unique terms	% of terms
Social tags	155	69	1.88 of total unique social tags
LCSH descriptors	77	27	27.55 of total unique LCSH descriptors

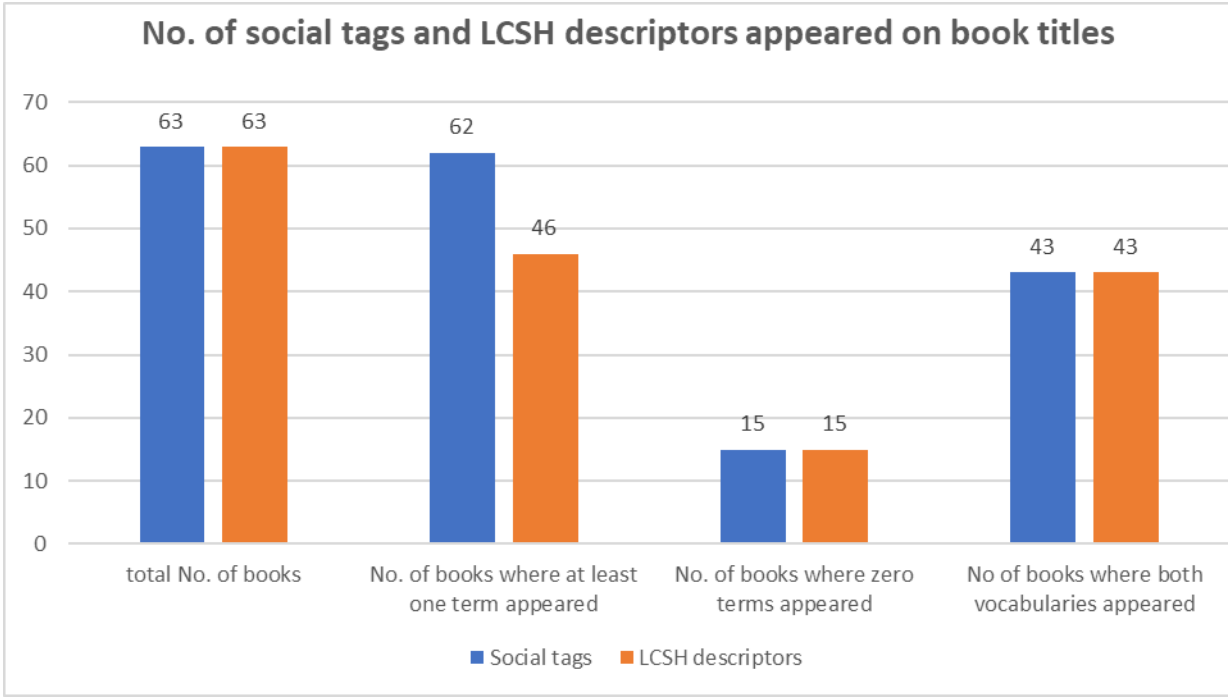


Fig.11:Title-wise social tags and LCSH descriptors appearing layout

Analysis of Terms from Book Titles in Social Tag Vocabularies and LCSH Descriptors

This study investigates the presence of title-based terms in both social tag vocabularies and LCSH (Library of Congress Subject Headings) descriptors for Tagore's literature. Specifically, it explores how many terms from book titles appear in each vocabulary and identifies which vocabulary is used more frequently to catalogue books with at least one title-based term.

Figure 11 shows that out of a total of 63 book titles, both social tag and LCSH descriptor vocabularies contain at least one title-based term. In contrast, there are 15 book titles where no terms from book titles appear in either vocabulary.

The study found that:

62 books used social tag vocabulary terms from the book titles.

46 books used LCSH descriptor vocabulary terms from the book titles.

43 books featured title-based terms in both social tag and LCSH descriptor vocabularies.

This indicates that users tend to focus more on title-based terms for tagging books compared to experts, who are less likely to use such terms for cataloguing.

Table 7 highlights the most frequent terms found in book titles across both datasets. It shows that in the 62 major books where users have used title-based terms, this is significantly higher than the 46% of books where experts have used title-based terms. Furthermore, there are 32 books (78 total books minus 46) where experts use terms that differ from those found in the book titles.

Overall, the study highlights that users tend to rely heavily on title-based terms for social tagging, reflecting their preference for keywords directly derived from book titles, whereas experts use these terms less frequently when working with LCSH descriptors, indicating a divergence in terminology preferences between general users and professionals.

Table 7:Top 10 most frequent terms found in book titles across both datasets

Social tags	Frequency	LCSH terms	Frequency
Tagore	23	Tagore	23
Rabindranath Tagore	15	Rabindranath	17
Stories	9	Literature	4
Poems	6	Modern	3
Letters	5	Poetry	3
Literature	4	Collections	3
Modern	3	Life	3
Anthology	3	Poets	2
Songs	3	Translations	2
Poetry	3	Short stories	1

Findings of the study

The present study explores the adoption of social tags within the realm of Tagore's literature, offering a comparative analysis between social tags and Library of Congress Subject Headings (LCSH) descriptors. The findings illustrate several significant aspects of the effectiveness and usability of social tags in library databases.

The comparison between social tags and LCSH descriptors reveals that social tag vocabulary is broader and more diverse than the LCSH database. Specifically, there are only 30 unique terms common to both vocabularies, which represents a mere 0.08% of social tags and 30.61% of LCSH descriptors. This indicates that while users predominantly apply LCSH descriptors as tags, experts utilize a much narrower range of terms. The study shows that 99.92% of the terms in the social tag vocabulary are distinct from those in the LCSH descriptors, highlighting the broader and more varied nature of user-generated tags compared to expert-chosen terms.

Further analysis of the top twenty terms from both datasets reveals that social tags include only twelve subject-based terms, whereas LCSH descriptors contain eleven. This suggests that users and experts nearly both tend to employ more subject-based terms. Additionally, social tags incorporate some personal tags (three instances), which may hinder the integration into library systems. With the appearance of comparatively same subject terms in social tags as LCSH descriptors the study suggests that incorporating social tags alongside LCSH descriptors could enrich library catalogues, catering to a wider array of search needs and enhancing subject access.

Examining the types of terms used in social tags versus LCSH descriptors reveals that social taggers predominantly use topical or geographic name entry element terms ($a) and topical subdivision terms($x). In contrast, form division-$v, chronological subdivision-$y, and geographic subdivision-$ztermsare use dinfrequently as tags. This finding indicates that both users and experts fav or topical and geographic terms but differ in their use of other ermtypes.The study also shows that users do not use chronological terms as tags, whereas experts do. This divergence suggests that integrating social tags could help libraries address a broader range of search queries, especially those related to chronological perspectives.

The comparison of title-based terms between social tags and LCSH descriptors reveals that social tags account for only 1.88% of terms found in book titles, across 62 major books, whereas LCSH descriptors appear in 27.55% of book

titles, but across fewer books (46). This discrepancy highlights that users are more inclined to use title-based terms, while experts avoid them. Combining social tags and LCSH descriptors could therefore enhance title-based searches and offer a more comprehensive approach to cataloguing Tagore's literature.

Overall, the study underscores that while social tag vocabulary shares some similarities with LCSH vocabulary, the two serve different purposes. Experts use LCSH to enhance subject-based retrieval, while social taggers create tags based on personal perspectives and needs. The varied nature of social tags—ranging from terms like 'wishlist' and 'to-read' to traditional subject terms like 'Poetry' and' Literature'—demonstrates their potential to complement LCSH descriptors. By adopting social tags in libraries, it is possible to create richer, more versatile catalogues that address diverse use rneeds and strengthen subject access.

Conclusion and further study

The study examined the application of social tags in libraries, particularly in medium and large libraries using Library of Congress Subject Headings (LCSH). It found that social tags and expert-generated descriptors, such as LCSH and Sears List of Subject Headings (SLSH), exhibit different vocabularies and perspectives. While expert-generated terms follow controlled vocabularies for precision and consistency, social tags reflect user perspectives, often including personal, ambiguous, or context-specific terms. This divergence indicates that users and experts perceive and describe the same documents differently, highlighting the potential of social tags to supplement expert metadata.

Controlled vocabularies, like LCSH, provide uniformity and help overcome issues in natural language searches, offering a structured approach to categorizing and retrieving resources. On the other hand, social tags, though prone to ambiguity, synonyms, and personal terms, can enhance recall by introducing a broader range of terms, including those that resonate with users.

However, the study cautions that personal and non-subject tags, such as "wishlist" or "to-read," can reduce catalogue efficiency if not properly managed. Despite the challenges, incorporating social tags can create a more user-centric library environment, fostering engagement, satisfaction, and community among patrons. Tagging allows users to express their perspectives, making the library catalogue more accessible and interactive. This approach has already been implemented in pioneering libraries, such as the University of Pennsylvania's PennTags and the University of Michigan's Mtagger, as well as in public libraries that have adopted systems like LibraryThing for Libraries (LTFL) and Social OPACs (SOPAC).

While social tags present inherent issues like inconsistency and lack of structure, several studies suggest their value in enhancing resource descriptions. For instance, Bischoff et al. (2008) found that social tags work well as metadata for music resources, while Syn and Spring discovered their effectiveness in describing academic papers. Noll and Meinel (2007) also found social tags superior to author-generated metadata in describing web content. The study stresses the importance of libraries establishing clear guidelines for incorporating social tags. Without proper management, issues like synonyms, homonyms, and irrelevant tags can complicate searches. Researchers such as Spiteri recommend that libraries adopt written guidelines on the use and form of social tags to ensure they are effectively integrated. Additionally, tools like automatic indexing software can help filter and manage social tags, addressing issues of ambiguity and inconsistency.

Ultimately, the study concludes that integrating both expert-generated and user-generated metadata is crucial to meeting evolving user needs. This "hybrid metadata ecology" would combine the precision and reliability of controlled vocabularies with the flexibility and inclusivity of social tags, offering users a broader range of access points to library resources. Social tags can enrich library catalogues by introducing user-driven terms that complement but do not replace, expert-generated descriptors. Such a system would enhance both the subject accessibility of collections and the overall user experience, making libraries

more relevant, dynamic, and user-friendly in the digital age. A future study on the lexical types of social tags—single-word, multi-word, subject-based, or personal—would provide deeper insights into their usage, semantics, and cultural relevance. Further research should explore the comparison of social tags with other controlled vocabularies, such as the Sears List of Subject Headings, to examine broader metadata applications. While this study focused on lexical matching, future research should assess social tags at the semantic level to better understand their effectiveness in capturing the nuances of literary works like Tagore's. Expanding the study to analyze retrieval performance in popular online databases and search engines would provide a more comprehensive evaluation of the effectiveness of social tags alongside LCSH descriptors.

References

Adler, M. (2009). 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), 309–331. https://doi.org/10.1080/19322900903341099

Bogers, T., & Petras, V. (2017). An In-Depth Analysis of Tags and Controlled Metadata for Book Search. In Proceedings of iConference *2017* iSchools. https://doi.org/10.9776/17002

Carman, Nicholas (2009). LibraryThing Tags and Library of Congress Subject Headings: a Comparison of Science Fiction and Fantasy Works. Open Access Te Herenga Waka-Victoria University of Wellington. Thesis. https://doi.org/10.26686/wgtn.16969825

Fireflies by Rabindranath Tagore. (n.d.). Retrieved July 20, 2024, from https://www.librarything.com/work/249496

Heymann, P. & Garcia-Molina, H. (2009). Contrasting Controlled Vocabulary and Tagging: Do Experts Choose the Right Names to Label the Wrong Things?. Second ACM International Conference on Web Search and Data Mining (WSDM 2009), Late Breaking Results Session (p./pp. 1--4), February, : Stanford InfoLab.

Lawson, K. G. (2009). Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers. The Journal of Academic Librarianship, 35(6), 574–582.

https://doi.org/10.1016/j.acalib.2009.08.020*LC* Catalog. (n.d.). [Webpage]. Retrieved July 20, 2024, from https://catalog2.loc.gov/index.html

Library of Congress Subject Headings PDF Files. (n.d.). [Web page]. Retrieved July 20, 2024, from //www.loc.gov/aba/publications/FreeLCSH/

Linking Folksonomy to Library of Congress Subject Headings: An Exploratory Study. (n.d.). Retrieved August 19, 2025, from https://www.researchgate.net/publication/235307151_Linking_Folksonomy_to_Library_of_Congress_Subject_Headings_An_Exploratory_Study

Lu, C., Park, J., & Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. Journal of Information Science, 36(6), 763–779. https://doi.org/10.1177/0165551510386173

MARC 21 Format for Bibliographic Data: 6XX: Subject Access Fields-General Information (Network Development and MARC Standards Office, Library of Congress). (n.d.). Retrieved July 20, 2024, from https://www.loc.gov/marc/bibliographic/bd6xx.html

Noll, Michael G., and Christoph Meinel. "Authors vs. Readers: A Comparative Study of Document Metadata and Content in the Www." Proceedings of the 2007 ACM Symposium on Document Engineering, 2007, pp. 177–86, https://doi.org/10.1145/1284420.1284465.

Rolla, P. J. (2010). User Tags versus Subject Headings Can User-Supplied Data Improve Subject Access to Library Collections? https://www.semanticscholar.org/paper/User-Tags-versus-Subject-Headings-Can-User-SuppliedRolla/57d3e362b83a2c9d553bdba01a15d9765900f3f4

Samanta, K., & Rath, D. (2019). Social Tags Versus LCSH Descriptors a Comparative Metadata Analysis in the Field of Economics. DESIDOC Journal of Library & Information Technology, 39, 145–151. https://doi.org/10.14429/djlit.39.4.14291

Tagore, R., & Chakravarty, R. (2010). The land of cards: Stories, poems, and plays for children. Puffin Books.

Thomas M, Caudle DM, Schmitz CM (2009), "To tag or not to tag?". Library Hi Tech, Vol. 27 No. 3 pp. 411–434, doi: https://doi.org/10.1108/07378830910988540

Vaidya, P., & Harinarayana, N. (2017, November 1). The role of social tags in web resource discovery: An evaluation of user-generated keywords. https://www.semanticscholar.org/paper/The-role-of-social-tags-in-web-resource-discovery%3A-Vaidya-Harinarayana/60d2092b7c7872084c90cc424f4d935cac7cf592

Wu, D., He, D., Qiu, J., Lin, R., & Liu, Y. (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(2), 169-187. https://doi.org/10.1177/0165551512451808 (Original work published 2013)

Zeitgeist | LibraryThing. (n.d.). LibraryThing.Com. Retrieved August 19, 2025, from https://www.librarything.com

Acknowledgments

The authors extend their heartfelt appreciation to the anonymous reviewers for their insightful feedback and suggestions, which significantly enhanced the quality of the manuscript.

Conflict of interest

The authors declare no conflict of interest.