Social Tags of Select Books Written by Mahatma Gandhi: A Comparative Study of Library Thing Tags and OCLC Fast Subject Headings

Sachin Katagi#,* and Bhakti Gala$

#Karnatak University, Dharwad - 580 003, India
$Central University of Gujarat, Gandhinagar - 382 030, India
*E-mail: sachinskmatagi@gmail.com

ABSTRACT

Retrieval to specific information is critical to user satisfaction in today’s semantic web environment. Users use various terms, nomenclatures and words for aiding mnemonic value to identify concepts for future access and sharing with community members. Social tagging has emerged as a popular option for the millennials to address this issue. Tagging is a convenient way used by today’s users for naming their web resources in their own terms. In this study a comparative study was carried out between the standardised Online Computer Library Centre (OCLC) Faceted Application of Subject Terminology (FAST) Subject headings and the popular user-generated tags observed on the LibraryThing website for select books written by Mahatma Gandhi, the iconic Father of Modern India, who was also a prolific writer. M K Gandhi wrote many books, letters, newsletters, essays and edited journals on various subjects like health, vegetarianism, diet, religion and social reforms. A close examination of the subject headings and tags given by users to works authored by M K Gandhi reveal that there is a vast amount of difference between the standard terms assigned using OCLC FAST Subject Headings and tags assigned by the users. The study reveals that neither subject headings nor tags are perfect systems by themselves, but they may complement each other in library catalogues. The inclusion of user-generated keywords into catalogues will greatly enhance representation, organisation and retrieval of resources in a library environment.

Keywords: Information retrieval systems; Folksonomies; Social tagging; OCLC fast subject headings; Controlled vocabularies; Library Thing Tags; Metadata; Mahatma Gandhi.

1. INTRODUCTION

Information professionals have long understood the role and importance of subject headings as metadata connecting user with the relevant resources. Graham & Abbas, (2006) state its role in aiding in the identification, description, management and location of information resources in both digital and non-digital environments. Libraries have traditionally used controlled vocabulary tools for aiding users meaningful retrieval of desired information. Controlled vocabularies contain selected words and phrases which are used by professionals to index the content for higher precision and recall in the search and retrieval process. These controlled vocabulary tools help in achieving a standardisation in the rendering of names of various concepts such as subject names, people names, geographic names, etc. Some of the more popular controlled vocabularies used by libraries globally are Library of Congress Subject Headings (LCSH), Sears List of Subject Headings (SLSH), Medical Subject Headings (MESH), Book Industry Standards and Communications (BISAC) and Faceted Application of Subject Terminology (FAST) designed by Online Computer Library Centre (OCLC) and termed as OCLC FAST in this study.

A disconnect with the subject headings given by the library cataloger using these standard vocabulary control tools has given rise to the concept of Folksonomy by the users. The current web environment further facilitates users to identify resources by assigning their own subject headings giving tags to resources, which further help the user and others to retrieve the resources.

The term Folksonomy, a portmanteau of the terms, ‘folk’ and ‘taxonomy’ was coined by Thomas Vander Wal as a popular web 2.0 tool in 2004. Thus folksonomy represents the ‘nomos’ i.e. management of ‘taxis’ which is ‘classification’ given by the ‘folk’ which are people. It is result of the process of tagging done by online users. Folksonomy is a set of large terms which have been purely generated by the users, giving freely chosen keywords. Various studies (Gupta, Li, Yin, & Han, (2011); Cordón-García, Alonso-Arévalo, Gómez-Díaz, & Linder, (2013) and Thomas, Caudle, & Schmitz, (2009) refer to folksonomy ‘People – Powered Metadata’. In folksonomy, classification is done by using labels which are popularly known as tags. Folksonomy is a relatively new aspect of personal organisation of information Steele, (2009). Users can assign tags to any document, online links, and online photograph. Tagging works with a process known as social bookmarking where user generated tags are assigned to these bookmarks.
Mohandas Karamchand Gandhi, popularly known as Mahatma Gandhi, was a global icon who led the non-violent struggle for independence in South Africa and India. A prolific writer, M K Gandhi authored innumerable literary works, in the form of books, letters, newspaper articles, and essays. He wrote eight books and was the editor of many newspapers and journals. He wrote mostly in Gujarati and translated his own works in English and Hindi. Among his notable literary works are Hind Swaraj (1909), and his autobiography My Experiments with Truth (1927) among others. The publications division of Government of India published The Collected Works of Mahatma Gandhi documenting Gandhiji’s words from the year 1884 till his assassination on January 30, 1948 which are found on the Gandhi Heritage Portal (https://www.gandhiiheritageportal.org/). He wrote on various subjects such as non-violence, social reformation, diet and vegetarianism, religion, the Bhagavad Gita, and health. This study has looked at the various social tags used by readers globally to identify and retrieve works of M K Gandhi and compared these tags with the controlled vocabulary terms used. As India celebrates the 150th birth anniversary of M K Gandhi, the study demonstrates the folksonomy or social tags used by readers to recollect and read works authored by M K Gandhi on the popular social tagging site LibraryThing.

2. LITERATURE REVIEW

The Online Computer Library Center’s (OCLC) FAST program is one approach to integrating access to library collections and services into the “flows” of Web users Nilges, C. (2006). For many years, the OCLC Office of Research has focused on the areas like prototyping classification tools for catalogers, developing automated classification software, applying and refining statistical mapping techniques, next generation library catalogs including tag clouds and multi-faceted searching and navigation and its recent focus is on emerging data models and new technologies (e.g., SKOS and linked data) which will provide new opportunities for publishing, linking, and sharing data in the years to come. Mitchell, Joan S. and Diane Vizine-Goetz. (2010)

The rapid growth and creation of enormous volume of e-resources in the present day has led to emergence of numerous metadata schemes. After analysing the requirements of the metadata scheme which is simple, easy to apply and to comprehend, OCLC came up with a new schema known as FAST. Within a short span of time various organisations both public and private have expressed the interest in implementing FAST in their catalogue work flow Chan, Lois Mai et al (2001). Further O’Neill, E. T., & Mai Chan, L. (2003) in their article they discuss about the development and syntax of FAST which consists of 8 distinct facets.

Web 2.0 is emerging as a centre of importance in today’s information world, there is a paradigm shift that the way information technology is used in present days for dissemination of information (Santosh, 2017). Faci, Maamar, Burêgio, Ugljanin, & Benslimane (2017) conducted a study on the application of web 2.0 tools in workplaces and highlighted that web 2.0 is very essential in today’s competitive environment. Here literature is carried out under two sections i.e. presence of social tags in libraries and comparison of social tags vs traditional subject headings.

Tags can be defined as a meaningful word, URL, phrase or code, a string of numbers with the purpose to describe the content of a physical document by the individual users. Although tags will not replace subject headings Gerolimos, (2013) there are advantages of discovery and retrieval at low cost, and finding everything in a single search box. In the rapidly evolving digital world tagging is the hottest trend in organizing personal resources in a digital environment and tagging has created a closer, more personalised relationship between user and their resources or collections (Samanta, K., & Rath, D., 2019). The modern cataloger can overcome limitations of the traditional catalogue with the aid of folksonomies (McFadden & Weidenbenner, 2010; Gupta et al., 2011) Tagging can be seen as best as a parallel value-added process rather the replacement of traditional subject headings in a library. In his thesis Kathuria, (2011) states that academic libraries which add social tags to institutional repository websites help in the retrieval of the elusive grey literature.

Literature on social tagging which is already implemented in some of the libraries raises concern about how the tags complement traditional subject access for the documents. In general, scholars found that social tagging is another way to approach documents and helps in subject access for materials. However, tags entirely cannot replace controlled vocabularies Vaidya & Harinarayana, (2016) because they provide greater information retrieval. Adler, (2009) in her study compared Library of Congress Subject Headings (LCSH) with user generated tags collected from LibraryThing on transgender people. In their study they found that controlled vocabularies like Library of Congress Subject Headings do not give any space for keywords which were user generated and in such instance a folksonomy comes to the aid of users. Most studies on social tagging have given the result that tags provide further value additions to the traditional subject headings.

Westcott, Chappell, & Lebel, (2009) in their study stated that Claremont University users were enthusiastic and happy about LibraryThing tags and were helpful in retrieving related documents. Further, in folksonomies users can add new terms very quickly and easily whereas in controlled vocabularies the process of adding new terms may take a period of months or years (Pirmann, 2012).

In his analysis on book tagging on LibraryThing, Bartley (2009) found that tags were matched with medical subject headings (MESH) and they found that only 20 per cent of the tags overlapped with controlled vocabularies. The study proved that there were lots of differences between the terms used by professional indexers and user generated tags. User generated tags yield better in showing the latest subject headings, thereby increasing natural language retrieval. Murphy & Rafferty, (2015) conducted a study and criticised about tags and mentioned that for the post structuralist project tagging is the imperfect analogy. Though all these studies state the benefit of the LibraryThing, it is also not free from criticisms. Folksonomies are basically uncontrolled in nature in particular, lexical anomalies, lack of synonyms and homonyms, no rules and suffer from high degree of messiness and inconsistencies.
3. RESEARCH METHODOLOGY

In the backdrop of the review of literature, the present study was guided by the following primary research question.

RQ 1: What are the similarities and differences in the tags given on the LibraryThing website and subject headings given by OCLC FAST for select books written by Mahatma Gandhi, a global cult figure?

The study compares the subject headings given using the OCLC FAST with the tags given on the website LibraryThing. OCLC FAST is basically derived from the Library of Congress Subject Headings and is designed to meet the general use of assigning subject terminologies. On its website (https://www.oclc.org/research/themes/data-science/fast.html) the purpose behind designing OCLC FAST has been stated as retaining the vast vocabulary of Library of Congress Subject Headings (LCSH) and making it easier to understand and learn with navigation friendly facets and modern in its design and application. OCLC FAST is mainly designed for an online environment which is a post-coordinated faceted vocabulary and used by library professionals with minimal training and experience. It consists of eight distinct facets: Topical, Geographic (Place), Personal Name, Corporate Name, Form (Type, Genre), Chronological (Time, Period), Title, and Meeting Name. OCLC FAST has a several advantages like mapping of subject data and cross-domain searching, more amenable to computer-assisted indexing and facilitate computer-assisted authority control. O’Neill, E. T., & Mai Chan, L. (2003)

LibraryThing is a website which provides an online service to people to catalogue their own books. LibraryThing was created by Tim Spalding. It provides an online service that allows the general users in helping catalogue their books and share their catalogue with other readers. Cordón-García et al., (2013). It connects people who are reading the same book and provides suggestions by other members for what to read next and so on Gisela & Dash, (2015). There are about 2,444,318 members with 137,890,018 catalogued books and 150,137,845 tags added to these resources in LibraryThing. [Data as on (11-09-2019 12:53)]

LibraryThing provides in-depth insight into how users assign meaning for the resources they have read by assigning social tags to the resources. The website allows users to search titles to see how many other readers have that book on their shelves, and how many have reviewed it. There are also suggestions of related books to read. Kathuria (2011) states that “it’s a virtual feast of information.” It also offers various features like when a tag in the LibraryThing is clicked, it leads to a new page with other books which have the same tag and other forms of the word used (Variations in the spelling of the Tags), provides information about how many members and other terms which are related to each other. “Tag Mashes” in a LibraryThing website provides a detailed list of users who have also used the same Tag. LibraryThing is integrated with Google “NGram Viewer” and gives statistical data about how many times the item has been used. Here users are also allowed to manipulate the years, language and how the usage of the term has changed over the years. Such features are very useful to build the users-oriented community atmosphere. Library thing allows searching of catalogues of Library of Congress and more than 80 world libraries. According to LibraryThing website, users can search, sort and they may use DDC and Library of Congress classification schemes to organise their personal collection Kathuria, (2011). Mendes, Quinonez-Skinner, & Skaggs, (2009) identified that LibraryThing provides three main features like tag clouds, links to other editions, translations, of a work and recommendations of books on the basis of books that are already read by users. LibraryThing for Libraries (LTFL) is a website which displays user-generated metadata from the “social cataloguing website” LTFL consists of mainly three types of data like Tags, links and recommendations to other similar works and he concluded that in a pool of knowledge where potential resources can be found easily by using tags.

In most of the articles reviewed by the researcher content analysis was the apt method used to analyse the tags and subject headings. Porter, (2011)

The researcher selected books that were written by M.K Gandhi because Mahatma Gandhi is a famous personality not only in India but also in the whole world and in 2018 India is celebrating the 150th birth Anniversary of Mahatma Gandhi. In India Mahatma Gandhi is also known as the Father of the Nation, because he has made a huge contribution in providing independence to India from the British Rule.

4. DATA COLLECTION

Secondary data were collected for the present study. Initially the researcher considered comparing LibraryThing tags with Goodreads (a popular social bookmarking site), Library of Congress Subject Headings, and IndCat. IndCat is an online union catalogue of Indian universities developed and maintained by INFLIBNET and National Library of India catalogue were also tested, but unavailability of subject headings for the selected book titles, the website server being down and complete lack of subject headings in the Machine-Readable Catalogue (MARC) records resulted in the selection of OCLC FAST. Thereafter, a comparative study between LibraryThing tags and OCLC FAST subject headings which were derived from British National Bibliography (BNB) was decided upon, as all the book titles selected were found to be present in the BNB. Secondary data such as the tags and the relevant subject headings was collected from LibraryThing website and British National Bibliography website. Web content analysis method was used to formalise this collected data. To collect tags and subject headings researcher choose “M.K Gandhi” as a keyword. All the data was collected in the first week of March to last week of April 2018.

Criteria for the selection of data (books) were:

- Books had to be written by M.K Gandhi, and
- Metadata of those books had to be available on both websites. (LibraryThing and British National Bibliography)

The following five books were selected which fulfilled the above criteria:

• The Bhagavad Gita according to Gandhi
• The Way to God
• Prayer
• The Penguin Gandhi Reader.

Data was collected firstly by retrieving the tags with the keyword “M.K Gandhi” in both the search box of British National Bibliography and LibraryThing websites and search results were retrieved, copied and pasted in a Microsoft Excel Sheet and the same Excel Sheets were used for further data analysis.

5. DATA ANALYSIS

The analysis of the secondary data collected was carried out using Microsoft Excel by comparing the following:
• Comparison between LibraryThing Tags and OCLC FAST subject headings
• In-depth analysis of LibraryThing Tags and
• Top five most repeated tags in all the five books.

For analysis of the data the researcher coded the collected data and formed the different categories of tags. The tags collected from both LibraryThing and OCLC FAST from British National Bibliography were broadly coded under the following groups:
• Exact synonyms: Exact synonyms are those where there is an exact identical match between LibraryThing tags and OCLC FAST subject headings
• Near synonyms: Near synonyms are those where there the terms matching between LibraryThing tags and OCLC FAST subject headings have minor variations
• Broad and narrow tags: Broader terms are those which indicate broadly the contents in a book, whereas narrow term is much more specific to the contents in a book
• Unique tags: Unique tags are those tags not available in OCLC FAST subject headings
• Stand-alone tags: Stand-alone tags are tags which were available in OCLC FAST subject headings but found to be present with minor variations embedded within the titles of the books selected for the study
• Irrelevant tags: Irrelevant tags are those tags present in LibraryThing which were assigned to the resources for personal reasons to manage their own collection, but they do not have relevance from the access perspective.

5.1 Comparison between LibraryThing Tags and OCLC FAST Subject Headings

Here the researcher conducted a comparative study of exact synonyms, near synonyms and broad and narrow terms between LibraryThing Tags and OCLC FAST Subject Headings.

Table 1 shows that there were about total 131 tags from LibraryThing and 103 tags were found relevant to the contents of the book i.e. 79 per cent of the tags were found relevant. It is also noted that within the relevant tags 6 were exact synonyms, 31 were near synonyms and 8 were broad and narrow terms, the proportion between LibraryThing tags and OCLC fast subject headings is 105:21.

5.2 Distribution of Tags According to their Percentage for all books

The formula used for the calculations of tags is as under:

\[
\text{Category of tag chosen for the study} \times 100 / \text{Tags which were found relevant to the contents of the book}
\]

Here category of the tag chosen for the study will be Exact Synonyms, Near Synonyms, Broad and narrow tags, Unique tags, Stand-alone tags, Irrelevant tags and Total number of tags found relevant for the study will be 103 and further by multiplying this equation with 100 we will get the total percentage of distribution of tags.

From the Fig. 1, over all there were about 131 tags collected from LibraryThing and 21 OCLC Fast subject headings. The comparison of LibraryThing with OCLC FAST subject headings show that 16 per cent of the tags are near synonyms, 3 per cent of the tags were exact synonyms, and 4 per cent of the tags were broad and narrow terms.

5.3 In-depth analysis of LibraryThing tags

From Fig. 2, over all there were about 131 tags collected from LibraryThing, In Depth analysis of LibraryThing Tags shows that 69 per cent of the tags were Unique Tags, 18 per cent were Non-relevant tags, 13 per cent of the tags were standalones tags.

5.4 Top Five Most Repeated Tags in all the Five Books.

Here the researcher conducted a study of the top 5 most repeated tags of only LibraryThing Tags.
From Fig. 3, it indicates that most repeated tags were Gandhi, Religion which were repeated in all the 5 books and Hinduism, non-fiction, philosophy were repeated in 4 books. Even though these are the most popular tags in all the 5 books but researcher has not found their presence in the OCLC FAST subject headings.

6. CONCLUSIONS

This article has set out to mark the distinction between the LibraryThing tags and OCLC Fast Subject headings in the comparative analysis, it is observed that, tags were more compared to subject headings and 79 per cent of the tags were found relevant to the study. Further in depth analysis of LibraryThing tags shows that 69 per cent of the tags were unique and these represents the books uniquely. Even though these were found relevant in identifying the resources they were not found in OCLC fast subject headings. In the analysis of top 5 most repeated tags it is observed that M. K Gandhi is a central part of all the data collected but “Gandhi” is not a preferred term in the OCLC FAST subject headings. From this it is concluded that apart from subject headings tags play a major role in retrieving the documents.

The study reveals some of the features of folksonomies as tags are richer and represent the core of the text with terms which were not found in a control vocabulary, they are also democratic in nature and offers changes in the vocabulary. Whereas controlled vocabularies are universal, and they will offer precision in finding the required resources. It is concluded that no system is perfect. Neither folksonomies nor controlled vocabularies, although tags are dynamic, user centric but they lack the precision.

Now a days tagging has become a ubiquitous part of the information landscape. In the modern era tagging system has to be used highly in a very robust manner. Nowadays more and more information is generated in all the disciplines of the universal knowledge with only fewer resources to annotate them, so tagging system needs to be strengthened so that excellent search and recommendation system has to be built for the 21st Century users.

This research has helped in identifying areas for further exploration and research on various other aspects of tagging. Present study has been conducted by taking the books written by M.K Gandhi similar study may also be conducted by using books written by another famous personality of Indian origin i.e. Rabindranath Tagore, who is also a Nobel laureate.

REFERENCES


12. Heymann, Paul & Garcia-Molina, Hector. Contrasting controlled vocabulary and tagging: Do experts choose the right names to label the wrong things? In Second ACM International Conference on Web Search and Data Mining (WSDM 2009), Late Breaking Results Session, 2009, 2, 9-13, Barcelona, Spain.


CONTRIBUTORS

Mr. Sachin Katagi completed his Master of Library and Information Science from the Central University of Gujarat. He is a recipient of Indian Library Association- Dr K Padma Umaphathy and KS Umaphathy Fellowship in Library and Information Science, Travel Grant Recipient of Indian Public Library Conference (IPLC) 2017, New Delhi and he also received the “Best Poster” award in the LIFE IISER Pune International Conference in 2017. His areas of Interests were Folksonomy, ETDs etc.

Dr Bhakti gala is an Assistant Professor at the Central University of Gujarat, India. She obtained her doctoral degree in library and information science from The Maharaja Sayajirao University of Baroda, Vadodara. Her research interests include information literacy, public libraries, metadata, digital scholarly communication and digital libraries. Her areas of research interests are Metadata, Open Access, Information Literacy, Public Libraries and Digital Libraries. She is also the recipient of the 2014 ASIS&T SIG III Best International Paper Contest award.