A. I. M. JAKARIA RAHMAN

Social tagging versus Expert created subject headings

Supervisor: Professor Ragnar Nordlie

Master thesis
International Master in Digital Library Learning
2012

Abstract

The purpose of the study was to investigate social tagging practice in science book context. In addition, it identified the usefulness of social tags as supplementary of controlled vocabulary to enhance the use of library resources. More specifically, this study examined to know to what extent the social tags match with controlled vocabulary, and whether or not any additional perception is provided by social tags to improve the accessibility and information retrieval in a digital environment. In both cases, the social tags were considered with respect to the appropriateness to the specific book. For the successful implementation of social tagging in library systems, there is a need to understand how users assign social tags to library collections, what vocabularies they use and how far the social tag relates to controlled vocabulary. This understanding can help libraries to decide on how to implement and review the social tagging.

This study used a combination of both qualitative and quantitative research approaches. The LibraryThing website and Library of Congress Subject Headings were considered as a research site. Social tags have been collected from the LibraryThing website and LCSHs has been considered as controlled vocabulary. Twenty books from the science genre have been chosen purposefully. The sample has further been considered to include only those books that have also been available in the Library of Congress catalogue. Ten books have been taken from the academic group and the remaining were from the non-academic group. This study took into consideration only those social tags that occurred at least twice.

A coding system has been developed to pull together all the similar social tags for further analysis. In the coding system, four broad categories have been defined, e.g., Social tags that match exactly with LCSHs, Social tags that match partially with LCSHs, Social tags that reflect bibliographic information and social tags that are user specific information. The last three categories were further sub-categorized.

It is found that there is a clear difference between assigning expert created subject terms and social tagging practice to library books. Cataloguers assigned relatively few terms per book through the use of restricted and established vocabulary following firm rules, whereas, the end users enjoyed liberty with unlimited terms. More than fifty percent of

social tags matched with expert created subject headings. The frequency of use of the social tags that matched with LCSHs terms was higher than the non-matched ones. The expert created subject headings were highly ranked in the social tags' lists, where end users more frequently assigned social tags that represented broader or narrower terms than the cataloguers' assigned subject headings. In addition, the social tagging represented other aspects that could not be either covered within the strict subject headings assigned rules or cataloguing rules. Such diverse impressions can be seen as an access point to the same library collections according to users' interest and opinions.

This study revealed that as a standalone tool neither the controlled vocabulary nor the social tagging practice can work like a satisfactory information retrieval tool. A hybrid catalogue with combining both LCSHs and social tags would give its patrons the best of both worlds in terms of access to materials. This kind of practice may give more significant outcome for local research or university libraries where the users are more concentrated on a defined number of disciplines. Adapting users' views in addition to controlled vocabulary through social tags may increase the efficiency of information retrieval process in library OPAC.

This study implied both qualitative and quantitative support for the use of social tags in the library OPACs. The findings support many of the previous theories proposed in literature about social tagging and LCSHs. The qualitative analysis of social tags disclosed the diverse way of looking at the library resources by the end users in addition to subject descriptors.

Keywords: Social tags, Controlled vocabularies, Library of Congress Subject Headings, LibraryThing, Science genre, Web 2.0, Online Public Access Catalogue (OPAC)

Declaration

I hereby declare that the subject matter of this thesis is the original research work done by me. The content of this thesis was not taken from any previous work of me or to the best of my knowledge not from anybody else. I have not submitted this thesis to any research degree in any other University/Institution.

I have duly acknowledged all the intellectual thoughts and opinions of others found relevant that have used in this thesis. The work has been under the guidance of Professor Ragnar Nordlie at the Oslo and Akershus University College of Applied Sciences, Norway. The thesis is submitted as a part of the International Master Program in Digital Library Learning that has been conducted in cooperation with Oslo and Akershus University College of Applied Sciences, Norway, Tallinn University, Estonia and Parma University, Italy under the Erasmus Mundus Scholarship.

………………………………………………..

A. I. M. Jakaria Rahman [Candidate's signature]

[Submitted electronically]

Acknowledgements

I would like to express my deepest gratitude to Professor Ragnar Nordlie, who has given his vigilant supervision for shaping up this work. It was his continuous encouragement and valuable direction, which enable me to look into the depth of the research problem. I am extremely indebted and thankful to him for his continuous support. .

My sincere gratitude goes to the European Commission, and the Erasmus Mundus Scholarship Program for financial support for the two-year International master program. I am grateful to the DILL Consortium for making it possible for me to be part of this great program.

My appreciation goes to all faculty members of DILL program, Ragnar Audunson, Ragnar Nordlie, Nils Pharo, Tor Arne Dahl, Sirje Virkus, Aira Lepik, Anna Maria Tammaro, Vittore Casarosa, Elena Corradini, Pat Dixon, and R. David Lankes. Thank you very much for opening my eyes to valuable knowledge and sharing your experiences. I am also grateful and acknowledge the significant role of visiting professors, and library professionals, for sharing knowledge and experience during the entire DILL program 2010-2012.

I would also like to thank the administrative coordinators Kersti Ahrén Heløe (Oslo), Merje Songe and Karin Oolu (Tallinn), Federico Monaco (Parma) for their support throughout the DILL program. I am thankful to DILL colleagues who made this two- year of my life memorable. I really appreciate for being friends with all of you.

Finally, my sincere thanks and appreciation are to my parents, younger brother, and my wife for the loneliness they experienced. Without their sacrifice, patience and support, the completion of the entire DILL program and thesis work would not have been possible.

Last but not the least; I want to say Tusen Takk! Aitah! Grazie! Thanks to all. ধন্যবাদ।

Oslo   A. I. M. Jakaria Rahman
June, 2012

Dedication

Dedicated to my beloved parents

Md. Fazlur Rahman and Sabera Khatun

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CSH   :   Canadian Subject Headings

DDC   :   Dewey Decimal Classification

ISBN   :   International Standard Book Number

LC   :   Library of Congress

LCSH   :   Library of Congress Subject Headings

LGBTQ   :   Lesbian, Gay, Bisexual, Transgender, and Queer

MARC   :   Machine-Readable Catalogue

MeSH   :   Medical Subject Headings

OCLC   :   Online Computer Library Center, Inc.

OPAC   :   Online Public Access Catalogue

PDF   :   Portable Document Format

SLSH   :   Sears' List Subject Heading

SPSS   :   Statistical Package for the Social Sciences

Chapter I: Introduction

1.1 Prologue

Social tagging is one of the most popular Web 2.0 applications and has attracted the interest of libraries and museums, which has developed services that facilitate user-community collaboration (Kim & Abbas, 2010). The rigidity of the underlying taxonomical structures and the difficulty of introducing change in the categories are common limitations of subject headings. In the digital environment, social tagging has the potentiality to overcome certain limitations of traditional subject headings (Yi & Chan, 2009; Carla S., 2010; Tripathi & Kumar, 2010). In order to make end users more dynamic participants, the libraries' Online Public Access Catalogues (OPAC) allow individuals to express their views through social tagging practice about the publications they read (Anfinnsen, Ghinea, & de Cesare, 2011). Social tagging represents a supporting technology to existing subject classification systems helping to describe library resources more flexibly, dynamically and openly (Jäschke et al., 2008). It is important to examine the differences and connections between social tags and experts' assigned subject terms, and what other kind of added value is expressed by those social tags in addition to subject description. This study led to know to what extent the social tags match with controlled vocabulary, and whether or not any additional benefits provided by social tags to improve the accessibility and information retrieval in a digital environment.

This study considers the term 'social tag' as the descriptors, which may be single words or phrases, assigned to a website or other resource (e.g. books), typically by the users of the site. The term 'expert created subject headings' refers to a set of terms or phrases (descriptors or keywords), created and maintained by the concerned authority of the controlled vocabulary.

1.2 Background and context of the study

1.2.1 Information retrieval practice

Today's information seekers are different from yesterdays, and libraries should look for ways to adapt to a changing world and keep their services relevant (Sadeh, 2007). In the

era of card catalogues, subject heading string systems were created and maintained according to commonly accepted rules. Earlier, the user could browse and locate 'subject card' with the terms. Later, libraries discontinued card catalogue and adapted automation and launched OPACs that are more user-friendly. This offered three main options of searching i.e. to browse subject headings, or to search using key words, or search by title or authors key words (Landry, Bultrini, O'Neill, & Roe, 2011). The users were obliged to search for general subjects while the more specific information was unavailable for them (Lopes & Beall, 1999; Chan, 2005). The detailed subject access to documents has become the vital need in the online environment.

1.2.2 Application of expert created terms in the OPAC

Subject headings present rigid taxonomies that make changes in categories quite difficult (Antelman, Lynema, & Pace, 2006). In fact, such systems undergo change based on the work and decisions taken by the concerned authority only. This policy has the benefit of making the categorization of concepts and subjects consistent across libraries. However, the sole adoption of subject headings as the basis of providing knowledge on published material reduces the possibility of other forms and levels of information (Munk & Mork, 2007, Kakali & Papatheodorou, 2010; Anfinnsen et al., 2011). In order to give the maximum benefit to the end users, libraries need active participation of the users. In addition, libraries need to adopt new web technologies and let the users express their preferences about the library resources they use. Users would be able to express their views about a certain publication by conceptually attaching keywords or phrase to it.

1.2.3 Criticism on expert created subject headings

Library of Congress Subject Headings (LCSHs) is the most widely subject indexing language in the world and has been translated into many languages and used around the world by libraries large and small (Landry et al., 2011). However, it has been both praised and criticized for over a century. Some reports have expressed the notion that LCSHs should be stopped (Bibliographic Services Task Force, 2005). It is claimed that it takes too long to train anyone to correctly apply the complex rules associated with LCSHs pre-coordinated subject strings; the specific terms and text strings are not understood by end-users or perhaps even reference librarians or cataloguers themselves; it is too slow to incorporate new terms (Anderson & Hofmann, 2006).

Earlier, criticisms (Kirtland & Cochrane, 1982) focused on the level of specificity: LCSHs is too specific and many of the specific terms are only used once in the library database and many are not – users would never guess or find the specific term. It is not rigorous enough for a true thesaurus, due to the form of headings and subdivision practices. It is too complicated and inconsistent, it has too many outdated terms, not keeping up with current terminologies, and has too many terms that reflect prejudice, and it requires an apprentice period to learn the principles properly, patterns, and rules, etc. In addition, the LCSHs terms are based on "literary warrant". Literary warrant has also been both praised and criticized i.e. praised for the ability to reflect the topics of materials being added to library collections and criticized for lacking a way for users to easily contribute to the terminology (Cataloging Policy and Support Office, 2007).

1.2.4 Emergence of new web technology

The emergence of Web 2.0 enhanced the idea to give chances to library users to assign their own keywords to describe their required library materials. Tim O'Reilly and John Battelle coined the term Web 2.0 in 2004 to describe the use of networks. Most notably the Internet and World Wide Web, as a platform where users can use, consume, and contribute data from multiple sources (O'Reilly, 2005; Kim & Abbas, 2010). O'Reilly also mentioned that 2.0 is not a version number in the same sense as in software releases, it is an expression meaning the current state of the art of web development.

In Web 2.0, the user is seen as a contributor when assigning some keywords to an object. It enhance response and interactivity between users and provide benefits to the community (Needleman, 2007; Ankolekar, Krötzsch, Tran, & Vrandečić, 2008). Community is an important part of Web 2.0. For example, popular social tagging sites like flickr.com, del.ico.us, youtube.com, LibraryThing.com, all have communities. At the same time, social tagging is done in an uncontrolled environment. End users do not need to master specific metadata standards or indexing rules for tagging (Lu, Park, & Hu, 2010). They apply their own keywords or phrase or descriptors to resources that is of interest to them.

For the successful implementation of social tagging in library systems, there is a need to understand how users assign social tags to library collections, what vocabulary they use and how far the social tags relates to a professional vocabulary. This understanding can

help libraries to make decisions, on how to implement social tagging and review the social tags.

1.3 Social tagging

Social tagging that is also known as collaborative tagging, social classification, and social indexing, allows end users to assign keywords to items. Typically, these items are web-based resources and the social tags become immediately available for others to see and use. Unlike traditional subject headings, social tagging keywords are usually freely chosen instead of using a controlled vocabulary (Tonkin et al., 2008; Arakji, Benbunan-Fich, & Koufaris, 2009). Social tagging is of interest to researchers because with a sufficiently large number of tags, useful terms will emerge that can either augment or even replace traditional subject terms. As a result, social tagging has created a renewed level of interest in manual indexing (Voß, 2007). For this study, social tagging has been considered as the act of tagging by the person consuming the information, and is done socially. It is not a direct collaboration between participants, but the result of this tagging is shared in the community and as a result, produces collective intelligence.

1.3.1 Social tagging practice

Since inception on the web in 2003 with the tagging system Del.icio.us, social tagging have become an enormously popular way to categorize large amount of information resources (Conradi, 2009). Social tags emerge from the aggregation of textual labels called tags that are affixed to digital objects of various formats within sites that allow for tagging. The social tagging resources can be of any type or in any format, such as web pages (e.g. del.ici.ous), videos (e.g. YouTube), photos (e.g. Flickr), academic papers (e.g. CiteULike), books (e.g. LibraryThing) and so on (Gabriela, 2009; Lu, Park, & Hu, 2010). Depending on the system, social tags are generated by the creator or owner of the content, or by the users of the content, or by a combination of the two (Smith, 2007).

Lambiotte & Ausloos (2006) described that the model of social tagging consist of three main components: users, tags and resources. This model has been useful in attempts to explain the relationships between aggregated tags, tagged resources and the community of users (Mika, 2007), also mentioned by Conradi (2009). Quintarelli (2005) discusses social tagging in terms of the 'Power Law distribution', stating that the power law reveals that

many people agree on using a few popular tags but also that smaller groups often prefer less known terms to describe their items of interest. Halpin & Place (2007) also showed that social tagging distributions tend to stabilize into power law distributions.

1.3.2 Criticism of social tagging

Social tagging have received criticism by those advocating top-down approaches to organize information (Peterson, 2006; Conradi, 2009). It is argued that the uncontrolled vocabulary of tags causes too many recall and precision problems to make them useful as information retrieval tools, and that the flat structure of folksonomies prevent users from seeing relationships between information items (Wetterstrom, 2008; Westcott, Chappell, & Lebel, 2009; Bates & Rowley, 2011). Social tagging is neither exclusive nor hierarchical, and therefore, in some circumstances, have an advantage over hierarchical taxonomies (Golder & Huberman, 2006). Simultaneously, Kipp & Campbell (2006) have exposed stable trends and patterns in which large user groups tag items. Therefore, analysis of social tags can consequently provide invaluable insight to information professionals on how precisely large groups of users view and describe digital information resources.

1.4 Expert created terms (Subject Headings)

A user cannot read what a user cannot find. End users commonly search by the subject of the book when a title or author search is not sufficient. Traditionally, libraries are using expert created subject headings, e.g. LCSHs, Sears List of Subject Headings (SLSHs) since date back to the 19th century. The main purpose of such a system is to ensure effective and efficient subject access to information (Landry, Bultrini, O'Neill, & Roe, 2011). The libraries and other memory institutions aimed to continually develop subject access tools and enabling users to find and discover information that will meet their search criteria. For this study, a subject heading is considered as a word or phrase from a controlled vocabulary, which is used to describe the subject of a document or a class of documents. It provides a complete indication of the subject covered in the book, including all of its various aspects and forms.

1.4.1 Function of expert created terms

The Subject heading is a human and intellectual endeavor, where trained professionals apply topic descriptions to items in their collections. It facilitates the uniform access and retrieval of items in any library in the world using the same search strategy (Kirtland & Cochrane, 1982; Wetterstrom, 2008; Broughton, 2011). Subject headings are applied to every item within a library collection, and facilitate user access to items in the catalogue that pertain to similar subject matter (Lopes & Beall, 1999; Chan, 2005). If users could only locate items by 'title' or other descriptive fields, such as 'author' or 'publisher', they would have to expend an enormous amount of time searching for items of related subject matter, and undoubtedly miss many items because of the ineffective and inefficient search facility.

1.4.2 Tools for the use of expert created terms

Naturally, every library may choose to categorize the subject matter of their items differently, without a uniform standard. There are several standard subject headings like Library of Congress Subject Headings (LCSHs). However, there are libraries where the use of LCSHs is not ideal or effective. To deal with special types of collections and user communities, other subject headings may be required. The United States National Library of Medicine developed Medical Subject Headings (MeSHs) to use for many health science databases and collection. Many university libraries may not apply both LCSHs and MeSHs to items. The National Library of Canada worked with LCSHs representatives to create a complementary set of Canadian Subject Headings (CSHs) to access and express the topic content of documents on Canada and Canadian topics. There are others like Sears list of Subject Heading (SLSHs). It made in response to the demands for a list of subject headings that were more suitable for the small public library rather than the LCSHs. However, it basically follows the form of the LCSHs (Islam, 2008; Miller & McCarthy, 2010 ; Broughton, 2011).

1.5 Purpose of the study

The purpose of the study was to investigate social tagging practice in science book context. In addition, it identified the usefulness of social tags as supplementary of controlled vocabulary to enhance the use of library resources. More specifically, this study examined to know to what extent the social tags match with controlled vocabulary, and

whether or not any additional perception is provided by social tags to improve the accessibility and information retrieval in a digital environment. For the successful implementation of social tagging in library systems, there is a need to understand how users assign social tags to library collections, what vocabularies they use, and how far the social tags relate to controlled vocabulary. This understanding can help libraries to decide on how to implement and review the social tagging.

1.6 Objectives of the study

The majority of the libraries is using controlled vocabulary to describe the subject of a document or classify the documents. On the other hand, social tagging encourages users' participation and collaboration in subject indexing and access. It is significant to consider how social tagging might be managed efficiently to improve the effectiveness of an library OPAC services (Voß, 2007). This study intended to focus on the following objectives as well:

- To identify the scope of social tags in comparing to expert created terms
- To investigate to what extent social tags can be used to enhance subject access to library collections in a digital environment
- To understand the benefits and limitations of social tagging for indexing and retrieval purposes
- To identify what other perceptions are reflected by social tags than subject description

1.7 Research questions

To achieve the objectives of this research the following research questions were framed:

i) To what extent does social tagging represent the same concepts as the expert created subject headings?

ii) Do social tags that match with the expert created subject headings receive higher frequency of use?

iii) Are the expert created subject headings highly ranked in the social tags' lists?

iv) What other perceptions are reflected by social tags than subject description?'

1.8 Scope of the study

The scope of this study is limited to the comparison of social tagging of books and assigned expert created terms on the same books in the Science genre. For collection of social tags the LibraryThings website has been used, and LCSHs has been considered as expert created subject headings. In addition, the Library of Congress OPAC has been consulted to gather MARC records for the sample books studied in this study.

1.9 Research methodology

The study focused on how closely social tags match with expert created terms and how much additional information it provides. The study examined the match and non-match of social tags with expert created terms (quantitative), and what other perceptions (qualitative) are reflected by social tags than subject description. Therefore, both the qualitative and quantitative approaches have been used for this study. Further details of research method have been discussed in chapter three.

Chapter II: Literature Review

2.1 Introduction

Since the last decade, social tagging has been applied in information organizations such as museums, libraries, and archives, which have developed services that facilitate user-community collaboration (Kakali & Papatheodorou, 2010). Many libraries have launched new online public access catalogues (OPACs) or web-based applications that incorporate social tagging. Trant (2006) mentioned that the Steve Collaboration, a group of art museums, were collectively exploring the role of social tagging to know what museum visitors see in works of art and what they judge as significant. Penn Museum and its archives is another example of social tagging adoption by a cultural heritage. Social tagging has grown in popularity. It encourages user participation and collaboration in subject indexing and access. As library, museum, archives are adopting social tagging, so many researches are going on to consider how tags might be managed efficiently to improve the effectiveness of an organization's services (Carla S., 2010; Olivier, 2011).

There have been researches on users' behavior in social tagging (Cattuto, Loreto, & Pietronero, 2007; Fugelstad et al., 2012; Hollenstein & Purves, 2012; Yang, Sun, Zhang, & Mei, 2012), the semantic value of the social tags (Tonkin et al., 2008; Hollenstein & Purves, 2012; Miotto & Orio, 2011), the automated development of semantic construction such as taxonomies and ontologies to organize social tags (Akther, Kim, Rawashdeh, & El Saddik, 2012; Da, Ke-qing, Rong, & Jian-xiao, 2011; Torres, Diaz, Skaf-Molli, & Molli, 2011), and the impact of social tagging on subject access and existing knowledge organization systems (Lin, Beaudoin, Bui, & Desai, 2006; Wetterstrom, 2008; Adler, 2009; Yi & Chan, 2009; Carman, 2009; Lu, Park, & Hu, 2010; Rolla, 2009).

This study has analyzed the potential and limitation of both social tagging and expert created vocabularies. In addition, how social tagging complements subject access as well as what other facets of information are provided by collaborative tagging have been considered. Therefore, the review of literature is aimed at gaining an insight of research done on social tagging and subject access. The review of literature was done through literature search of both print and electronic materials on the topic social tagging and subject access.

2.2 Inadequacy of controlled vocabulary

A controlled vocabulary typically offers a distinct level of specificity for item description (Peters, 2009). To ensure uniformity and universality in library catalogues and other information retrieval systems, controlled vocabularies are designed in a way so that locating information can be predicted and precise (Adler, 2009). Only subject cataloguers at the Library of Congress allow the process of creating and authorizing new headings. New headings are established on the basis of literary warrant, defined as "the use of an actual collection of materials or body of literature as the basis for developing an indexing or classification system" (Chan, 2005, p. 518). The LCSHs developed on the literary warrant, are strongly connected with Library of Congress collections. For over 100 years, the traditional subject access in catalogues has been enhanced using specialized controlled vocabularies. Simultaneously, it has been criticized for lack of up to date, biased language, and uncommon syndetic structure. Drabenstott, Simcox, & Fenton (1999) found that only in 36% of cases, users correctly interpret the meaning of the headings, and only 52%-55% librarians were able to correctly interpret the meaning of various subject headings. Though, LCSHs is widely used in libraries all over the world, it focused the principle of stability for the convenient of the user and this principle make it harder to fully erase bias and limits in the tools (Olson, 2000; Chan, 2005; McTavish, 2011). It is always time consuming and expensive to create and maintain professional metadata. Moreover, traditional cataloguing and indexing has been considered as a one-time operation. In addition, most institutions face the difficulty to provide adequate description for a large collection with diverse content (van Hooland, 2006). Stvilia & Joergensen (2010) suggested that the library catalogue should be always kept update in respect of users' expectations.

The library of congress policy states: "Establish a subject heading for a topic that represents a discrete, identifiable concept when it is first encountered in a work being catalogued, rather than after several works on the topic have been published and catalogued" (Library of Congress, 2008). Adler (2009) stated that the authorization of new terms is considered as beneficial to adding against the cost of implementing the changes, including adjusting existing bibliographical records. Antell & Huang (2008) found that the University of Oklahoma's students performed keyword searches fourteen times more often than subject searches in the OPAC. Moreover, for the last couple of years, many librarians doubted the value of the complicated pre-coordinated subject strings that made up an

LCSHs, where in the Web 2.0 environment users are more habituated to keyword search on the Internet (Rolla, 2009). The above studies indicated that due to the advent of interactive websites and search facilities, the expectations of users go beyond the traditional library OPACs. The LC working group on the future of bibliographic control (2008) recommended that adapting social tagging in the OPAC will make catalogues more relevant to users and definitely improve the access to library collections.

2.3 Overview of social tagging

In the Web 2.0 era, for recognition and management of resources social tagging has become a popular solution (Da et al., 2011). It has been popularized through the use on websites like Flickr and Delicious. (Garcia-Plaza, Zubiaga, Fresno, & Martinez, 2012) and now appears on thousands of sites–from business to e-commence. Even memory institutions like libraries, museums, archives that heavily relied on controlled vocabulary for their collections have begun to leverage social tagging in their OPAC (Bearman & Trant, 2005; Trant, 2009)

Apart from the organization of content for personal means, users of tagging systems are motivated by the idea of social contribution and the desire to share with others (Ames & Naaman, 2007; Hollenstein & Purves, 2012). Fugelstad et al. (2012) found that there are general volunteer motivations, pro-social behavioral history, and community-specific motivations that predict both the amount of use and specific types of activities users engaged in after joining the community. However, the majority of social tags represent informal metadata, they are neither structured nor correspond to a formal ontology (Tonkin et al., 2008). Moreover, social tagging can serve as a way of bridging the 'semantic gap' that often exists between a specialized or controlled vocabulary and the non-specialized language of users (Pirmann, 2011). Feinberg (2011) mentioned social tagging 'wisdom of the crowed', and state that

> "[The] combined knowledge of a group of people will be more accurate than the knowledge of any individual, even an expert individual. While the editor of a controlled vocabulary may miss a term that a particular user might associate with a concept, a wide user base constantly adding and applying terms will be more likely to include it. In addition, this broad user base will add new terms to the system quickly, bypassing the lag associated with formal vocabulary development." (p. 5)

It is true that all the social tagging practices do not describe an item (Yi, 2010). In addition to content-related tags, users also assigned tags relating to the use of an item (e.g., to read), ownership of an item (e.g. bought) etc. and such tags are not likely be of use to anyone other than the persons tagged the item (Golder & Huberman, 2006). Munk & Mork (2007) noted that users can see social tags assigned by other users and thus over time a consensus may emerge regarding the preferred terminology(ies) for a topic or concept. Whereas a small number of domain experts develops controlled vocabularies, social tagging actually benefit from large numbers of users being involved in their creation.

Most of the authors generally experienced that social tagging can supplement controlled vocabulary. Social tagging would permit patrons to personalize the library OPAC and foster online communities organized around the library. It can also help users feel more connected to the library's website.

2.4 Evaluation studies of social tagging

Social tagging has been quickly gaining ground because of its ability to recruit the activity of web users into effectively organizing and sharing vast amounts of information (Cattuto, Loreto, & Pietronero, 2007). Wetterstrom (2008) compared user-assigned tags on the general collection of the National Library of New Zealand with LCSHs. He found that 75% of tags did not match any subject headings and only 15% of tags have a match with a subject heading terms. On the other hand, Thomas et al. (2009) analyzed social tags assigned to the titles in LibraryThing and found that 35% of tags represented synonyms or related concepts that are not used in LCSHs. They also indicated that there is a notion that social tags have significant potential to enhance subject access. Yi & Chan (2009) compared social tags from Delicious with the LCSHs authority file, and found that in a set of 300 tags, approximately 61% has a complete-word match with a subject heading, and an additional 28% of tags were "very close in form" to at least one established subject heading.

Rolla (2009) showed that users' and cataloguers' approach descriptors very differently. For these differences, users' tags can enhance subject access to library materials, but they cannot entirely replace controlled vocabularies. Carman (2009) compared social tags from LibraryThing with LCSHs, and found that the most frequently used social tags were those that matched with the LCSHs, but there were a significant number of non-matching tags that offered useful additional information about a book. Conversely, Lu, Park, & Hu

(2010) claimed that there are possibilities to use social tags to improve the accessibility of library collections, but carefully mentioned that the existence of non-subject-related tags may obstruct the application of social tagging in traditional library cataloguing systems. Simultaneously, Kipp & Campbell (2010) compared user-assigned tags, author assigned keywords, and controlled vocabulary terms assigned by professional indexers to articles indexed in PubMed and tagged in CiteULike. They found that some tags and author-assigned keywords have matched exactly with the controlled vocabulary while others do not match but provides additional access points.

Kakali & Papatheodorou (2010) found that social tags express directly the evolution of a scientific domain. They stated that libraries should reconsider and evaluate the organizational schemes of subject indices, and to renew their content by adding new terms or relations, substitute the current subject headings with more appropriate ones. Bates & Rowley (2011) found that LibraryThing tags offer benefits over LCSHs, where LCSHs did not represent well about lesbian, gay, bisexual, transgender, and queer (LGBTQ) resources. Social tagging contain its own biases in worldview and subject representation, they added. Voorbij (2012) determined the percentage of recently published books provided with tags drawn from LibraryThing for Libraries. Random samples of 600 records were examined to determine whether they carry tags. The study found that about one third of the records were provided with tags; 80 percent of the tags are subject terms; 50 percent of the subject tags are covered by a keyword in the record; 25 percent are broader than a keyword and another 25 percent are related, narrower, or new. However, the author did not examine whether the tags properly reflect the content of the book or not.

The majority of the researchers generally showed a positive attitude toward social tagging. However, some of them also recognized the inherent limitations and problems of social tagging, while others indicated that libraries should consider allowing users to add social tags in OPACs.

2.5 Information retrieval with social tags

A large number of libraries are already using social tags as a way of information retrieval through catalogue search either by developing their own tagging systems like PennTags (http://tags.library.upenn.edu) at the University of Pennsylvania, MTagger (http:// www. lib.umich.edu/mtagger) at the University of Michigan (Srinivasan, Boast, Furner, & Becvar, 2009; Steele, 2009), or by inclusion of tagging feature in cataloguing system like

VuFind, WorldCat Local (Emanuel, 2011; Bertot et al., 2012). The use of social tagging can help in clustering similar resources tagged in similar ways. Golder & Huberman (2006) indicated that users show their interest by adding and categorizing new materials in tagging system. It helps library professionals to discover and improve content and services.

Hassan-Montero & Herrero-Solana (2006) found that users tend to assign general social tags to items that are more suitable for browsing rather than for very specific queries about the item. Furner (2007) stated that social tagging is easier to achieve indexer-searcher consistency which generally robust indicator of retrieval effectiveness. Morrison (2008) found that social tags were least effective for searching a specific item or queries requiring a short, factual answer. Furthermore, Good, Tennis, & Wilkinson (2009) stated that as the expanding pool of human annotators acting to fulfill wide range of purposes and in possession of a broad range of expertise, the social tags should prove useful to a wider community.

Kipp & Campbell (2010) studied on CiteULike and showed that users are interested to use social tags as an aid to discover more resources. Matthews et al. (2010) investigated ways of enhancing social tagging via knowledge organization systems. They found that knowledge organization systems that supplement social tagging have increased the effectiveness of non-specialist users (without training) in subject indexing. Kiu & Tsui (2011) found that social tagging is more useful for digital object that has less rich metadata such as maps, videos, pictures etc.

In spite of all criticism on social tagging, many researchers agreed that end users are indexers and searchers at the same time. Therefore, there is high possibility that users and professional indexer will agree on subject description of an item, and the combination of them will increase the availability of the resources in a more convenient way.

2.6 Social tags and subject access

Social tagging has potential to be used as a means of enhancing subject access of library materials (Lund & Washburn, 2009; Rolla, 2009; Thomas et al., 2009). Social tagging have added value to library or archive collections that have already been catalogued. The Library of Congress made a number of its photographic collections available on Flickr in 2008. The library was pleased to allow social tagging in a photo archive from the

1930-1940's (Vaughan, 2010). In New Zealand, the War Art Online website has also allowed users to create tags to help describing their photographic collection (Carman, 2009).

Several studies concerning the relationship of social tags to controlled vocabularies, especially with LCSHs has been done. Matusiak (2006) found that social tags could serve as a means of providing additional access points for images, although they likely cannot replace the depth of description found in controlled vocabularies. Smith (2007) studied LibraryThing tags and the subject headings assigned to a small sample of documents and found that the social tags identified precisely latent subjects. Moreover, LibaryThing tag and survey participants produced tags found better than the LCSHs used in public libraries (Weaver, 2007). Wetterstrom (2008) found that there were relatively few matches in between LCSHs and social tags, but social tags complemented the LCSHs by providing supplementary access point. Co-existence of social tags and LCSHs can serve as a bridge between them.

Thomas et al. (2009) found that there are 35% of social tags that represented synonyms or related concepts which are not mentioned in LCSHs. Carman (2009) remarked that addition of social tagging to catalogue records may be particularly valuable for items that have less meaningful subject headings, such as fiction and fantasy works. Social tagging is useful for describing of items like transgender or women studies which are not adequately described in LCSHs (Adler, 2009; Pirmann, 2011). On the other hand, Kakali & Papatheodorou (2010) confirmed that, the social tags enrich the subject description of documents. They found that a significant number of social tags were identical to authority records, but not used for the thematic description of the particular records. Some studies also look for the overlapping between social tagging and existing controlled vocabularies. For example, Lawson (2009) found that social tags have some overlap with already existing subject headings. Lu et al. (2010) indicated that there is 50% probability of a given LCSHs term being adopted by the end users as a social tag.

The above studies indicated that social tagging could potentially enhance searching, browsing, and locating information in library catalogues. The majority of the researchers agreed that several social tags constitute either new concepts or neologisms. There may be some overlap between controlled vocabularies and social tags, but a large number of users generated terms are not captured in traditional indexing languages as well. Moreover,

social tagging could help library professionals to approach the users' way of thinking and facilitate them more effectively as well as to observe the community's terminology evolution.

2.7 Research methods used by others

This study further focus on the research methods applied in previous researches where comparison between social tagging and controlled vocabulary has been studied. It is found that some researchers used only quantitative methods (Wetterstrom, 2008; Carman, 2009; Lu, Park, & Hu, 2010), some used qualitative methods (Yi, 2008; Bates & Rowley, 2011), while others used both qualitative and quantitative methods (Adler, 2009; Lu, Park, & Hu, 2010; Pirmann, 2011; Rolla, 2009). In addition, it is observed that several researchers like Voorbij (1998), Kipp (2006), Wetterstrom (2008), Carman (2009), Thomas, Caudle, & Schmitz (2009), have used different kinds of code or scale for categorization of social tags.

Voorbij (1998) designed a seven point scale to compare subject descriptors and title keywords based on their degree of similarity to each other, like 1 = Descriptor is exactly or almost the same as word from title; 2 = Descriptor is synonym of word from title; 3 = Descriptor is broader than word from title; 4 = Descriptor is narrower than word from title; 5 = Descriptor is related to word from title; 6 = Descriptor has a certain relation to word from title, but it is difficult to distinguish between 2, 3, 4 and 5; and 7. Descriptor does not appear in title at all. He determined the degree of match using the hierarchical relationships of a thesaurus.

Kipp (2006) modified Voorbij's scale as 1 = exact matches; 2 = synonyms; 3 = broader terms; 4 = narrower terms; 5 = related terms; 6 = terms with an undefined relationship; and 7 = terms that were not related at all, to compare degree of overlap between tags and subject headings.

Later Thomas, Caudle, & Schmitz (2009), modified Voorbij and Kipp scale as 1 = Same, 2 = Synonym, 3 = Natural Language Synonym, 4 = Broader term., 5 = Narrower term, 6 = Related term, 7 = LCSH not assigned, 8 = Related, and 9 = Not Related to see which extent social tags replicate the LCSHs. They compared the social tags and LCSHs for ten popular books in areas where LCSHs is weak.

Wetterstrom (2008) developed three broad categories like Match (A), Partial match (B) and No match (C). Partial match was sub-categorized as cross-reference (CR), Spelling variation (SV) and Tag appears in subdivision (SD), where No matched was subcategorized as Broader term (BT), Narrower term (NT), Related term or different point of view (RT), American vs NZ vocabulary (NZV), Currency of term (CoT), and Popular language (PL). He investigated the complementarity of user-assigned tags and LCSHs assigned by cataloguers in a New Zealand library context.

Carman (2009) modified Wetterstrom's coding system in five broad categories like Match with LCSHs (A), Non-subject Match with OCLC record (B) Partial match (C), No match (D), No match between tag and LC subject heading (E). Further the 'Non-subject Match with OCLC record' categories was subcategorized as Author/Title information (B-AT), Publishing details (B-PD) Format Information (B-FI). Cross-reference (C-CR), Spelling variation (C-SV), Tags appear in sub-division (C-SD), Related term or different point of view (C-RT), Popular language (C-PL) were subcategories of Partial match. No match (D) was subcategorized as Plot details (D-PLo), Format Information (D-FI). Literary Genre Information (D-LG), Place and character names (D-PC), Other non-specific info (D-BT), User specific (D-US), Unclear (D-UC). He considered social tags on Science Fiction and Fantasy works that presence in LibraryThing, and compared with LCSHs.

It is found that the coding system is an extension of scaling system, but serves the same purpose. Both the systems have been developed, and modified by others to compare the users assigned terms with controlled vocabulary. Such coding or scaling helped to pull out all the social tags under defined categories by the researchers for respective studies. Moreover, it helped them to compare the tags with controlled vocabularies and showed match and non-match between these two with ease.

2.8 Conclusion

From the above literature review, it is found that the limitation of controlled vocabularies has been identified in early 1980s. More limitations of controlled vocabularies were noticed due to the adoption of information technology as well as introducing online catalogues for the libraries. These limitations were more focused while keyword based searching made available in the Internet. The researchers were looking for more feasible solutions to reduce the gap between user expectations and expert created subject terms. The emergence of web 2.0 technologies facilitated the use of social tagging for libraries.

Several studies on evaluation of the use and application of social tags have been done and indicated social tagging as a tool to overcome the shortcomings of controlled vocabularies. The advantages of social tagging have been identified as the terms that originate from the user base and reflect the users' natural language. In addition, social tagging allows users to immediately assign terms to an information object based on their own knowledge. Contrariwise, the use of controlled vocabulary often involves expertise to understand and use the same.

Social tagging also received criticism. In addition to content-related tags, there is the presence of many personal assigned social tags, which would not be of use to anyone other than the person assigned. Despite such criticism, some researchers have demonstrated the potential of social tagging to be used as a means of enhancing subject access to materials in libraries, archives, and museums. Moreover, several studies on information retrieval indicated that social tagging could increase frequency of use of items that are traditionally described using controlled vocabularies.

It is evident from the literature review that researches on social tagging have been conducted from a different point of view. Comparison of social tagging with controlled vocabulary is one of them. It is found that researchers, who made comparisons between these two, either took the genres and audiences in general or focused on a specific genre – fiction, fantasy work, non-fiction or specific topics like transgender, gay, lesbian, etc. Moreover, those studies compared the matching of social tagging with controlled vocabulary at the collection level and considered the tags cloud as a whole.

According to literature consulted in this study, it is found that almost no study has been conducted in academic and non-academic books, especially, science and research oriented genre like Science, Social Science, Applied Science, etc., Moreover, the appropriateness of social tags to the assigned items and comparison with the assigned subject heading of that book has not been studied yet. In addition, many researchers examined the enhancement of subject access with social tagging. Among those, some indicated about the personalization of social tags or use of personal tags rather than a subject descriptor. Therefore, one more unexplored factor has been identified, i.e. what other perceptions do social tags in addition to subject descriptions reflect. The literature review encouraged that

it could be very significant to make a comparative study between social tagging and expert created subject terms of the science genre to see how far social tagging represents the same concepts as the expert created subject headings, and further to look at the differences between academic and non-academic books of the same genre. In addition, it needs to examine what kind of perceptions reflected by social tags than subject description. Both qualitative and qualitative approaches should be applied for such study. Use of a modified version of previous authors' coding practice might help to fulfill the purpose of the study.

Chapter III: Research Design and Methodology

3.1 Introduction

The methodology followed in this study is a combination of both qualitative and quantitative approaches. Quantitative research represents the human experience in numerical categories, sometimes referred as statistics. On the other hand, qualitative research provides detailed description and analysis of the quality, or the substance of the human experience (Marvasti, 2003, Christensen, Johnson, & Turner, 2010). Sometimes one research question (or subsidiary question) demands quantitative data while another requires qualitative data; in other cases both qualitative and quantitative data may be required to answer one research question (Creswell, 2008; Bob Matthews & Ross, 2010). In addition, combination of both methodologies provide an in depth look at context, processes, and interactions and precise measurement of outcomes (Lodico, Spaulding, & Voegtle, 2010). The focus of this study was how closely social tags match with expert created subject terms, and what other perceptions provided by social tags. This study found that the most practical way to answer the research questions would be to include a combination of a qualitative and a quantitative research approach. To identify the match and non-match of social tags with expert created terms quantitative data is required. Simultaneously, qualitative analysis is required to identify what kind of supplementary information provided in addition to subject description by social tags.

3.2 The Research sites

The review of related literature in the previous chapter showed that there was no harmony among the researchers regarding genres and audience to be included in assessing social tagging in comparison with LCSHs. Since this study is meant as an initial venture into this arena, it initially examined a small number of books. Two criteria have been used for choosing the social tagging system. The first criterion was concerned about the objects of the tags in the tagging system. It was desirable to find a tagging system in which the objects of the tags were books. Conradi (2009) mentioned that there are websites that allow users to assign social tags to books including the online sales company like Amazon (http://www.amazon.com), university library project like PennTags (http://tags.library. upenn.edu), social book cataloguing site LibraryThing (http://www.librarything.com),

aNobii (http:// www. anobii.com), Goodreads (http://www.goodreads.com, tags are called shelves), Shelfari (http://www.shelfari.com) and WeRead (http://weread.com) etc.

The second criterion of the selection process concerned the amount of the social tags. Tim Spalding (2007) found that LibraryThing users generate ten times more tags per book than Amazon users do, even though LibraryThing receives ten times less traffic. Spalding concluded that users assigned social tags to books for their own purposes, and has no real incentive for a commercial entity. He also remarked that social tagging works well when users do it for their own needs, but it fails when they are asked to do it for someone else. "To do anything useful with tags, you need numbers …[and] with a larger number of tags, clear patterns emerge", Spalding mentioned, summarizing the significance of large quantity of social tags (Conradi, 2009; Thomas et al., 2009). As of March 2012, LibraryThing has over 1.53 million users, and more than 72 million catalogued books, and more than 87 million social tags have been added (source http://www.librarything.com/ zeitgeist). These statistical data indicates LibraryThing as a larger social book-marking site than all of the above-mentioned sites. Therefore, for this study the LibraryThing website has been selected as research site to collect social tags.

The Library of Congress Subject Headings is a controlled vocabulary for representing the subject and form of the books and serials in the Library of Congress collection, with the purpose of providing subject access points to the bibliographic records contained in the Library of Congress catalogues. It has become a tool for subject indexing of library catalogues in general as an increasing number of libraries all around the world have adopted LCSHs system (Library of Congress Authorities, 2011). The expert from a carefully selected list of words and phrases chooses the subject headings terms. Library professionals summarize the subject content of the resource, and then match the subject content with the best subject representation captured in LCSHs. The extensive use and acceptance of the LCSHs facilitates the uniform access and retrieval of items in libraries using the same search strategy. if the correct headings have been applied to the item by the library (Lu, Park, & Hu, 2010). This study has chosen LCSHs as controlled vocabulary since the terms have been created by an expert authority, as well as the Library of Congress Online Public Access Catalogue (OPAC) is using it.

3.3 LibraryThing

Designed as a social cataloguing website, LibraryThing allows users to assign descriptive metadata to books in the form of tags. Tim Spalding, a freelance web designer with experience in the publishing industry, founded LibraryThing in August 2005. Although it began purely as an online tool with which individuals could catalogue their personal libraries, the site has been quickly developed into a full-fledged social network. It allows members to share book recommendations, compare collections, post book reviews, comment on one another's libraries and reviews, contribute to discussions, participate in author chats, and indulge in other matters of a bookish nature. The most frequently reviewed book is Stephenie Meyer's Twilight; the author with the largest number of copies in members' personal libraries is J. K. Rowling. The section labeled 'Zeitgeist' discloses these useful tidbits (Conradi, 2009; Johnson, 2010).
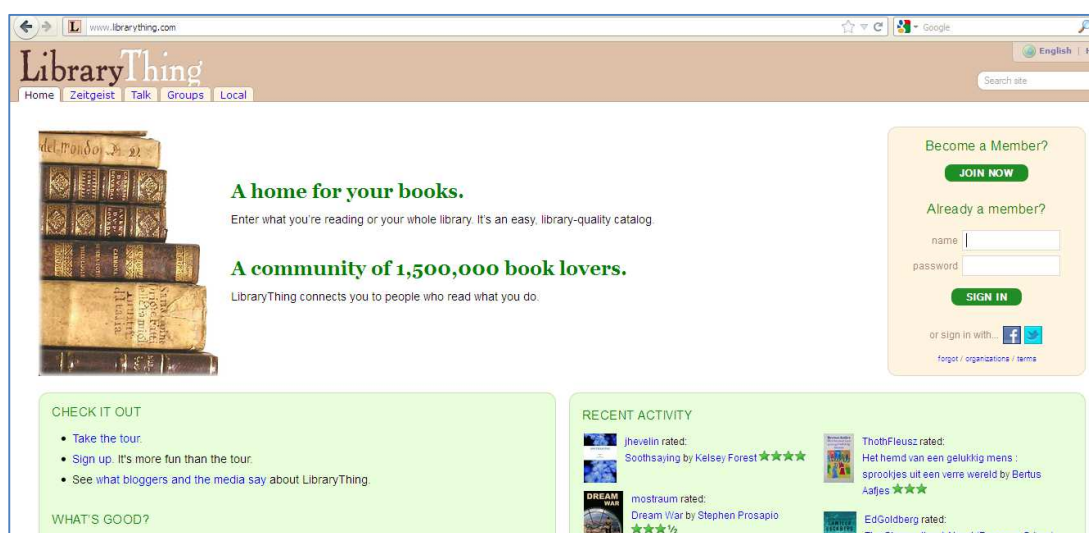

Figure 1: Screenshot of the LibrayThing home page

The cataloguing feature, the heart of the site, is both straightforward and easy to use. From the "Add Books" page, users can enter details such as title, author, ISBN, or LC card number, add whatever descriptive social tags they would like to assign to a book. At the same time, users can choose their copy cataloguing source from amazon.com, Library of Congress, and amazon UK etc for example, there are 690 international sources in total. Click on a title in the list of search results will add the book to one's (members only) collection. LibraryThing will indicate immediately whether the book is a duplicate entry or not. It is useful for the end users who own many books, but face difficulty to remember

them all. Moreover, it shows a list of how many other members own copies of the same book.

Existing books in the system can be searched by title, ISBN, author, and social tag. For each book, additional information is provided with bibliographic details, alternate titles, editions available, users average rating, links to reviews, links to other sites like Amazon, WorldCat, Google Books etc., and most importantly social tags that others end users already assigned to describe the book. Click on any one of these social tags will bring up other books that fit with the same description. LibraryThing cannot be beaten for its user-friendliness, versatility, and literary enthusiasm. It is both fast and readily accessible from any web browser or mobile device (Conradi, 2009; Bates & Rowley, 2011).

Figure 2: Screen shot social tags occurrences in LibraryThing

The above figure shows excerpts of a book's social tags occurrences. The social tags are presented in tag clouds. The most popular social tags are displayed in larger font than the less popular ones. In LibraryThing, each book has a "social information" page that displays basic publication information about the book, reviews, and a social tag cloud that enumerates the most commonly used social tags. These social tags function as a discovery mechanism through which users can locate other titles that have been assigned the same social tag.

3.4 Library of Congress Subject Headings

Since 1898, LCSHs has provided a set of terms for a comprehensive, broad range of topics. LCSHs has been translated into many languages and is used around the world by libraries large and small (Anderson & Hofmann, 2006). Since 2009, the LCSHs has also been available online (http://id.loc.gov/search). It is an important guide to find the keywords to use in the subject search option in the library OPAC. By consulting the LCSHs, one can often save time, and locate useful subject terms that described the subject of the book. Broughton (2011) described that a subject heading may consist of one or more words. A one-word heading represents a single concept, whereas a multiple-word heading may represent either a single concept or multiple concepts.

Chan (2005) indicated that three types of relationships are represented in the cross-reference structure of LCSHs i.e. equivalence, hierarchical, and associative. These relationships are expressed in terms of USE, UF (Used for), BT (Broader term), NT (Narrower term), RT (Related term), and SA (See also) references. Each reference links a term or heading with another heading or with a group of headings USE references are made from unauthorized or non-preferred terms to authorize or valid headings. On the other hand, in the form of UF (Used-for), references are made under the valid headings.

Headings related hierarchically are connected by means of reciprocal BT (Broader term) and NT (Narrower term) references. A heading is linked to the level immediately above it and the level immediately below it in the appropriate hierarchy. Types of hierarchical relationships include Genus/species (or class/class member), Whole/part, and Instance (or generic topic/proper name). Under each valid heading, other headings representing concepts on a level immediately above in the hierarchy are listed as BT (Broader term), except when the heading in question represents the "top term" in the hierarchy, or when the broader term cannot be readily identified. Headings related in some manner other than by hierarchy are linked with RT (Related term) references (Stone, 2000; Library of Congress, 2008).

The LCSHs subject term 'Sports' has been taken here as an example for further clarification of the above discussion. Headings in boldface may be used for subject searching. For example, 'Sports' may be used as a subject heading. *May Subd Geog or Not*

Subd Geog indicate whether or not the subject heading may be subdivided geographically. Subdivided subject headings may be topical, by form, chronological, or geographic.

```
Sports (May Subd Geog)
   UF Field sports
      Pastimes
      Recreations
   BT Recreation
   RT Athletics
      Games
      Outdoor life
      Physical education and training
   SA subdivision Sports under military
         services, e.g., United States. Army
         --Sports; and under ethnic groups
   NA Aeronautical sports
      Age and sports
      Aquatic sports
      Ball games
```

UF = Used For, refers to related subject headings. In this example, 'Field Sports', or 'Past time' or 'Recreations'.

BT = Broader Term. A subject of broader scope which may include some or all aspects of 'Recreation'

RT = Related Topic. Indicates other associated headings, e.g.: 'Athletics'

SA = See Also; indicates other useful terms for searching the subject.

NT = Narrower Topic; indicates more specific headings than the boldface subject heading.

-- (Dash); indicates a subdivision of the main (boldface) subject heading. There will be more information in the catalogue listed under this subdivision. This is useful for narrowing a topic.

Many names of places and people (i.e., proper nouns) are not listed in LCSHs, however, they may be used as subject headings.

3.5 The Sample

The study selected twenty books from the science genre. It is felt that the information value of social tags would be better to judge from well-known genre. The sample has been further considered to include only those books that have also been catalogued in the Library of Congress catalogue. The sampling procedure was purposive, as this study want to explore the research questions in depth. The sample size was limited to twenty books due to limited time and resources available.

On the LibraryThing site, first a search has been done with the keyword 'Science'. It provided the search results in various categories like 'works', 'authors', 'series', 'tags', 'common knowledge', 'classification', 'social group', 'publisher information', etc. Among those options, the 'tags' has been selected and found 'Tag search: Science' automatically,

as the first keyword search was 'Science'. From the results of this stage only the social tag 'Science' has been chosen. Purposefully, more specific tags like: mathematics, biology, chemistry, etc. have been chosen to select books from the whole science genre.

3.6 Sample selection

The selection of books has been done by the criteria that the book should be classified within 500 – 599 according to Dewey Decimal Classification (DDC), and should contain at least 150 social tags. The chosen book's ISBN has been checked in the library of congress OPAC to see whether the same book is available there. It was very important to get the same book in both the place (LibraryThing and OPAC of the library of congress) to compare the same. Among the twenty books (Appendix - I), carefully ten books have been chosen from the academic group and the reaming from the non-academic group. To identify the academic and non-academic books, the DDC class and assigned subject headings in the library of congress catalogue have been considered. Each book has been assigned a code like 'Book 1', 'Book 2', 'Book 3' etc to minimize the repetition of the title for further analysis and keeping track of the same.

3.7 Data collection methods

For collecting and recording social tags, each selected books' page on LibraryThing has been navigated. The entire social tags cloud including the frequency of use was harvested by copy into a text editor to remove the hyperlinks from the social tags. Then all the text was converted into a single font and same size. Each social tag and its accompanying frequency of use were spaced out in respective lines in plain-text document and tag delimited. The social tags for each book have been transferred into an MS Excel Spreadsheet.

In addition, a screenshot of the page for each book on LibraryThing was taken and converted as a PDF. The LibraryThing records are subject to change over time due to alterations made by the End users. PDFs were made to ensure that the researcher could always refer back to the records, as they existed during the data collection process. The same book's ISBN has been searched in the library of congress catalogue. While the catalogue displayed bibliographical information the 'Full record' and 'MARC tags' has been printed one after another as a PDF. Both files have been preserved to supplement each other for clear understanding.

This study took into consideration only those social tags that occurred at least twice. It means at least two people assigned the same social tags. Furthermore, spelling variation or wrong spelling, singular or plural form of the social tags word has not been tried to fix. The reason behind keeping them intact is to see further if there is a possibility to automatic adaptability of social tags in support of the library catalogue or not. After that, all the social tags are organized in descending order based on their frequency of use. The one time occurred social tags have not been considered for further analysis. It should be mentioned that all the social tags for a single book has been kept isolated from other books, as the study exclusively considered the appropriateness of social tags of that particular book.

The MARC record has been considered to determine LCSHs term assigned to each book. The subject entries of Field 650 in all MARC records conform to LCSHs. Expert cataloguers based on LCSHs terms, assign the subject terms included in Field 650.

Figure 3: MARC Tags in Library of Congress OPAC

LCSHs has subject headings and subdivisions. When encoded in MARC, the LCSHs subject headings and subdivisions are structured into several subfields of 650. The terms containing several subfields together form a full subject heading. For instance, the MARC record (shown in Figure III) has four 650 fields, representing four subject headings which are composed of the terms contained in their subfields. In this study, the LCSHs terms

contained in each subfield have been treated as separate keywords. Social tags were compared with these separate subject terms instead of the full complex subject headings.

3.8 Coding methods

After completion of the collection of data, a tool was required to make comparison and analyze social tags with LCSHs terms. There is a need to use of code that should help to pull together all the similar concepts for further analysis. The objective of the coding process was to make sense out of Social tags. It was an inductive process of categorization of social tags into a few themes. An initial evaluation has been made to find out what kind of information the social tags reflect. A pilot study with two books was conducted to confirm the feasibility of the coding method. After careful observation, it is found that the social tags fell into four broad categories: Social tags that match exactly with LCSHs, Social tags that match partially with LCSHs, Social tags that reflect bibliographic information and social tags that are user specific information. The last three categories were further sub-categorized.

In literature review, it is found that several authors like Voorbij (1998), Kipp (2006), Wetterstrom (2008), Carman (2009), Thomas, Caudle, & Schmitz (2009) have used different kinds of code for categorization of tags (Discussed in the section 2.7). Based on those literatures, a modified coding system has been developed for this study. A letter designated each category in the coding system. The categories were then given additional one/two letters (e.g.: BT) to further sub-categorize. The coded categories have been depicted below:

Code   **Explanation**

A   **Exact match with LCSH**
          The tag is an exact match with the LCSH

  A-A   **Exact match**
              The social tag which is the same as assigned in LCSHs (headings in
              boldface)

B   **Partial match with LCSH**
          The social tag that appeared in any of the following sub-categories of the exact
          match LCSH term of that particular book

  B-S   **Synonyms (Used for)**
              The social tag matches an established "used for" (UF) term in LCSHs and
              refers to related subject headings

B-BT   **Broader term**
The social tag matches with the broader term in LCSHs and refers to more general subject headings.

B-NT   **Narrower term**
The social tag matches with the narrower term in LCSHs and refers to more specific headings than the boldface heading.

B-RT   **Related term**
The social tag matches with related terms in LCSHs and provide ideas of other topics to investigate.

B-SD   **Subdivision/See also**
The social tag matches either with See also term in LCSHs that refers to other ways of looking up the same topic, or appeared with -- (a dash) refers to a subdivision of the boldface subject heading for narrowing the topic.

C   **Bibliographical information**
The social tag matches with bibliographical information available in the Library of Congress catalogue record other than LCSH.

C-AT   **Author/Title**
The social tag represents the author/s or title of the book, e.g. 'Halliday', '$E=mc^2$' etc.

C-PD   **Publishing details**
The social tag provides publishing information, e.g. date of publication, publisher etc.

C-FI   **Format information**
The social tags represent the availability of a variety of formats of the book, e.g. Paperback, E-book etc.

C-LP   **Language of publication**
The social tags represent that the book has been published in other languages in addition to English, e.g. Spanish, German etc.

D   **User specific information**
The social tag that has not been felt in any of the above mentioned sub-categories, but provides meaningful information related to user or use specific and go well with any of the sub-categories.

D-UC   **User community**
The social tag represents the user community of the book, e.g. bachelor level, university, school etc.

D-OL   **Tags in non-English**
The social tags that appeared in a language other than English e.g. Lumière (French) mean Light (English)

D-BI **Background information**
The social tag represents background information of author/s, e.g. Nobel prize winner, British etc.

D-RC **Related concept**
The tag represents related concept but did not with match LCSH, e.g. 'Mathematics' has been tagged as 'number theory'

D-NR **Not related**
The social tag is too vague to express the subject of the book, e.g. A biology book has received a social tag 'Astronomy'.

D-PC **Partial topic or chapter**
The social tag represents a partial topic or chapter discussed in the book.

D-US **User or use specific**
The social tags that are not useful descriptor, and do not provide any access to the book, but represent to particular users or use specific information, e.g.: recommended', 'science dept', 'we have 3different editions' etc.

D-NS **Unclear/ Non-specific**
The tag is cryptic or difficult to interpret or does not fit into any category, e.g. '@', '#105', 'wormholes' etc.

3.9 Coding of social tags

Each social tag was first compared with the LCSHs term of the books assigned in the library of congress OPAC. If matched, then it fell in the Category A (exact match). If not, then it has been considered for the category B (partial match), and LCSHs tool has been carefully consulted to identify the social tags with Used for, Broader term, Narrower term, Related term, Subdivision/See also etc, assigned under LCSHs terms for the book. If the social tag has not matched either category A (exact match) or B (partial match), then the appropriateness of those social tags with the sub-categories of C (bibliographic information) and D (user specific information) have been considered.

The Library of Congress OPAC and the book both have been consulted to identify the match of 'author/title (C-AT)', 'publishing details (C-PD)'. To determine 'format information (C-FI)' and 'language of publication (C-LP)', amazon.com and WorldCat.org have been consulted for further clarification. Under the broad category D (user specific information) the sub-category 'Tags in non-English (D-OL)' has been identified and translated in English with the Google translator service to understand the meaning. The

social tags that referred to the sub-category 'Background Information (D-BI)' has been checked with the author's biography, book reviews, etc. The social tags that fell in the sub-category 'Partial topic or chapter (D-PC)' has been consulted with the content page of the book. As the coding was the heart of this study, the supervisor of the study carefully observed and double checked the sample of categorization made by the researcher.

3.10 Process of code analysis

This study has analyzed social tags comparing to LCSHs terms to identify how many of them matched exactly or partially. Furthermore, it also investigates what other supplementary information was provided by the social tags in addition to subject descriptive terms. In both cases, the social tags were considered with the appropriateness to the specific book. After assigning codes for each social tag, the 'count of tags' and the 'sum of times appears' have been calculated from pivot table of MS Excel for each book. It has been done for further analysis of the social tags of respective book.

	Values	
Row Labels	Count of Tags	Sum of Times Appears
C-AT	2	6
D-NS	6	13
A-A	3	89
C-FI	2	5
D-PC	4	71
D-CI	2	4
D-NR	10	137
D-US	14	201
B-BT	2	382
D-RC	4	98
D-OL	6	24
Grand Total	55	1030

	Values	
Row Labels	Count of Tags	Sum of Times Appears
⊟ A-A	3	89
Einstein	1	84
Einstein Albert	1	3
mathematical physics	1	2
⊟ C-FI	2	5
audio	1	2
hardcover	1	3
⊟ D-PC	4	71
biography	1	64
force	1	2
Mass	1	3
nuclear physics	1	2
⊟ D-CI	2	4
American	1	2
engineering	1	2
⊟ D-NR	10	137
astronomy	1	2
Biography/Autobiography	1	2
biology	1	2

Figure 4: Snapshot of pivot table of code analysis for social tags

The figure four shows that how code practice pull all the similar social tags together for each book, and how each social tag has been counted under each sub-category.

Later the entire social tags of twenty books have been assembled together in one Spreadsheet. Again, the 'count of social tags' and the 'sum of times appears' has been calculated by pivot table of the MS Excel for the total sample.

Values		
Row Labels ▾	Count of Tags	Sum of Times Appears
A-A	50	4078
B-BT	38	9164
B-NT	18	141
B-RT	4	209
B-S	5	88
B-SD	3	9
C-AT	17	208
C-FI	59	285
C-L	11	38
C-PD	30	98
D-CI	14	75
D-NR	94	983
D-NS	413	1427
D-OL	65	265
D-PC	118	2091
D-RC	47	618
D-UC	13	32
D-US	366	5336
Grand Total	1365	25145

Figure 5: Snapshot of pivot table of code analysis for twenty books

Afterward, the result of the pivot table has been transferred to the table 2, and percentage of each sub-category has been calculated. The social tags under each sub-category have been further investigated according to the research questions of this study.

3.11 Data presentation

MS Excel, and SPSS (Statistical Package for the Social Sciences) software were used to get an accurate analysis of the related data. Pearson's coefficient of correlation has been calculated in related cases. In this study three tables, twenty graphical presentation has been produced using MS Excel to explain the analysis.

3.12 Ethical considerations

The main ethical considerations in this study were connected with keeping all the social tags intact as collected from the LibraryThing site. Before assigning the code to each social tag, the library of Congress OPAC, Library of Congress Subject Headings, book review, preface, and content page of the book, amazon.com, WorldCat.org and also related websites for the information about the authors of the book have been consulted. The set criteria for choosing books as a sample have been strictly followed to avoid personal favor to any particular book or author or subject.

Chapter IV: Analysis and Findings

4.1 Data analysis process

In this study, the social tags and LCSHs terms were analyzed at the book level. First the assigned LCSHs terms of each book has been figured out in LCSHs volumes. Later, each of the social tags of that particular's book has been analyzed according to coding practice mentioned in the methodology chapter (Discussed in the section 3.8). It should be clear that a LCSHs term can be a subject heading, but the same term available as either 'use for'/'broader term'/'narrower term'/'related term'/ 'subdivision' under another subject has not been considered. The only syntactic form of the terms has been considered. A social tag and a LCSHs term were considered equivalent only if they were syntactically identical, e.g.: the tag 'physics' and the LCSHs term 'Physics' are considered equivalent, but the social tag 'Ecology' and the LCSH term 'Food chains' were considered as two different terms even though shared the core semantic meaning.

4.2 Social tags, and LCSHs terms

The attention-grabbing difference between social tags and library-cataloguer assigned terms was that users assigned more social tags to books than the cataloguers assigned LCSHs, Golder & Huberman (2006), Rolla (2009) and Lu, Park, & Hu (2010) also remarked in their studies. This study found that in total, the considered 20 books has been assigned 55 LCSHs terms while the same books received 6,600 social tags.
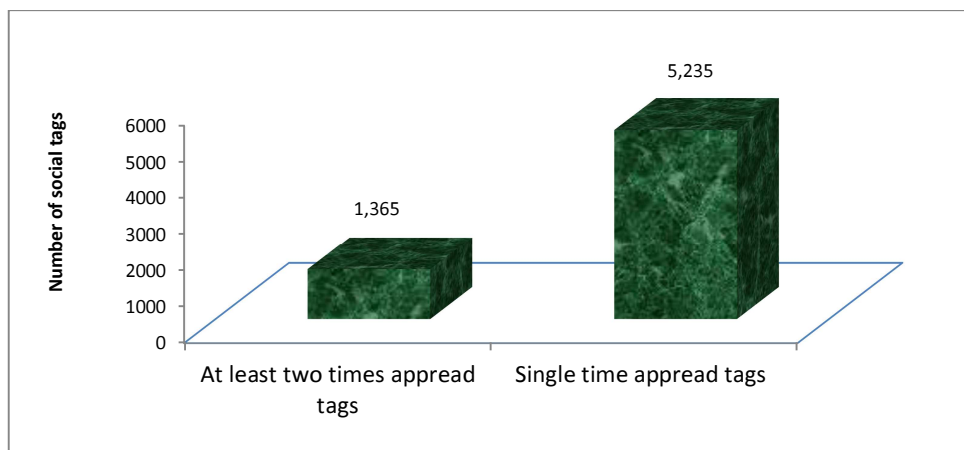
Figure 6: Distribution of social tags

The above figure shows that there were 5,235 (79%) social tags that appeared only once, and these were considered as single time used tags. These single time assigned social tags occurred due to either spelling variation or due to extreme personalization, and the same social tags have not been used by others. This study excluded those single times occurred social tags to keep the consistency of comparison with the limited amount of LCSHs terms. Therefore, it is found that there were 1,365 (21%) social tags that appeared at least twice, and these multiple times assigned social tags has been considered in this study. It is also mentionable that on an average 2.75 LCSHs terms has been assigned by library cataloguers, whereas on an average 68.25 social tags (excluding single time assigned tags) have been assigned by the end users to the selected books.

4.3 Number of social tags occurrence

In total, 30,380 social tags have been found. The figure below shows that social tags that appear multiple times have been used 25,145 times by the end users, while obviously single time (5, 235) assigned social tags remain the same.
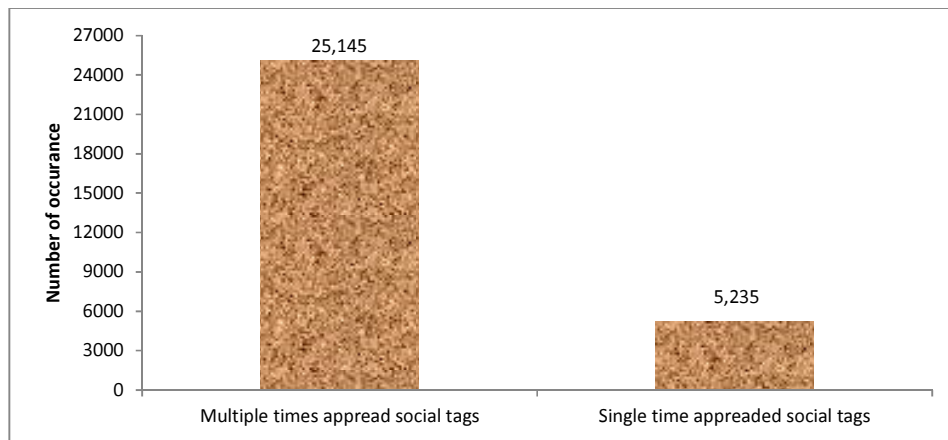
Figure 7: Number of social tags occurrence

The social tags that are assigned multiple times are less frequent in the count of unique social tags than those appear only once (Discussed in section 4.2). It is obvious that the more times the same social tags has been assigned, the greater the indication of collaborative value (Wetterstrom, 2008).

4.4 Distribution of tag's occurrences per book

The percentage of tag's distribution for each book has been calculated. The figure below shows that the book 4 received the highest number (8,949; 35%), while the book 12 received the lowest number (194; 0.77%) of social tags.
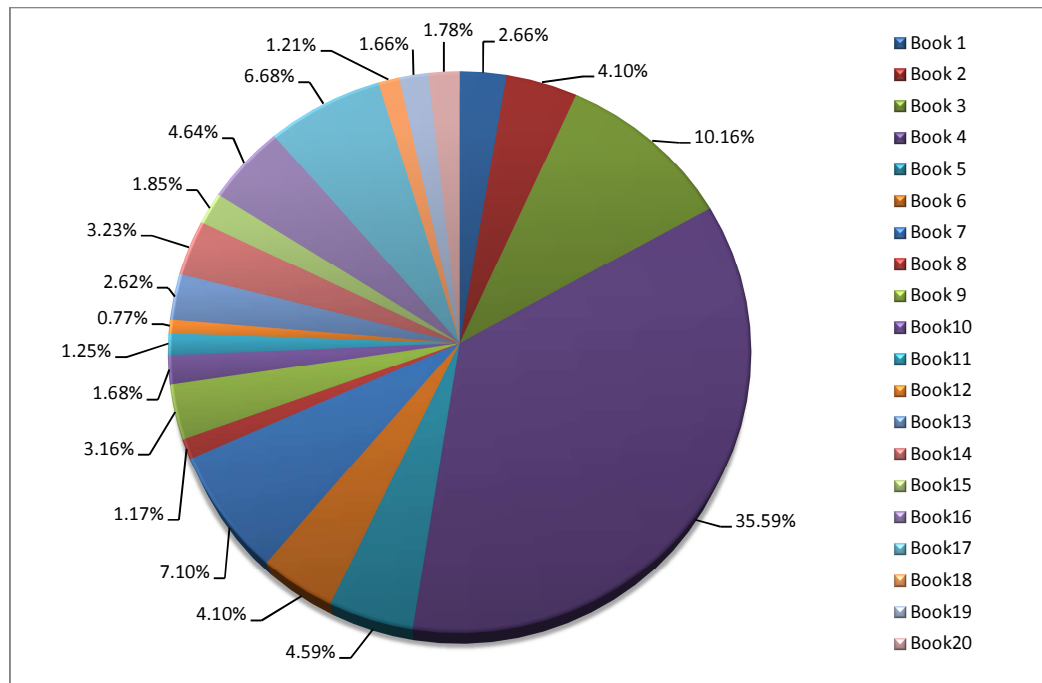
Figure 8: Distribution of tag's occurrences per book

It also shows that other than book 3 and 4, all other books have received less than a double figure percentage. It is apparent that the majority of the books individually has received less than four percent of the total number of social tags.

4.5 Academic books versus non-academic books

As mentioned in methodology, this study considered 20 books as a purposive sample. Among those, the first 10 books have been selected from non-academic, and the rest was from the academic book point of view. There is a significant difference between these two groups of books considering the number of occurrences of social tags. It is found that non-academic books received a majority of social tags (18,685; 74%) than academic books (6,460; 26%). Non-academic books received nearly three-times more social tags than academic books. The cataloguers assigned 27 (49%) LCSHs terms to non-academic books, and 28 (51%) terms to the academic books. The numbers of assigned LCSHs terms have no major impact upon the number of social tags occurrence between these two

groups. Therefore, it is marked that the end users assigned comparatively more social tags in non-academic books than academic books.

Table 1: Distribution of LCSHs terms and Social tags in each book group

	Non-academic books			Academic books	
Book Code	Number of LCSH terms	Number of Occurrence of Social tags	Book Code	Number of LCSH terms	Number of Occurrence of Social tags
Book 1	5	669 (3.58%)	Book11	1	314 (4.86%)
Book 2	4	1030 (5.51%)	Book12	1	194 (3.00%)
Book 3	2	2,554 (13.67%)	Book13	1	659 (10.20%)
Book 4	1	8,949 (47.89%)	Book14	6	811 (12.55%)
Book 5	1	1,155 (6.18%)	Book15	4	464 (7.18%)
Book 6	3	1,031 (5.52%)	Book16	3	1,167 (18.07%)
Book 7	1	1,786 (9.56%)	Book17	3	1,680 (26.01%)
Book 8	2	294 (1.57%)	Book18	5	305 (4.72%)
Book 9	6	794 (4.25%)	Book19	3	418 (6.47%)
Book10	2	423 (2.26%)	Book20	1	448 (6.93%)
Total:	27 (49.09%)	18,685 (74.31%)	Total:	28 (50.91%)	6,460 (25.69%)
Mean:	2.70	1868.5	Mean:	2.80	646
Std. Deviation:	1.76	2575.21	Std. Deviation:	1.81	462.25

The above table shows that other than book 3 and 4, all other books of the non-academic group received less than double figured percentage. Contrariwise, from the academic book group, the book 17 received the highest number of social tags (1,680; 26%), and the rest received less than five percent of the respective group's social tags. There was no significant difference in assigning LCSHs terms between non-academic and academic books by the cataloguers. It is obvious that the cataloguers treated the academic and non-academic books equally in terms of assigning LCSHs.

There are variations among non-academic books in terms of receiving the number of social tags. Though the highest numbers of social tags are available in one non-academic book, it is found that the Standard Deviation of non-academic books was higher (2575.21) than academic books (462.25). Therefore, it is apparent that end users assigned similar social tags to academic books, whereas more diverse social tags to non-academic books.

4.6 Distribution of Social tags according to tag categories

This study has analyzed 25,145 social tags considering the appropriateness for particular book where those assigned. The distribution of social tags into the four main categories, A (Exact Match), B (LCSHs Partial match), C (Bibliographical information), D (User specific information) showed that the vast majority of social tags qualified for the last category 10,827 (43%). Whereas, 4078 (16%) social tags were exactly match, the LCSHs

terms assigned by the experts. At the same time, 9,611 (38%) social tags were partially matched with LCSHs terms, i.e. either with synonyms or broader term, or narrower term, or related term, or sub-division. The Figure IX shows the distribution of social tags in broad categories.
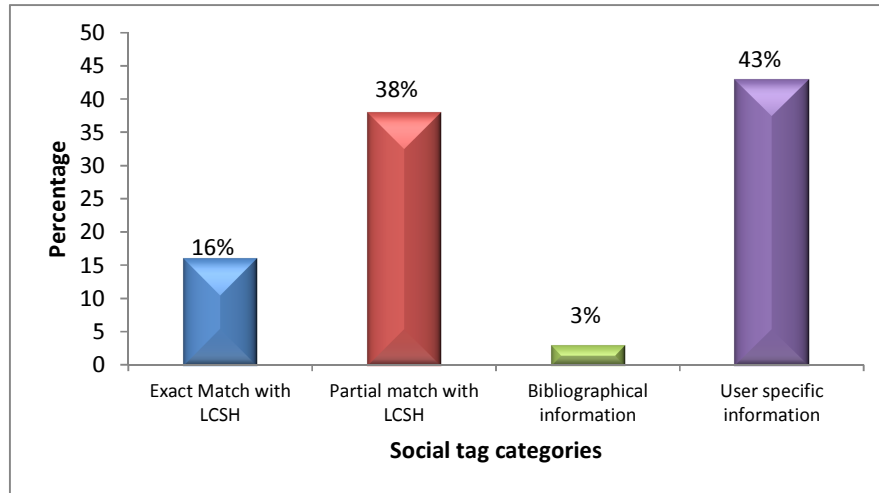
Figure 9: Distribution (in percentage) of social tags in broad categories

It is also found that there were 629 (3%) social tags that fell in the category of bibliographical information which were sub-categorized as author/title, publishing details, format information, and language of publication. A very small number of social tags fell in this category, but it indicates that end users also focused on basic bibliographical details of books. It is interesting that in some previous research, bibliographical information either has been considered as a non-match term or was not considered at all. The broad category, user specific information has been sub-categories, namely 'user community', 'tags in non-English', 'background information', 'related concept', 'not related', 'partial topic/chapter', 'use/user specific', 'unclear/ non-specific'. A large amount of social tags fell in this category that indicates end users like to assign own terms to the books, also remarked by Golder and Huberman (2006), and mentioned by Rolla (2009). The broad category, 'user specific information' provides other perceptions of social tags apart from LCSHs, how end users look at that book other than subject descriptive terms, will be discussed later.

Table 2: Distribution of social tags in all categories

Books	Total Tags	Exact Match with LCSH	Partial match with LCSH					Bibliographical information				User specific information							
		A-A	B-S	B-BT	B-NT	B-RT	B-SD	C-AT	C-PD	C-FI	C-LP	D-UC	D-OL	D-BI	D-RC	D-NR	D-PC	D-US	D-NS
		Exact match	Synonyms	Broader term	Narrower term	Related term	In sub-division	Author/Title	Publishing details	Format information	Language of publication	User Community	Tags in non-English	Background information	Related concept	Not related	Partial topic/chapter	Use/user specific	Unclear/Non-specific
Book 1	669	234 34.98%	--	129 19.28%	7 1.05%	--	5 0.75%	2 0.30%	3 0.45%	8 1.20%	--	--	2 0.30%	--	109 16.29%	21 3.14%	11 1.64%	113 16.89%	25 3.74%
Book 2	1030	89 8.64%	--	382 37.09%	--	--	--	6 0.58%	--	5 0.49%	--	--	24 2.33%	4 0.39%	98 9.51%	137 13.30%	71 6.89%	201 19.51%	13 1.26%
Book 3	2554	251 9.83%	--	1133 44.36%	--	--	--	11 0.43%	19 0.74%	29 1.14%	2 0.08%	--	26 1.02%	5 0.20%	61 2.39%	38 1.49%	378 14.80%	491 19.22%	110 4.31%
Book 4	8949	673 7.52%	--	4109 45.92%	6 0.07%	--	--	113 1.26%	11 0.12%	95 1.06%	22 0.25%	4 .04%	118 1.32%	14 0.16%	78 0.87%	210 2.35%	1041 11.63%	1796 20.07%	659 7.36%
Book 5	1155	83 7.19%	--	460 39.83%	--	--	--	10 0.87%	--	11 0.95%	--	2 0.17%	--	2 0.17%	10 0.87%	176 15.24%	86 7.45%	265 22.94%	50 4.33%
Book 6	1031	92 8.92%	25 2.42%	490 47.53%	27 2.62%	--	--	--	8 0.78%	7 0.68%	4 0.39%	--	2 0.19%	--	--	50 4.85%	97 9.41%	198 19.20%	31 3.01%
Book 7	1786	416 23.29%	11 0.62%	396 22.17%	19 1.06%	198 11.09%	--	--	35 1.96%	49 2.74%	6 0.34%	--	7 0.39%	47 2.63%	21 1.18%	78 4.37%	27 1.51%	377 21.11%	99 5.54%
Book 8	294	31 10.54%	--	97 32.99%	--	4 1.36%	--	--	4 1.36%	6 2.04%	--	--	4 1.36%	3 1.02%	--	17 5.78%	--	72 24.49%	56 19.05%
Book 9	794	374 47.10%	--	--	2 0.25%	--	--	2 0.25%	--	23 2.90%	--	--	2 0.25%	--	16 2.02%	3 0.38%	97 12.22%	211 26.57%	64 8.06%
Book10	423	176 41.61%	--	82 19.39%	10 2.36%	--	--	--	--	4 0.95%	--	--	--	--	2 0.47%	31 7.33%	5 1.18%	91 21.51%	22 5.20%
Book11	314	22 7.01%	--	--	--	--	--	--	--	--	--	--	--	--	151 48.09%	2 0.64%	33 10.51%	91 28.98%	15 4.78%
Book12	194	47 24.23%	23 11.86%	63 32.47%	--	--	--	--	--	--	--	--	--	--	--	--	--	48 24.74%	13 6.70%
Book13	659	205 31.11%	--	100 15.17%	26 3.95%	7 1.06%	--	--	2 0.30%	--	--	15 2.28%	6 0.91%	--	4 0.61%	8 1.21%	16 2.43%	251 38.09%	19 2.88%
Book14	811	318 39.21%	--	97 11.96%	--	--	--	--	--	10 1.23%	--	--	48 5.92%	--	2 0.25%	30 3.70%	10 1.23%	242 29.84%	54 6.66%
Book15	464	55 11.85%	23 4.96%	181 39.01%	--	--	--	--	6 1.29%	--	--	2 0.43%	4 0.86%	--	--	11 2.37%	23 4.96%	119 25.65%	40 8.62%
Book16	1167	38 3.26%	6 0.51%	685 58.70%	--	--	--	56 4.80%	--	13 1.11%	--	--	22 1.89%	--	18 1.54%	17 1.46%	60 5.14%	212 18.17%	40 3.43%
Book17	1680	453 26.96%	--	623 37.08%	37 2.20%	--	4 0.24%	2 0.12%	3 0.18%	15 0.89%	2 0.12%	--	--	--	28 1.67%	56 3.33%	37 2.20%	351 20.89%	69 4.11%
Book18	305	82 26.89%	--	67 21.97%	--	--	--	--	5 1.64%	--	--	2 0.66%	--	--	11 3.61%	3 0.98%	50 16.39%	72 23.61%	13 4.26%
Book19	418	258 61.72%	--	--	--	--	--	3 0.72%	--	6 1.44%	2 0.48%	--	--	--	9 2.15%	26 6.22%	43 10.29%	62 14.83%	9 2.15%
Book20	448	181 40.40%	--	70 15.63%	7 1.56%	--	--	3 0.67%	2 0.45%	4 0.89%	--	7 1.56%	--	--	--	69 15.40%	6 1.34%	73 16.29%	26 5.80%
Total	25145	4078 16.22%	88 0.35%	9164 36.44%	141 0.56%	209 0.83%	9 0.04%	208 0.83%	98 0.39%	285 1.13%	38 0.15%	32 0.13%	265 1.05%	75 0.30%	618 2.46%	983 3.91%	2091 8.32%	5336 21.22%	1427 5.68%
		4078	9611					629				10827							

The Table 2 shows the number of social tags fell in different sub-categories and the corresponding percentage based on the total number of social tags of that book. The figure below shows the percentage of each sub-category of social tags based on the entire tags count.
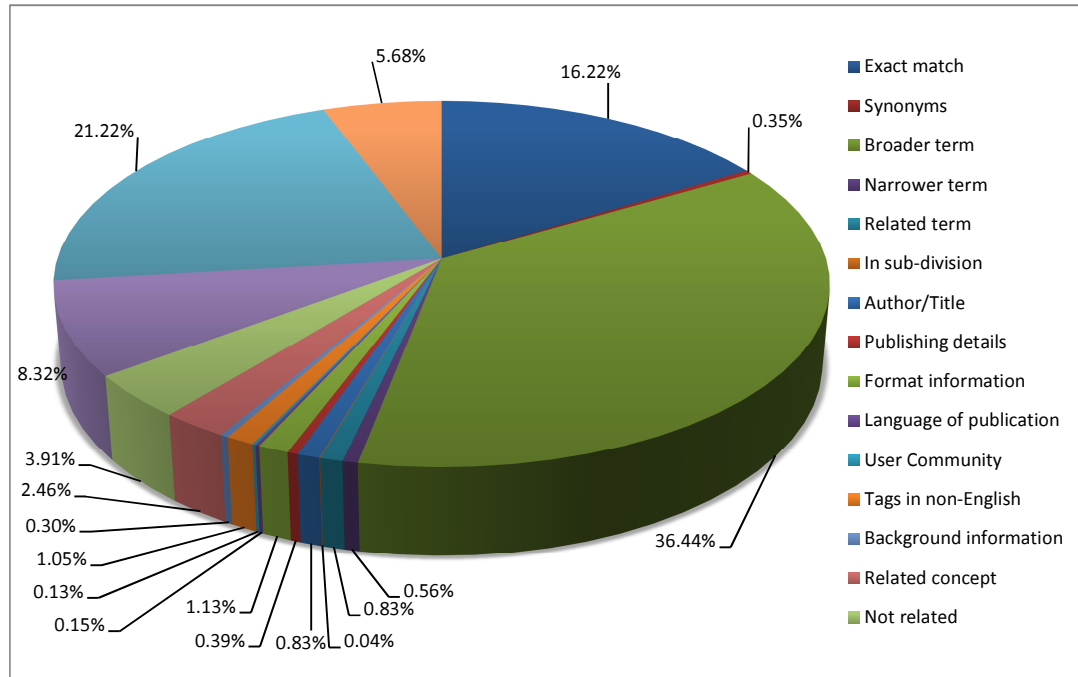
Figure 10: Appearance of social tags in different categories

The above figure shows that the majority (36%) of the social tags fell in the 'broader term' sub-category, and the second major sub-category is 'use/user specific' (21%). Apart from these, 'exact match' category received 16% of social tags, and the remaining sub-categories gained less than 10% of social tags. The variances of social tags in different sub-categories reflect that end users chooses diverse social tags in addition to subject descriptive terms.

4.6.1 Exact match with LCSHs

It is found that 16% of social tags were an exact match of LCSHs terms. The study also verified whether all the LCSHs terms used by the cataloguers have been assigned by the end users as social tags or not. All the LCSHs terms were found in the social tags list. The percentages of exact match social tags have been calculated for each book (Figure 11). The presence of the assigned LCSHs terms in the social tags list indicates that LCSHs terms are also assigned by the end users.
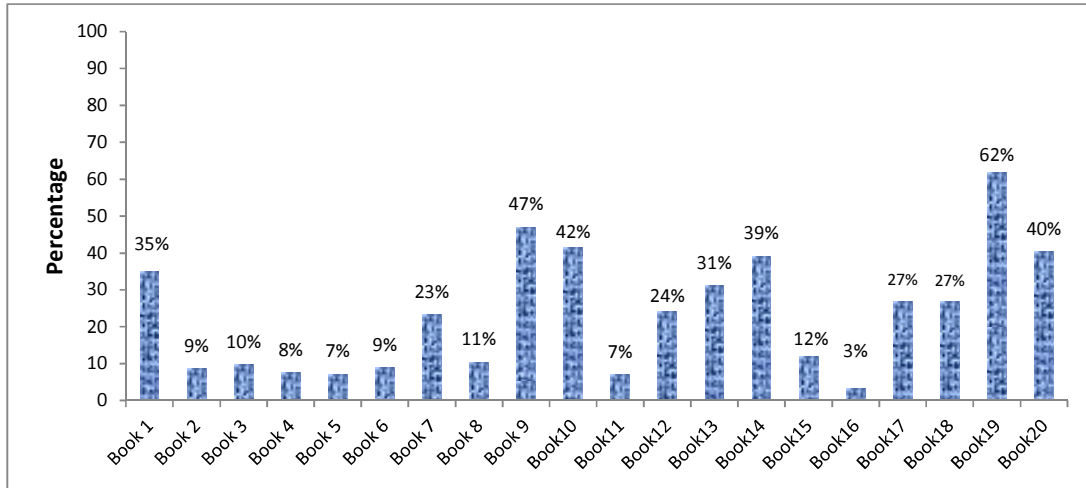
Figure 11: Percentage of exact match in each book

In the book code 19, it is found that 62% social tags exactly matched with LCSHs terms, which is a very high percentage. On the other hand, in the book code 16, only 3% of social tags matched with LCSHs terms. It is observed that both the books are representative of the academic book group. Conversely, the other five books, where less than 10% social tags matched exactly with LCSHs terms, were from the non-academic book group. It is apparent that comparatively more social tags match exactly with LCSHs terms in case of academic books rather than non-academic books.

4.6.2 Partial match with LCSHs

The Table 2 also shows 9,611 social tags partially matched with LCSHs, 38% of the entire social tags. This is more than double of exact match. It was mentioned in the methodology chapter that while evaluating social tags for the partial match, only the social tags that matched with either synonyms or broader term or narrower term or related term or subdivision of the exact match LCSHs term of that particular book has been considered. Under this broad category, the sub-category broader terms received 9,164 (95%) social tags. This is an extremely high percentage where the remaining sub-categories received only 5% of social tags in total.
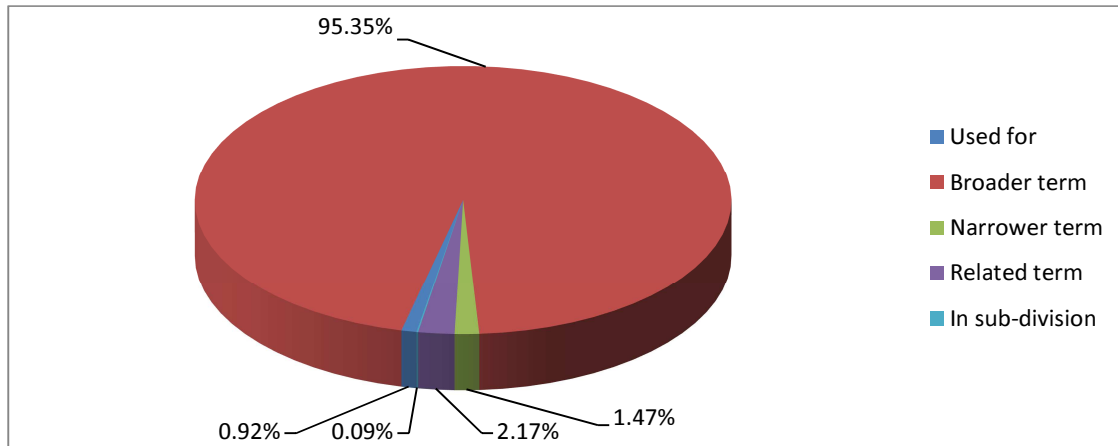
Figure 12: Distribution of social tags in sub-categories of partial match

For an example, a representative of the non-academic books "*The trouble with physics: the rise of string theory, the fall of a science, and what comes next*," where the assigned LCSHs terms were 'physics' (appeared 184 times in social tags), and 'string models' (appeared 69 times in social tags), Whereas, the broader term 'science' appeared 129 times as a social tag on that book. Considering a representative from the academic book group, "*QED: the strange theory of light and matter*" the assigned LCSH terms were 'Photons', 'Electrons', 'Quantum electrodynamics' appeared 3, 3, and 32 times in social tags respectively. Whereas broader terms like 'light' (14 times), 'physics' (373 times), 'quantum field theory' (7 times), 'quantum physics' (29 times) and 'science' (262 times) appeared in social tags. Therefore, it is noticeable that the number of appearance of broader terms was higher than assigned LCSHs terms.

On the other hand, Wetterstrom's (2008) study showed that narrower terms appeared more frequently than boarder terms, which are opposite findings of this study. But this present study's finding is similar to Rolla's (2009) study. Rolla mentioned that Wetterstrom's study group consisted of twenty individuals and a small number of persons contributed social tag, which could be one of the reasons, and other reasons could be that the group was too focused in creating social tags other than personalized tags. This indicates that end users assigned broader terms as social tags to express the subject matter of a book, which is also remarked by Golder and Huberman (2006) and Munk & Mork (2007).

4.6.3 Bibliographical information

It is a very common practice by the end user to search for a book either by author or by title, sometimes with the publisher's name. In addition, users like to know what is the format of the book, either e-book or print or audio book, etc., and even some others like to know whether the same book is available in their own language or not. Considering these approaches, this study found some social tags indicate bibliographical information of books. This broad category was sub-categories as 'author or title', 'publishing details', 'format information', 'language of publication', etc. The Figure IX shows that there were only 629 (3%) social tags that represent the bibliographical information of sample books. The possible reason behind the small number of such social tags could be the bibliographical information of the particular book is already available to the user while searching in LibraryThing.
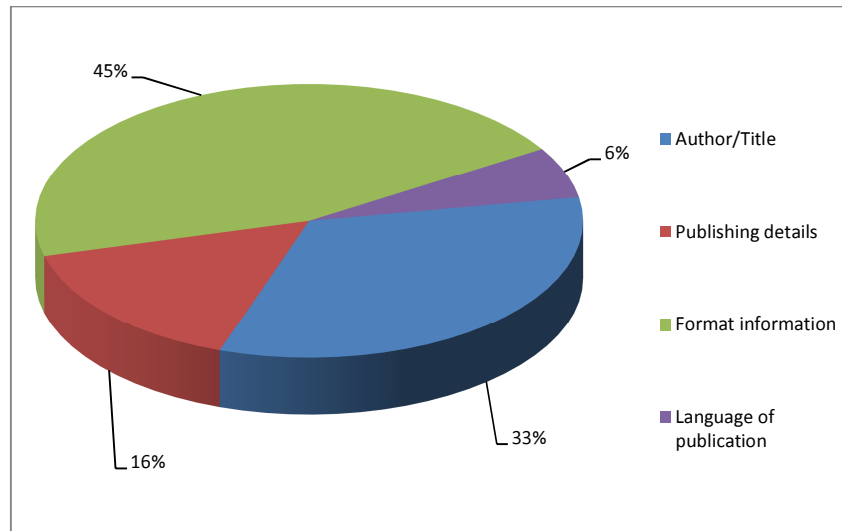
Figure 13: Distribution of social tags in sub-categories of bibliographical information

The above figure shows that the sub-category 'author/title' received 33%, and 'format information' received 45%, while the reaming two sub-categories e.g. 'publishing details' and 'language of publication' together received 22% of social tags. It is noticeable that altogether 'author/title', and 'format information' sub-categories received the majority (78%) of bibliographical information social tags. For example, the book '*The fabric of the cosmos: space, time, and the texture of reality*' has the following bibliographical details according to the library of congress online catalogue.

Figure 14: Snapshot of the library of congress catalogue and full record and MARC tags

The Library of congress catalogue shows 'print', 'microform', 'electronic' as the format of the book, where the social tags mentioned 'audible' and 'audio book' (13 times), 'e-book' (5 times), 'hardcover' (5 times) and 'paperback' (4 times). The term electronic did not indicate precisely whether this book is available as e-book or audible. The term print usually indicates either hardcover or paperback. Usually the prices of paperback are lower than hardcover. It is important for the users as well as to the libraries for the book selection process to know the specific format of the book for acquisition purpose. To make sure that the book available in the formats that mentioned in the social tags, the amazon.com has been consulted.
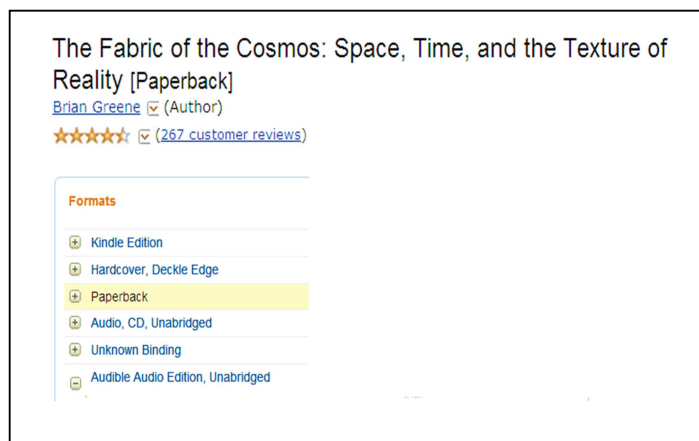
Figure 15: Snapshot from amazon.com

As a business-oriented site amazon.com always ensure the availability of the all the formats. It is found that all the formats mentioned in the social tags were available in the amazon.com (Figure 15). Social tags indicates format information of a book more precisely than the OPAC.

Another example, the book 'A brief history of time' has been assigned social tags like English (9 times), Finnish (2 times), French (2 times), German (4 times), Spanish (5 times) etc. The library of Congress OPAC has not provided any information whether the book was translated in other language or not. That is why, the WorldCat (www.worldcat.org, last accessed on 14 May, 2012) has been consulted. According to WorldCat, the book was available in 27 different languages, including the five languages mentioned in the social tags. The cataloguer input bibliographic information like 'language' or 'format' based on the library materials. Translations or different formats of a book are made available later than the original one. However, end users always have the advantages to know such information and assigned social tags to the library material after it became available in the library database. The presence of such kind of social tags indicates that some social tags provide more updated bibliographical information than the library OPAC.

4.6.4 User specific information

This study evaluated the social tags according to suitability to particular books they were assigned to. It was found that there were some social tags that matched with different sub-categories of the broad category 'user specific information'. Golder and Huberman (2006) identified such social tags as personal or individual terms, Rolla (2009) remarked as personal tags, where Wetterstrom (2008) and Carman (2009) put those under the broad category of no match with either LCSHs or OCLC records. The Table 2 shows 10,827 (43%) number of social tags fell in this broad category. Usually, these social tags do not provide any subject access to the books. However, they might have strong value to the end users who assigned them (Rolla, 2009; Wetterstrom, 2008).
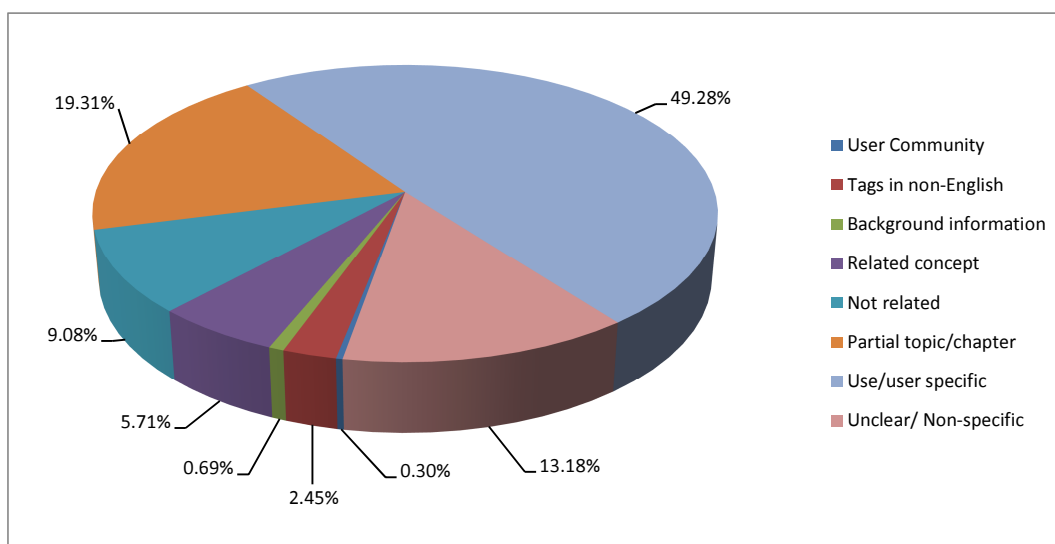
Figure 16: Distribution of social tags in the sub-categories of user specific information

The above figure shows that the sub-categories were 'user community' (0.30%), 'tags in non-English' (2.45%), 'background information' (0.69%), 'related concept' (5.71%), 'not related' (9.08%), 'partial topic/chapter' (19.31%), 'use/user specific' (49.28%) and 'unclear/non-specific' (13.18%), and the percentage mentioned in parenthesis reflects the amount of social tags assigned in consideration of the broad category. Each of the sub-categories has been discussed below:

4.6.4.1 User community

There were 0.30% of social tags that belonged to 'user community' sub-category; examples of such tags are tertiary, undergrad, university, school, high school, college, etc. These social tags represented user community for whom the book is appropriate. It is observed that the same book received several of such tags like high school, university, undergrad, etc. It indicated that end user put the social tags based on their respective community. The number of such social tags was very few, indicating that only a few users have added that in the books. These social tags may help other end users to determine the appropriate level of study of the book.

4.6.4.2 Tags in non-English

LibraryThing is open for worldwide use. Any member of it can assign social tags to any book. Currently, it allows assigning social tags in 55 different languages letting a wide universe of users to assign social tag in their own language. This study found 265 (2%) social

tags that were assigned in different languages other than English. The translation of those social tags has been done with the google translator. Some of the examples of these social tags have been shown below with English translation in parenthesis:

> **Danish**: naturvidenskab (science); **German**: physikalische tabellen (physical tables), bohr (drilling), chemische Daten (chemical data), datensammlungen (data collections), fakta (facts), große vereinheitlichte theorien (grand unified theory), handbücher (manuals), konstanten (constants), lichtgeschwindigkeit (speed of light), luftspiegelung (mirage), naturwissenschaften (natural sciences), physikalische daten (physical data), raum (space), raumzeit (space-time), relativitätstheorie (relativity theory), sammellinse (focusing lens), tabellenwerke (table works), unschärferelation (uncertainty principle), welle-teilchen-dualismus (wave-particle duality), zellbiologie (cell biology); **Dutch**: zwaartekracht (gravity); **Finnish**: aika (time), alkuräjähdys (big bang), avaruus (space), mailmankaikke uden synty (the birth of universe), maailmankaikkeus (universe), mustat aukot (black holes), suhteellisuusteoria (theory of relativity), tiede (science); **French**: lumière (light), particules (particles), temps (time), vulgarisation scientifique (popular science); **Italian**: divulgazione scientifica (popular science), divulgazione (disclosure), fisica: quantistica relativistica (physics: quantum relativistic); **Portuguese**: tempo (time); **Swedish**: svenska (swedish), vetenskap (science)

This study has not evaluated the appropriateness of non-English social tags for the respective books, as it considered only the social tags that were in English. It is obvious that these social tags appeared multiple times in a respective book. These social tags could have practical value in a library's OPAC environment where multinational or multilingual end users use the same library resources.

4.6.4.3 Background information

The study found some social tags that indicated background information either about the author of the book or any other information that associated with the book. This sub-category is only 0.69% of social tags of the broad category 'user specific information'.

For example, social tags like 'American' and 'Engineering' were assigned to the book '*E=mc²: a biography of the world's most famous equation by David Bodanis'*. It is found that the author is an American, and the subject content best suit to engineering discipline. In the book '*A brief history of time by Stephen Hawking*', presence of social tags like 'British', 'British author', and 'Oxford' reflected that the author 'Stephen Hawkings' is a

British, and graduated from Oxford University. '*The greatest show on earth: the evidence for evolution by Richard Dawkins'*, has received social tags like 'British', 'Charles Darwin', 'Darwin', etc. The author of the book is a British, and book content explained the evidence for evolution concerning Charles Darwin. Another book that was '*Electric universe: the shocking true story of electricity*' received social tags e.g. 'Faraday'. Noticeably, the social tags indicated about 'Michael Faraday', best known for his work regarding electricity and magnetism. Such social tags were not been available in the library OPAC, but obviously, these indicate background information that sometimes plays in the users' mind while they searched for books.

4.6.4.4 Related concept

The book '*The trouble with physics: the rise of string theory, the fall of a science, and what comes next*' received social tags like 'string theory', 'super string', 'superstring theory', 'quantum physic'. Conversely, the assigned LCSHs terms in this book were 'Physics', and 'String models'. The LCSHs terms did not reflect related concept where users assigned 'superstring theory' as a social tag, which is the version of string theory. The book '*Genome: the autobiography of a species in 23 chapters*' was assigned LCSHs terms like 'Human genome' and 'Human genetics'. The book discussed the mapping of the twenty-three pairs of chromosomes, and selected one newly discovered gene from each pair of chromosomes and described about it. Social tags like 'genes', 'human biology', 'human evolution', 'molecular biology', 'species',' etc. were assigned in that book which are related terms in respect the content of the book.

The above mentioned social tags are the examples of the 'related concept' sub-category. None of these social tags matched with LCSHs terms. Explicitly, such social tags reflect content of the book, while the cataloguer might consider the topics discussed as a whole and assigned LCSHs terms that described the book in general. This sub-category could be a good example of social tags that indicates the concepts not brought out by LCSHs terms.

4.6.4.5 Not related

The study considered the appropriateness of social tags in the particular book. There are 9% of social tags that fell on the 'not related' sub-category. It is found that social tags that fell in this sub-category were inappropriate to reflect the subject matter of the book. For example,

the book '*Zero: the biography of a dangerous idea*', is about the number 'zero' with a lot of mathematics and philosophy as well as a bit of physics. Social tags like 'discovery', 'fiction', 'history of ideas', 'nothing', 'social history' has been assigned to the book. These social tags did not reflect the actual subject discussed in the book. In the social tags list for the book 'The greatest show on earth: the evidence for evolution', received social tags, e.g. 'skepticism', 'sociology', 'history' etc. Such social tags do not fit according to the subject matter of the book. The book explained the evidence for evolution while exposing the irrationality of the creationist argument. Another book that is '*Nature's numbers: the unreal reality of mathematical imagination*', which explained the key concepts in mathematics and their implications, pointing out that mathematics is an entirely mental construct. This book also received some irrelevant social tags like 'biology', 'history', 'physics'. It could be assumed that these social tags may have strong individual value to the end users who assigned those, especially in consideration of end users who is doing some multidisciplinary things.

4.6.4.6 Partial topic/chapter

The study examined whether there were any social tag that reflected the chapter or a partial topic discussed in the book. It is found that there were 2,091 (8%) social tags fell under the sub-category 'partial topic/chapter'. To identify the appropriateness of such social tag on the specific book, the abstract or description of the book, content pages, and reviews have been consulted from amazon.com, LiraryThings, World Cat, Googlebooks etc. For example, the cataloguers assigned LCSHs terms for the book '*Molecular biology of the cell (4th edition)*' were 'cytology', 'molecular biology', 'cells'. It is found that beside discussion of cells, genomes, and cell chemistry to DNA, cellular organization, function and cell immunity, the authors also incorporated new discoveries in biotechnology, infection and immunity. The assigned social tags like 'biotechnology' 'genetics', 'medicine' reflected such contents of the book. The book '*Introduction to electrodynamics (3rd Edition)*', discussed electricity and magnetism covering: electrostatics, electric fields, magneto statics, magnetic fields, radiation, relativistic electrodynamics. For this book, the cataloguers assigned only one LCSHs term, i.e. 'Electrodynamics'. As the book is more introductory, rather than technical, a user might miss this book if not searched by the assigned LCSHs term. Whereas the assigned social tags like 'electricity', 'electricity and magnetism', 'electromagnetic theory', 'electromagnetic', 'electromagnetism', 'magnetism' etc. provided a deeper view regarding the subject matter of

the book. It is apparent that the cataloguers select LCSHs terms that reflect the main subject of the book being catalogued, whereas social tags identify very specific and perhaps minor aspect of the same book. Moreover, cataloguers assign LCSHs terms without reading the book entirety, whereas users usually assigned social tag after reading them, also mentioned by Wetterstrom (2008).

4.6.4.7 Use/user specific

The study noticed that there were 5,334 social tags that fell under the 'use/user specific' sub-category. It covered 49% of the 'user specific information' broad category, and it is almost half of that broad category. These social tags do not provide any subject access to the book (Adler, 2009; Lu, Park, & Xiaohua Hu, 2010; Rolla, 2009). Examples of such tags are below:

> alternative-read, bedroom, bookshelf 4, borrowed, bought, currently reading, dissertation, donate, duplicate, favorite, finished, genome project, gift, gone, handbook, have read, home library, imported, in storage, inventory, loan, lost, male author, manhattan project, missing, multiple copies, need to read, not started, not yet read, office, own, owned, partially read, personal, pictorial , popular mathematics, primary source , read, read but unowned, read in 2011 , read not owned, reading, recommended, review, science dept, share, storage, textbook, to read, unowned, unread, want, want to read, we have 3 different editions, wishlist

The social tags like 'read not owned', 'unread', 'read in 2011', 'need to read', 'primary source' etc. may have retrieval value for the particular end user who assigned those. But those social tags do not have any importance to other end users. At the same time, the social tags like 'imported', 'inventory', 'recommended', 'science dept', 'we have 3 different editions', 'wishlist' etc were user specific, and may be several users of a group were taking advantage to locate the book to meet their purpose.

It is apparent that if a group of people found a book useful for their common purpose could assigned social tags with their familiar words to facilitate other members of the group to find the same straightforwardly (Golder & Huberman, 2006; Tonkin et al., 2008; Carman, 2009; Rolla, 2009; Thomas, Caudle, & Schmitz, 2009).

The idea behind such social tags could be that a user can find book tagged by other users and therefore increase respective list of relevant resources, which may not be found otherwise

(Wetterstrom, 2008). In spite of the relatively limited presence of such social tags, there is still a indication of the possibility of forming a user group that create useful social tags for themselves irrespective of taking account of subject access (Weinberger, 2007).

4.6.4.8 Unclear/Non-specific

The last but not the least sub-category was 'unclear/non-specific'. These social tags did not fall into any other sub-categories categories mentioned in this studym e.g. 'dressing', 'G', 'worldview' 'wormholes',' etc. Most of the time, such social tags appeared as either cryptic or difficult to interpret like '@', ' * ', ' #1010 ' etc. More examples of such tags are shown below:

> 1, 2, * , @ , 4c, 5-2 , A1 , acquired 2010, adults, all , B, B1, basic data, BCE, bed3shelf4front, Blue, Box, BP, bye, chem, Chris, class, Cover Clear, CSUF, D3, dave, dfw , DH, dj, dressing, E&M, English literature, evo, from_half_022508, FTL, G, gw, H, hard, HC, health, his , HMH, huge, impossible, in, jacket, L:Science, Loc:SERC, main, Mark, mooched, NF, NIL, NR, O, PDF, pharmacy, phil, phys, pl, politics, promo, psyhics, Qb, QFT, QM, R, Red, Sagan, sci, SET DEC, SO, Tables, tbt, tech, telepathy, tieto, universe, war, wissen, work, worldview, wormholes

This study took into consideration only those tags that appear at least twice. It means at least two people assigned the same social tags. Furthermore, spelling variation or wrong spelling, singular or plural form of social tags has not been corrected before analysis. Therefore, there might be some social tags of this sub-category that might fall in other sub-categories if those variations were fixed. The reason behind keeping them intact is to determine the automatic adaptability of social tags in support of the library OPAC. In this respect, these apparently 'unclear/non-specific' social tags may have value for personal use only to those who assigned them. Therefore, it may conclude in a way that the appearance of such social tags is normal in tagging practice, useful for individual use only, and do not have retrieval value for other users.

Chapter V: Discussion and Conclusion

5.1 Introduction

The study compared social tagging and expert created subject terms of the science genre. The methodology followed, and the process of data analysis, helped to figure out findings that are discussed in the Chapter IV. This chapter integrates the analysis and findings, and discuss further according to the research questions set for this study in the introductory chapter.

5.2 Occurrence of social tags and subject terms

Twenty books have been considered for this study, and it is found that these books have been assigned 55 LCSHs terms while the same books received 6,600 unique social tags (Discussed in the section 4.2). This means that on an average (per book) 2.75 LCSHs terms has been assigned by library cataloguer, whereas on an average 330 social tags has been assigned by the users. The end users assigned more social tags to the books than library cataloguers assigned subject headings (Discussed in the section 4.5), which is similar to the findings of Lu, Park, & Hu (2010), Carman (2009) and Rolla (2009). This difference reflects the fact that the cataloguer has considered only the LCSHs terms that best described the subject matter of the book. Therefore, it is observed that users liked to assign social tags, and at the same time it reflected that there is diversity between users in terms of choosing social tags.

5.3 Specificity versus frequency of social tags

This study found that there were a large number of social tags that appeared only once for a book. For the purpose of this study, only the social tags that appeared at least twice have been considered. However, it found that the same single occurring social tag has appeared for several books. For example, the social tag 'modern physics' appeared for three different books e.g. '*The trouble with physics: the rise of string theory, the fall of a science, and what comes next*', '*A brief history of time*', '*QED: the strange theory of light and matter*'. 'Modern Physics' is neither a LCSHs term nor had it been assigned by end users for books apart from the above three. An investigation of those single tags gives an idea that those social tags might have personal value for specific users. But this value may be too personalized to be useful for others even if those occur more than once.

5.4 Academic versus Non-academic books

As mentioned in the methodology chapter, out of 20 books, 10 books were chosen from non-academic group and the rest 10 from academic group. In this study, academic books are concerned with course curriculum of school, college or universities, and non-academic books are not concerned with specific course curriculum but they could be of the same topic/subject as academic ones and used for general reading or to gain additional knowledge. It is found that the non-academic books received majority of social tags (74%) in comparison to academic books (26%). It indicated that non-academic books have been assigned at least three times more social tags than academic books. Contrariwise, the cataloguers assigned a slightly lower number of LCSHs term for non-academic books (49%) comparing to academic books (51%). Therefore, it is noticeable that social tags were more frequently assigned in non-academic books than academic books, where there is almost no difference in assigning LCSHs terms between these two (Discussed in section 4.5). However, the amount of LCSHs terms has no significant impact upon the number of social tags occurrence in the non-academic and academic books groups. In addition, non-academic book's social tags were more diverse than academic books. Therefore, it is obvious that end users agreed on more similar social tags in academic books than non-academic books.

5.5 Findings related to the research questions

The first research question was '***To what extent does social tagging represent the same concepts as the expert created subject headings***?' This question tried to identify the amount of social tagging that matched with the expert created subject headings. It is found that 16% of the social tags were exactly matched with LCSHs that has been assigned in the Library of Congress OPAC by the cataloguer. These terms were exactly the same as the LCSH terms assigned by the experts. At the same time, 38% of the social tags were partially matched with LCSHs and fell under the subcategories as either synonyms or broader term or narrower term or related term or sub-division (discussed in section 4.6). In both the cases, the social tags, represented the same concept as the expert created terms
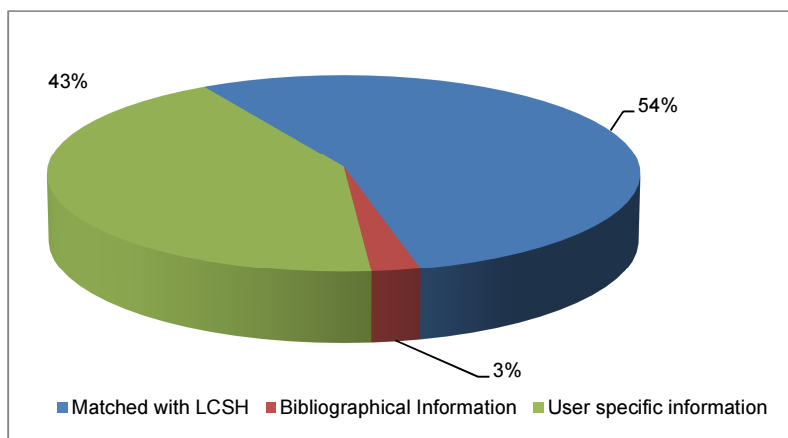
Figure 17: Match and non-match of social tags with LCSHs

The figure 17 shows that there were 54% social tags exactly matched with LCSHs and the 3% reflected bibliographical information and the reaming 43% social tags were user specific information. The last two categories were considered as non-matched social tags with LCSHs, and the combination of them indicated that 46% social tags did not match with LCSHs.

The data presented in the Table 2 indicates that social tags matched with LCSHs terms varied between 68% to 44% of the individual books, except one book (Discussed in section 4.6.1; see figure 11). In the book, '*Introduction to Electrodynamics (3rd Edition)*', only 7% social tags matched with the LCSHs term. From the above discussion, it is apparent that 54% social tags exactly represented the same subject terms available in LCSHs, and 46% social tags did not match with LCSHs. Consequently, more than half of the social tags represented similar concepts as LCSHs.

It is mentionable here again that this study has analyzed social tags in terms of the appropriateness of each book. Obviously, if all the books were considered together, some social tags would match exactly with more LCSHs terms. For example, the book '*Concrete mathematics: a foundation for computer science*', has been assigned social tag 'Information Technology'. Apparently, for the title, it seemed that the term is appropriate with the book, and it is also available in the LCSHs, and the book might have discussed something about computers that fall under the broad category of Information Technology. 'Information Technology' is a LCSHs term where the instruction has been given as "*Here are entered works on the acquisition, processing, storage and dissemination of vocal, pictorial, textual*

and numerical information by microelectronics, computers and telecommunication". According to the instructions of LCSHs, it is clear that the term cannot be considered as an appropriate subject heading for this book. This study considered that assigning inappropriate subject terms as social tags to a book has no retrieval value for other end users.

The second research question was '***Do social tags that match with the expert created subject headings receive higher frequency of use***?' This question tried to know whether the LCSHs are more frequently used by the users or not in social tags list.

This study has considered 1,365 unique socials tags, and each of them has been assigned by the users at least twice (Discussed in section 4.2). To identify the most frequently occurring social tags, the number of social tags that appeared in different sub-categories has been considered. The occurrences refer to the number of unique tags in the sample. It is found that the most frequently occurring social tag categories were those that did not match with LCSHs, similarly as remarked by Carman (2009).
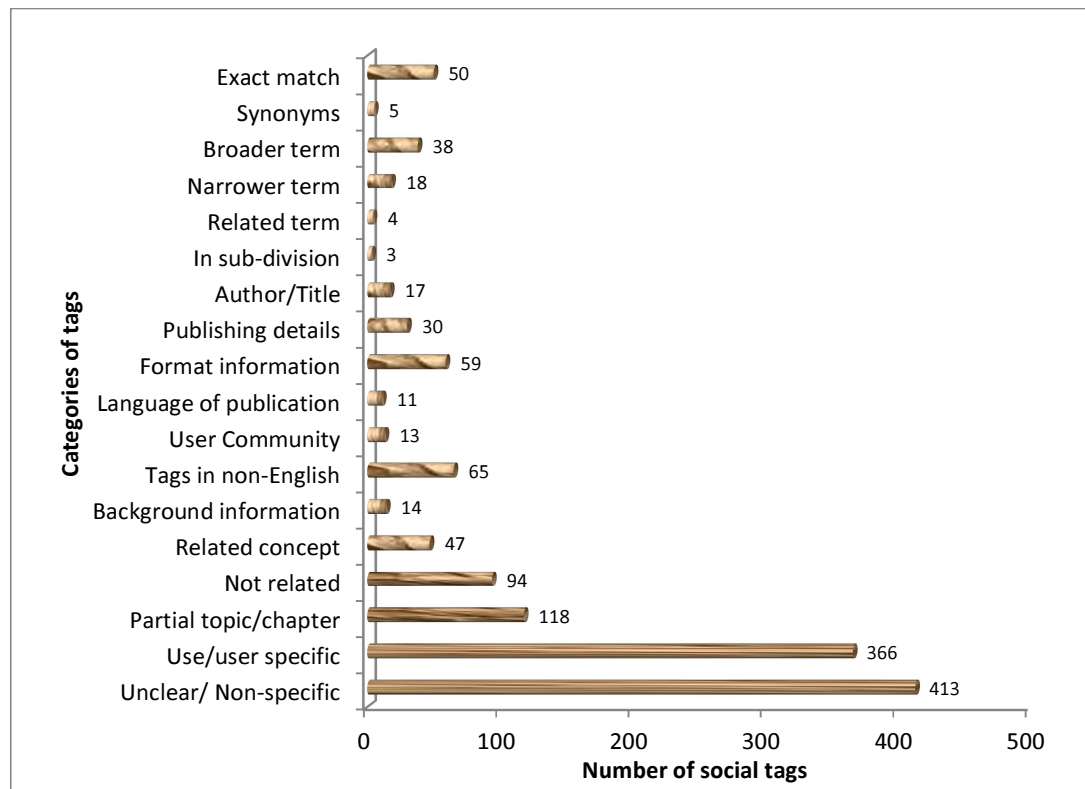
Figure 18: Number of unique social tags occurrences

The figure 18 shows that the numbers of 'unclear/non-specific' tags were highest among all the sub-categories, followed by 'use/user specific' sub-category. For example, social tags occurrence in sub-categories like 'partial topic/chapter' 'not related', 'use/user specific', and 'format information' are higher than in the exact match category.

At the same time, an attempt has been made to identify social tags of which sub-category were most frequently used by end users, considering all sub-categories. It is common that the more popular a social tag among the end users, the more frequently it is assigned to a collection of books. The frequency or popularity of a social tag is measured by the number of times it is assigned by the users.
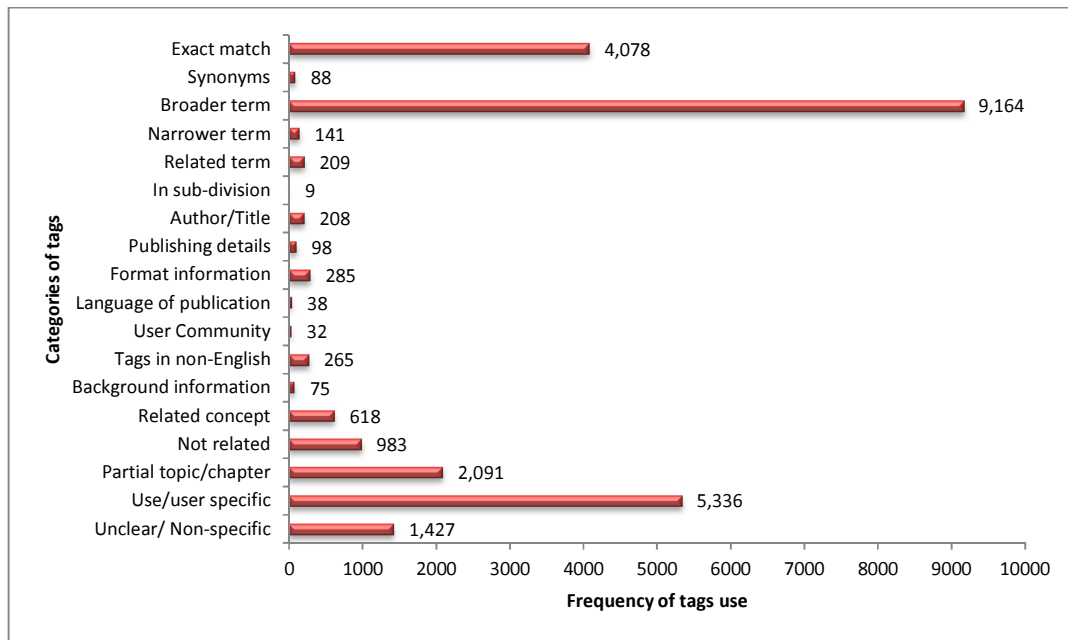
Figure 19: Frequency of use of social tags in each sub-category

It is observed that 'exact match' (50 terms) and 'broader term' (38 terms) sub-categories has been used 4,078 and 9,164 times by the users (Figure 19). While for the sub-categories with highest number of separately occurring terms like 'unclear/non-specific' (366 terms) and 'use/user specific' (413 terms), the frequency of use were 1,427, and 5,336 times respectively. It is obvious that 'exact match' and 'broader term's matched social tags have been assigned more frequently than the other, non-matched subcategories. End users assigned a relatively higher number of social tags that particularly matched with LCSHs terms.

Pearson's correlation coefficient has been calculated between numbers of occurrences and usage frequency of social tags. It helped to assess how well the relation were between the occurrences and usage frequency of social tags, and whether the social tags that matched with expert created subject headings were more frequently used by the users or not. Pearson's correlation coefficient value is in the interval [–1, 1]. If Y tends to increase when X increases, the Pearson correlation coefficient is positive. If Y tends to decrease when X increases, the Pearson correlation coefficient is negative. A greater absolute value of Pearson's correlation coefficient indicates a stronger correlation. A Pearson correlation of zero indicates that the two rankings are completely independent (Brase & Brase, 2006) .
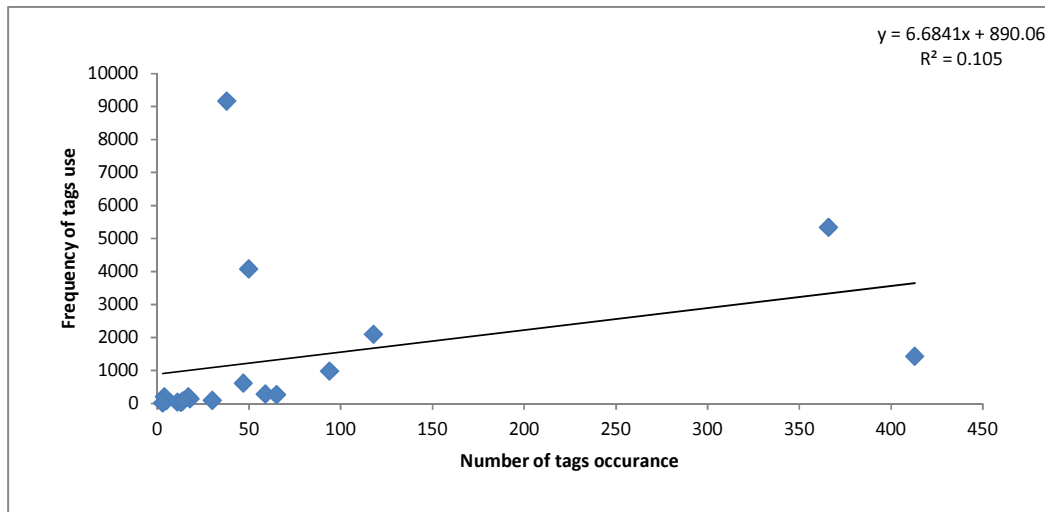
Figure 20: Pearson's correlation coefficient between numbers of occurrences and use frequency of social tags

In this case, the Pearson correlation coefficient of these two groups is 0.105. This indicates a positive, but very low correlation between the social tags' occurrence and frequency of use in each sub-category.

From the above discussion, it appears that though the numbers of occurrences were comparatively lower in LCSHs match categories (exact match and partial match), their frequency of use was higher than non-matched categories (bibliographical information and user specific information). Moreover, there were almost no correlations between the occurrence and frequency of use of social tags. Therefore, it is clear that social tags that matched with the expert created subject headings were more frequently assigned by the end users.

The third research question was '***Are the expert created subject headings highly ranked in the social tags' lists***?' This question investigates the average number of times a social tag was used for all the broad categories, how often end users concepts were reflected in LCSHs. It has been calculated based on the number of unique tags, and how many times on an average those social tags appeared in different broad categories. Later the categories were ranked according to the frequency of uses.

It is found that in total 50 unique LCSHs match exactly with social tags, and appeared 4,078 times in the social tags cloud. In addition, 68 unique LCSHs which were available either under the 'Synonyms' or 'Broader term' or 'Narrower term' or 'Related term' or 'sub-division' has been considered, and matched with the social tags cloud. These partially matched social tags were available under the exact match subject headings, but cataloguers may not assigned those terms as subject headings for the respective books in compliance with rules of LCSHs. On the other hand, it is found that there were 117 unique social tags that matched with 'Bibliographical information' of the books and appeared 629 times, and 1,130 unique social tags that matched 'User specific information' and appeared 10,827 times.

Table 3: Ranking of different categories based on average of appearance

Broad Categories	Number of unique terms	Number of Social tags	Average appearance	Rank
Exact match	50	4078	81.56	2
Partial match	68	9611	141.34	1
Bibliographical information	117	629	5.38	4
User specific information	1130	10827	9.58	3

To identify how many times the unique terms has been appeared in social tag's cloud, the average appearance has been calculated by dividing the 'number of social tags' with 'number of unique terms'. From the above table 3, it is found that social tags that matched partially with LCSHs terms has appeared on an average 141.34 times, and ranked first. It indicated that the majority of the social tags matched partially (e.g. Synonyms or broader term, or narrower term, or related term, or sub-division) with LCSHs. Where social tags that exactly match with LCSHs terms appeared on an average 81.56 times, and ranked second. Consequently, User specific information (9.58 times) social tags, and bibliographic information (5.38 times) social tags has been ranked as third and fourth respectively. The last

two categories showed that the numbers of unique social tags were larger than the first two categories, but the average appearance of those social tags was relatively small.

It indicates that other than subject descriptive social tags, there is a wide disparity between the end users to assign social tags. In one way, it may be considered as weakness of social tagging. However, in this case, it might be considered that those non-subject descriptive tags may have social values as at least two users agreed on the same tag. As the social tags that either partially or exactly matched with LCSHs ranked first and second, respectively, therefore, it is evident that the expert created subject headings are highly ranked in the social tags' lists.

The fourth research question was '***What other perceptions are reflected by social tags than subject description?***' This question effort to identify what other supplementary information is provided by the non-matched social tags.

This study found that there were 25,145 social tags and 13,689 of them are exactly or partially matched with LCSHs. Therefore, it was significant to discover what other notions has been provided by the remaining 11,456 social tags. The small pilot study (discussed in section 3.9) indicated that there were social tags that did not fit with LCSHs terms, and broadly they could be defined as non-LCSHs terms or non-matched terms. According to the findings of the pilot study, for the analysis of social tags two broad categories has been created. The first one was 'Bibliographical information' and sub-categorized as 'author/title', 'publishing details', 'format information', 'language of publication' etc. the second one was 'user specific information' and sub-categorized' as 'user community', 'tags in non-English', 'background information', 'related concept', 'partial topic/chapter', 'use/user specific', 'not related', 'unclear/ non-specific', etc. The collected social tags were well distributed, and none of the social tags matched with more than one sub-category. It is mentionable that not all the books taken into consideration for this study received social tags in all the sub-categories.

This study found that there were 629 social tags that represented the bibliographic data of sample books (Discussed in section 4.6.3). On an average, each book received 31.45 social tags. These tags fell in to different sub-categories under the broad category 'Bibliographical Information'. It is a very common practice that apart from subject description terms, the

general users search for required books by bibliographical information, e.g. author, title, publisher, etc. The possible reason behind the small number of such social tags could be that the bibliographical information of the particular book is already available to the user while searching in LibraryThing.

It is interesting that the social tags also represented 'format information' and 'language of publication'. If the cataloguer has no readily available information about the publication of the book (in hand) in other 'language' or 'format', s/he could not input any information regarding this. Moreover, sometimes the translations or different formats of the book are made available later than the original one. However, users always have the advantages to know such information better than the cataloguers as they always tag the same book after it become available in the library database. Anyway, the existence of such social tags indicates that those social tags have bibliographic informative value, and provided more updated bibliographical information than library databases.

A large number of social tags have been identified that were related to use/user specific. It is found that there were some social tags that represented user community for whom the book was appropriate like tertiary, undergrad, university, school, high school, college, etc. (Discussed in section 4.6.4.1). Such social tags definitely provided an indication to users about the appropriate level of study of the book. These kind of social tags are significant for the users of the scientific books indeed.

It is observed that some social tags have been assigned in different languages other than English. This study considers social tags that were in English, and has not evaluated non-English social tags. For the sake of analysis, the translation of those non-English terms has been done, and it was found that many of them went along with the respective book (Discussed in section 4.6.4.2). It indicated that these social tags have real value for OPAC environment where users might like to tag in their own language. Moreover, such kind of social tagging practice could be particularly useful in the multilingual environment where users come from different countries.

This study also found that some social tags indicated background information either about the author of the book or any other information that has linkages with the book's author or topic, e.g. 'American', 'British', 'British author', 'Oxford',' etc. (Discussed in section 4.6.4.3). Such

social tags group background information that may sometimes play in the users' mind. For example, there might be some users who would be influenced to read books written by 'British author', or 'American authors'. Such social tags helped to pull out the similar books together as well as provide some insight information about the author or the topic. However, such terms are not being available in library catalogue metadata arrangement, but obviously, such social tags matched with users' preferences.

The LCSHs term assists cataloguer to provide a very general subject classification. It was found that sometimes the cataloguers assigned broader terms than the specific topic discussed in the book. It often did not point out the narrow concepts that have been discussed in the book (discussed in section 4.6.4.4). It is observable that social tags described the book with appropriate narrower subject terms. This variation might occur as the cataloguer assigned subject headings based on a technical reading of a book, e.g. title and subtitle, preface, content page, etc., whereas, end users probably read the book before tagging and assigned more narrow terms. Such kind of social tags could be a good example that implies the concepts not brought out by cataloguers.

This study considered the appropriateness of social tag for the particular book where it has been assigned. It is found that there were some social tags assigned to books which were totally irrelevant to the subject matter of the book (discussed in section 4.6.4.5). For example, a book of math has received social tags like 'biology', 'history', 'physics', which were inappropriate to the content of the book. However, the presence of such kind of social tags indicated that these might have strong individual value to the users who put them. More specifically, for the users, who were interested on multidisciplinary studies and tried to assigned social tags that would help them to pull all the books that were relevant to their studies. In addition, it was found that there were some social tags that would not go with the book's description. For example, the tag 'non-fiction', 'textbook', 'reference book', assigned to all the books that has been considered for this study. These social tags might be useful for the users who assigned them, but they do not fit into the traditional controlled vocabulary systems.

After searching and getting the bibliographical information about a book, it is common to look at the available format of the book (Discussed in the section 4.6.3). Until now, the print

version is very common for books. Some users who look for the availability of the same book as 'E-book' for their convenience of use and portability. Furthermore, due to the advent of new technologies use of the e-reader like 'Kindle' is not uncommon. A number of books has been assigned the term 'Kindle'. This study also checked the appropriateness of the social tags – whether the book was available for kindle or not. In all the cases, it is found that the tags suited well for the book. Such kind of social tags helped the users to look at the entire book collections that were available for 'Kindle'. For example, if anyone likes to buy books that are suitable for 'Kindle' will directly benefit from such social tags. Moreover, searching for books in rare formats using such tags as 'audio book' could be more helpful for the users who are looking for such a format.

It is usual to get some misleading social tags. While considering the suitability of assigned social tags for a particular book, it is found that there were some social tags that do not describe any feature of the book. For example, social tags like 'dissertation', 'essay', 'project', 'recommended' etc did not match with any book's description. Presence of such social tags is not exceptional, though they may mislead others while searching for books. However, such social tags may not have value for collaborating tagging, but users may put such social tags for personalization. For example, if someone works on dissertation or a project or writing an essay or someone recommends few books on a topic, then such kind of personalization will give the maximum benefit to those persons who assigned them.

Some social tags indicated double meaning, and the users needed to further be confirmed what those actually meant for. Social tags like 'American' or 'British' could be easily identifiable as related to country or citizenship, usually where the author came from. These tags gave a clear message in respect of those queries. Contrariwise, social tags like 'Spanish', 'Portuguese', 'Italian', often showed double indications. Further clarification was required to identify whether those tags meant the origin of authors birth place or language of the book published or translated. However, such social tags may encourage users to further look on, but apparently those help to pull together books that have been written in a specific language or written by authors came from a specific country. Some users may have interest to read books in his/her own language, and at the same time, fascination may be there to read views of authors who are from the same country. Such social tags may not be suitable for inclusion

as LCSHs, but they are interesting and useful to some readers. These social tags might be easily co-existed with the LCSHs, also suggested by Wetterstrom (2008).

It is found that social tags covered the oversight of professional cataloguers. Sometimes cataloguers overlooked to incorporate some LCSHs term that were appropriate for a book. For example, the book '*Introduction to electrodynamics*' has been assigned only one subject terms in the Library of Congress OPAC i.e. 'Electrodynamics'. It is possible only to find this book while search with the mentioned subject key word. There is more possibility to miss the book by the end user who is searching the same book with the broader term like 'physics'. The term 'Physics' has been assigned 124 times as social tags, whether the term 'Electrodynamics' has been assigned only 22 times for the same book. It is apparent that sometime the end user assigned broader subject headings as social tags than the cataloguers assigned.

5.6 Limitation of the study

This study was a small-scale research with time bindings; therefore, the sample was relatively small. Moreover, this study has not considered how many end users provided the social tags. Studies with a larger and randomly selected sample may vary from these findings. This study focused on social tags that has been collected from LibrayThing website and only compared with LCSHs terms as well as considered the Library of Congress's OPAC. However, Social tags collection from other than LibraryThing site and comparison with other controlled vocabulary in addition to LCSHs, and consideration of other large collection library's OPAC may indicate different result than this study. Moreover, the samples of this study were concentrated to Science genre only. Furthermore, the social tags analyzed in this study have been collected on a specific time frame. The number of social tags may increase by the time the study completed or afterwards. Therefore, caution should be taken before applying these results in consideration of social tagging as a whole.

5.7 Suggestions for further study

This study considered LibraryThings website for social tags collection. It did not consider the number of end users who assigned the social tags, age group or sex difference or from where the users came from. It would be interesting to compare the results of this study to similar studies with those demographical data. Most interestingly, there might be differences in social tagging practice between 'generation X' and 'generation Y'. Kajewski (2007) highlighted and also mentioned by Wetterstrom (2008) that the 'generation Y' is more competent with Web 2.0 technologies . At the same time they are the potential users of the future libraries. Therefore, it might be interesting to explore same kind of study with different age groups. In addition, as this study investigated books from Science genre, similar studies may be done with other genres like Social Sciences; Applied Science etc to compare the results with this study. Furthermore, the awareness and actual use of social tags in library OPACs are still a developing area (Rolla, 2009). Researchers should study the incorporation of social tagging in the next generation library websites. Moreover, it could be more appealing to examine adaptability of social tagging practice in digital library's website where the collections are not limited to books only.

5.8 Conclusion

There is clear difference between assigning expert created subject terms and social tagging practice to library books. Social tagging for a group of books and the cataloguer assigned subject headings for the same books showed that users and cataloguers approach these descriptors very differently, also indicated by Rolla (2009) and Chan (2011). Cataloguers assigned relatively few terms per books using restricted and established vocabulary by following firm rules, whereas, the end users enjoyed liberty of assigning unlimited terms. In the library environment, the subject heading provider and the end user are typically different individuals, but this study found that more than fifty percent social tags matched with expert created subject headings. Simultaneously, the frequencies of use of social tags that matched with LCSHs terms were higher than the non-matched ones. Moreover, the expert created subject headings were highly ranked in the social tags' lists, where end users more frequently assigned social tags that represented broader or narrower terms than the cataloguer assigned subject headings. In addition, the social tagging represented other aspects that could not be either covered by following the strict subject heading rules or cataloguing rules. Such diverse

impressions can be seen as an access point to the same library collections according to users' interest and opinions.

This study revealed that as a standalone tool neither the controlled vocabulary nor the social tagging practice can work like a satisfactory information retrieval tool. The match between social tags and expert created subject terms indicated that both the cataloguer and the end users were competent to identify basic subject heading for most of the books (Carman, 2009). The end users used LCSHs terms as social tags. At the same time, they are also looking at the library items from a different point of view. The libraries cannot avoid users approach in the current era. The social tags non-matched with LCSHs indicated that some users liked to trace books according to their own interest, need and reading habits. Moreover, this study found that there were more detailed and diverse way of looking in the same book by the end users. Such diverse social tags do not go with the principle and purpose of controlled vocabulary as subject descriptor. In addition, such diverse details cannot be categorized in hierarchical way. But, libraries cannot ignore end users' information seeking point of view, and such preferences should be considered for designing future OPAC to give the maximum benefit of social tags.

Controlled vocabularies may be updated and expanded based on the need of the users after a certain time. It is possible to classify subject terms more and add latest subject terms that suit best to the newly emerged concepts of a discipline. However, the process of updating old catalogue entries is time consuming as well as have economical involvement and labor. Whereas end users may like to tag library resources for their own interest and enrich the metadata by collaborating social tags. It is advisable that the libraries should look at those significantly used social tags, and later the expert cataloguer can adopt those terms in addition to LCSHs terms to enhance subject access. This kind of practice may give more significant outcome for local research libraries or university libraries where the users are more concentrated on a defined number of disciplines. Adopting users' views in addition to controlled vocabulary through social tags may increase the efficiency of information retrieval process in library OPAC.

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. doi:10.1080/19322900903341099

Akther, A., Kim, H.-N., Rawashdeh, M., & El Saddik, A. (2012). Applying Latent Semantic Analysis to Tag-Based Community Recommendations. In L. Kosseim & D. Inkpen (Eds.), *Advances in Artificial Intelligence*, Lecture Notes in Computer Science (Vol. 7310, pp. 1–12).

Ames, M., & Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07 (pp. 971–980). New York, NY, USA: ACM. doi:10.1145/1240624. 1240772

Anderson, J. D., & Hofmann., M. A. (2006). A fully faceted syntax for Library of Congress Subject Headings. *Cataloging & Classification Quarterly*, *43*(1), 7–38.

Anfinnsen, S., Ghinea, G., & de Cesare, S. (2011). Web 2.0 and folksonomies in a library context. *International Journal of Information Management*, *31*(1), 63–70. doi:10.1016/j.ijinfomgt.2010.05.006

Ankolekar, A., Krötzsch, M., Tran, T., & Vrandečić, D. (2008). The two cultures: Mashing up Web 2.0 and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, *6*(1), 70–75. doi:10.1016/j.websem.2007.11.005

Antell, K., & Huang, J. (2008). Subject searching success - Transaction logs, patron perceptions, and implications for library instruction. *Reference & User Services Quarterly*, *48*(1), 68–76.

Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a twenty-first century library catalog. *Information technology and libraries*, *25*(3), 128–139.

Arakji, R., Benbunan-Fich, R., & Koufaris, M. (2009). Exploring contributions of public resources in social bookmarking systems. *Decision Support Systems*, *47*(3), 245–253. doi:10.1016/j.dss.2009.02.007

Bates, J., & Rowley, J. (2011). Social reproduction and exclusion in subject indexing: A comparison of public library OPACs and LibraryThing folksonomy. *Journal of Documentation*, *67*(3), 431–448. doi:10.1108/00220411111124532

Bearman, D., & Trant, J. (2005). Social Terminology Enhancement through Vernacular Engagement. *D-Lib Magazine*, *11*(09). doi:10.1045/september2005-bearman

Bertot, J. C., Berube, K., Devereaux, P., Dhakal, K., Powers, S., & Ray, J. (2012). Assessing the Usability of Worldcat Local: Findings and Considerations. *Library Quarterly*, *82*(2), 207–221. doi:10.1086/664588

Bibliographic Services Task Force. (2005). *Rethinking how we provide bibliographic services for the University of California*. The University of California Libraries.

Brase, C. H., & Brase, C. P. (2006). *Understandable Statistics* (8th ed.). Brooks Cole.

Broughton, V. (2011). *Essential Library of Congress Subject Headings* (1st ed.). Neal-Schuman Publishers, Inc.

Carla S., R. (2010). Social Bookmarking in Academic Libraries: Trends and Applications. *The Journal of Academic Librarianship*, *36*(3), 219–227. doi:10.1016/j.acalib.2010.03.004

Carman, N. (2009). *LibraryThing Tags and Library of Congress Subject Headings: a Comparison of Science Fiction and Fantasy Works*. Victoria University of Wellington. Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/1272

Cataloging Policy and Support Office. (2007). *Library of Congress Subject Headings Pre- vs. Post-Coordination and Related Issues:* Report for Beacher Wiggins, Director, Acquisitions & Bibliographic Access Directorate, Library Services, Library of Congress.

Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1461–1464. doi:10.1073/pnas.0610487104

Chan, L. M. (2005). *Library of Congress Subject Headings: Principles and Application* (4th ed.). Libraries Unlimited.

Chan, L. M. (2011). Social bookmarking and subject indexing. In P. Landry, L. Bultrini, E. T. O'Neill, & S. K. Roe (Eds.), *Subject Access: Preparing for the Future*, IFLA Series on Bibliographic Control (Vol. 42, pp. 127–142). De Gruyter.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2010). *Research Methods, Design, and Analysis, 11th Edition* (11th ed.). Allyn and Bacon.

Conradi, E. (2009). *to_be_classified: A Facet Analysis of a Folksonomy* (Master thesis). Retrieved from https://oda.hio.no/jspui/handle/10642/313

Creswell, J. W. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Sage Publications, Inc.

Da, N., Ke-qing, H., Rong, P., & Jian-xiao, L. (2011). A Multi-dimensional Social Tagging Method for Semantic Web Services. *Services Computing (SCC), 2011 IEEE International Conference on* (pp. 735 –736). doi:10.1109/SCC.2011.49

Drabenstott, K. M., Simcox, S., & Fenton, E. G. (1999). End-user understanding of subject headings in library catalogs. *Library Resources & Technical Services*, *43*(3), 140–160.

Emanuel, J. (2011). Usability of the VuFind Next-Generation Online Catalog. *Information Technology and Libraries*, *30*(1), 44–52.

Feinberg, M. (2011). An examination of authority in social classification systems. *Advances in Classification Research Online*, *17*(1), 1–11. doi:10.7152/acro.v17i1.12490

Fugelstad, P., Dwyer, P., Filson Moses, J., Kim, J., Mannino, C. A., Terveen, L., & Snyder, M. (2012). What makes users rate (share, tag, edit...)?: predicting patterns of participation in online communities. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12 (pp. 969–978). New York, NY, USA: ACM. doi:10.1145/2145204.2145349

Furner, J. (2007). User tagging of library resources: Toward a framework for system evaluation. *World library and information congress: 73rd IFLA general conference and council*. Durban, South Africa.

Gabriela, G. (2009). To use or not to use web 2.0 in higher education? *Procedia - Social and Behavioral Sciences*, *1*(1), 478–482. doi:10.1016/j.sbspro.2009.01.087

Garcia-Plaza, A. P., Zubiaga, A., Fresno, V., & Martinez, R. (2012). Reorganizing clouds: A study on tag clustering and evaluation. *Expert Systems with Applications*, *39*(10), 9483–9493. doi:10.1016/j.eswa.2012.02.108

Golder, S. A., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. doi:10.1177/0165551506062337

Good, B. M., Tennis, J. T., & Wilkinson, M. D. (2009). Social tagging in the life sciences: characterizing a new metadata resource for bioinformatics. *Bmc Bioinformatics*, *10*. doi:10.1186/1471-2105-10-313

Halpin, H., & Place, B. (2007). The complex dynamics of collaborative tagging. *In Proceeding of International World Wide Web Conference* (pp. 211–220). ACM Press.

Hassan-Montero, Y., & Herrero-Solana, V. (2006). Improving tag-clouds as visual information retrieval interfaces. *International Conference on Multidisciplinary Information Sciences and Technologies* (pp. 25–28).

Hollenstein, L., & Purves, R. (2012). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, (1), 21–48. doi:10.5311/josis.v0i1.13

Islam, K. M. S. (2008). *Essentials of Cataloguing and Classifications*. Dhaka: New Progati.

Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., & Stumme, G. (2008). Discovering shared conceptualizations in folksonomies. *Web Semantics: Science, Services and Agents on the World Wide Web*, *6*(1), 38–53. doi:10.1016/j.websem.2007.11.004

Johnson, S. L. (2010). Review of LibraryThing. *Journal of Library & Information Services in Distance Learning*, *4*(4), 220–221. doi:10.1080/1533290X.2010.524832

Kajewski, M. A. (2007). Emerging technologies changing our service delivery models. *Electronic Library*, *25*(4), 420–429. doi:10.1108/0264047070779835

Kakali, C., & Papatheodorou, C. (2010). Exploitation of folksonomies in subject analysis. *Library & Information Science Research*, *32*(3), 192–202. doi:10.1016/j.lisr.2010.04.001

Kim, Y.-M., & Abbas, J. (2010). Adoption of Library 2.0 Functionalities by Academic Libraries and Users: A Knowledge Management Perspective. *The Journal of Academic Librarianship*, *36*(3), 211–218. doi:10.1016/j.acalib.2010.03.003

Kipp, M. E. I. (2006). Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator, and Intermediary Keywords. *Proceedings of the Canadian Association for Information Science*. York University, Toronto. Retrieved from http://eprints.rclis.org/bitstream/10760/8771/1/mkipp-caispaper.pdf

Kipp, M. E. I., & Campbell, D. G. (2006). Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. Retrieved from http://eprints.rclis.org/handle/10760/8720#.T9IDPJjpURo

Kipp, M. E. I., & Campbell, D. G. (2010). Searching with Tags: Do Tags Help Users Find Things? *Knowledge Organization*, *37*(4), 239–255.

Kirtland, M., & Cochrane, P. (1982). Critical Views of LCSH - Library of Congress Subject Headings A Bibliographic and Bibliometric Essay. *Cataloging & Classification Quarterly*, *1*(2-3), 71–94. doi:10.1300/J104v01n02_04

Kiu, C.-C., & Tsui, E. (2011). TaxoFolk: A hybrid taxonomy–folksonomy structure for knowledge classification and navigation. *Expert Systems with Applications*, *38*(5), 6049–6058. doi:10.1016/j.eswa.2010.11.014

Lambiotte, R., & Ausloos, M. (2006). Collaborative tagging as a tripartite network. *Lecture Notes in Computer Science*, *3993*, 1114–1117. doi:10.1007/11758532_152

Landry, P., Bultrini, L., O'Neill, E. T., & Roe, S. K. (Eds.). (2011). *Subject Access: Preparing for the Future*. IFLA Series on Bibliographic Control (Vol. 42). De Gruyter.

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

Library of Congress. (2008). *Subject cataloguing mannual: subject headings*. Washington DC: Library of Congress.

Library of Congress Authorities. (2011). Retrieved February 24, 2012, from http://authorities.loc.gov/help/auth-faq.htm

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in Educational Research: From Theory to Practice* (2nd ed.). Jossey-Bass.

Lopes, M. I., & Beall, J. (Eds.). (1999). *Principles underlying subject heading languages* (Vol. 21). K G Saur Verlag Gmbh & Co.

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. doi:10.1177/0165551510386173

Lund, W., & Washburn, A. (2009). Patrons Cataloging? The Role and Quality of Patron Tagging in Item Description. *Proceedings of the Fourteenth National Conference of the Association of College and Research Libraries* (pp. 263 – 271). Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/national/seattle/papers/263.pdf

Marvasti, A. B. (2003). *Qualitative Research in Sociology*. Sage Publications Ltd.

Matthews, Brian, Jones, C., Puzon, B., Moon, J., Tudhope, D., Golub, K., & Nielsen, M. L. (2010). An evaluation of enhancing social tagging with a knowledge organization system. *Aslib Proceedings*, *62*(4-5), 447–465. doi:10.1108/00012531011074690

Matthews, Bob, & Ross, L. (2010). *Research Methods: a practical guide for the social sciences*. Longman.

Matusiak, K. K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services*, *22*(4), 283–298.

McTavish, J. (2011). A domain-analytic perspective on sexual health in LCSH and RVM. *NASKO*, *3*(1), 83–93.

Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, *5*(1), 5–15. doi:10.1016/j.websem.2006.11.002

Miller, J., & McCarthy, S. (Eds.). (2010). *Sears List of Subject Headings* (20th ed.). Hw Wilson Co.

Miotto, R., & Orio, N. (2011). Accessing Music Digital Libraries by Combining Semantic Tags and Audio Content. In M. Agosti, F. Esposito, C. Meghini, & N. Orio (Eds.), *Digital Libraries and Archives*, Communications in Computer and Information Science (Vol. 249, pp. 26–37). Springer Berlin Heidelberg.

Morrison, J. P. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing & Management*, *44*(4), 1562–1579. doi:10.1016/j.ipm.2007.12.010

Munk, T. B., & Mork, K. (2007). Folksonomies, tagging communities, and tagging strategies - An empirical study. *Knowledge Organization*, *34*(3), 115–127.

Needleman, M. (2007). Web 2.0/Lib 2.0—What Is It? (If It's Anything at All). *Serials Review*, *33*(3), 202–203.

O'Reilly, T. (2005). What is Web 2.0—Design patterns and business models for the next generation of software. Retrieved May 8, 2012, from http://oreilly.com/web2/archive/what-is-web-20.html

Olivier, G. (2011). Folksonomies: Spontaneous crowd sourcing with online early detection potential? *Futures*, (0). doi:10.1016/j.futures.2011.10.008

Olson, H. A. (2000). Difference, Culture and Change: The Untapped Potential of LCSH. *Cataloging & Classification Quarterly*, *29*(1-2), 53–71. doi:10.1300/J104v29n01_04

Peters, I. (2009). *Folksonomies. Indexing and Retrieval in Web 2.0* (1st ed.). De Gruyter.

Peterson, E. (2006). Beneath the Metadata: Some Philosophical Problems with Folksonomy. *D-Lib Magazine*, *12*(11). doi:10.1045/november2006-peterson

Pirmann, C. M. (2011). Using Tags to Improve Findability in Library OPACs: A Usability Study of LibraryThing for Libraries. Retrieved from http://hdl.handle.net/2142/18918

Quintarelli, E. (2005). *Folksonomies: power to the people*. Paper presented at ISKO Italy-UniMIB meeting, Milan, Italy, June 24, 2005. Retrieved from http://www.iskoi.org/doc/folksonomies.htm

Rolla, P. J. (2009). User Tags versus Subject Headings. *Library Resources & Technical Services*, *53*(3), 174–184.

Sadeh, T. (2007). Time for a change: new approaches for a new generation of library users. *New Library World*, *108*(7/8), 307–316. doi:10.1108/03074800710763608

Smith, T. (2007). *Cataloging and You: Measuring the Efficacy of a Folksonomy for Subject Analysis* (Conference Paper). School of Information Resources and Library Science and Learning Technologies Center, University of Arizona. Retrieved from http://arizona.openrepository.com/arizona/handle/10150/106434

Spalding, T. (2007). When tags work and when they don't: Amazon and LibraryThing « The Thingology Blog. Retrieved from http://www.librarything.com/blogs/thingology/2007/02/when-tags-work-and-when-they-dont-amazon-and-librarything/

Srinivasan, R., Boast, R., Furner, J., & Becvar, K. M. (2009). Digital Museums and Diverse Cultural Knowledges: Moving Past the Traditional Catalog. *Information Society*, *25*(4), 265–278. doi:10.1080/01972240903028714

Steele, T. (2009). The new cooperative cataloging. *Library Hi Tech*, *27*(1), 68–77. doi:10.1108/07378830910942928

Stone, A. T. (2000). *The Lcsh Century: One Hundred Years With the Library of Congress Subject Headings System* (1st ed.). Routledge.

Stvilia, B., & Joergensen, C. (2010). Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr. *Journal of the American Society for Information Science and Technology*, *61*(12), 2477–2489. doi:10.1002/asi.21432

Thomas, M., Caudle, D. M., & Schmitz, C. M. (2009). To tag or not to tag? *Library Hi Tech*, *27*(3), 411–434. doi:10.1108/07378830910988540

Tonkin, E., Corrado, E. M., Moulaison, H. L., Moulaison, H. L., Resmini, A., & Zhang, Q. (2008). Collaborative and Social Tagging Networks. *Ariande*, *54*. Retrieved from http://www.ariadne.ac.uk/issue54/tonkin-et-al

Torres, D., Diaz, A., Skaf-Molli, H., & Molli, P. (2011). Semdrops: A Social Semantic Tagging Approach for Emerging Semantic Data. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 340 –347). doi:10.1109/WI-IAT.2011.51

Trant, J. (2006). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia*, *12*(1), 83–105. doi:10.1080/13614560600802940

Trant, Jennifer. (2009). Studying Social Tagging and Folksonomy: A Review and Framework. *Journal of Digital Information*, *10*(1), 1–44.

Tripathi, M., & Kumar, S. (2010). Use of Web 2.0 tools in academic libraries: A reconnaissance of the international landscape. *The International Information & Library Review*, *42*(3), 195–207. doi:10.1016/j.iilr.2010.07.005

van Hooland, S. (2006). From spectator to annotator: possibilities offered by user-generated metadata for digital cultural collections. *Presented at the CILIP Cataloguing Indexing Group Annual Conference, University of East Anglia, UK.*

Vaughan, J. (2010). *Insights into the Commons on Flickr. Portal: Libraries and the Academy*. Las Vegas: Johns Hopkins University Press.

Voorbij, H. (2012). The value of LibraryThing tags for academic libraries. *Online Information Review*, *36*(2), 3–3.

Voorbij, H. J. (1998). Title keywords and subject descriptors: A comparison of subject search entries of books in the humanities and social sciences. *Journal of Documentation*, *54*(4), 466–476. doi:10.1108/EUM0000000007178

Voß, J. (2007). Tagging, Folksonomy & Co-Renaissance of Manual Indexing? *Proceedings of the International Symposium of Information Science* (Vol. 234–254). Cologne, Germany.

Weaver, M. (Trans.). (2007). Contextual metadata: faceted schemas in virtual library communities. *Library Hi Tech*, *25*(4), 579–594. doi:10.1108/07378830710840527

Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books.

Westcott, J., Chappell, A., & Lebel, C. (2009). LibraryThing for libraries at Claremont. *Library Hi Tech*, *27*(1), 78–81. doi:10.1108/07378830910942937

Wetterstrom, M. (2008). The Complementarity of tags and LCSH - a tagging experiment and investigation into added value in a New Zealand library context. *The New Zealand Library and Information management Journal*, *50*(4), 292–306.

Working group on the future of bibliographic control. (2008). *On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control* (pp. 1–44). Retrieved from http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

Yang, L., Sun, T., Zhang, M., & Mei, Q. (2012). We know what @you #tag: does the dual role affect hashtag adoption? *Proceedings of the 21st international conference on World Wide Web*, WWW '12 (pp. 261–270). New York, NY, USA: ACM. doi:10.1145/2187836.2187872

Yi, K. (2008). A conceptual framework for improving information retrieval in folksonomy using Library of Congress subject headings. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1–6. doi:10.1002/meet.2008.1450450368

Yi, K. (2010). A Semantic Similarity Approach to Predicting Library of Congress Subject Headings for Social Tags. *Journal of the American Society for Information Science and Technology*, *61*(8), 1658–1672. doi:10.1002/asi.21351

Yi, K., & Chan, L. M. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, *65*(6), 872–900. doi:10.1108/00220410910998906

Appendix-I: List of books

Book Code	Title and author	Year of publication	Dewey Decimal Class number
Book 1	The trouble with physics : the rise of string theory, the fall of a science, and what comes next by Lee Smolin	2006	530.14
Book 2	E=mc² : : a biography of the world's most famous equation by David Bodanis	2000	530.11
Book 3	The fabric of the cosmos : space, time, and the texture of reality by Brian Greene	2004	523.1
Book 4	A brief history of time by Stephen Hawking	1988	523.1
Book 5	Zero: the biography of a dangerous idea by Charles Seife	2000	513
Book 6	The red queen: sex and the evolution of human nature by *Matt Ridley*	2003	599.93
Book 7	The greatest show on earth: the evidence for evolution by Richard Dawkins	2009	576.8
Book 8	Electric universe: the shocking true story of electricity by David Bodanis	2005	537
Book 9	Physics of the Impossible: A Scientific Exploration into the World of Phasers, Force Fields, Teleportation, and Time Travel by Michio Kaku	2008	530
Book 10	Nature's numbers: the unreal reality of mathematical imagination by Ian Stewart	1995	510
Book 11	Introduction to Electrodynamics (3rd Edition) by David J. Griffiths	1999	537.6
Book 12	Biochemistry by Lubert Stryer	1995	574
Book 13	Biology by Neil A. Campbell	1987	570
Book 14	CRC Handbook: Chemistry & Physics by David R. Lide	1999	540
Book 15	Molecular biology of the cell (4th edition) by Bruce Alberts	2002	571.6
Book 16	QED: the strange theory of light and matter by Richard Feynman	1985	539.7
Book 17	Genome: the autobiography of a species in 23 chapters by Matt Ridley	1999	599.93
Book 18	The elements: a visual exploration of every known atom in the universe by Theodore W. Gray	2009	546
Book 19	Concrete mathematics: a foundation for computer science by Ronald L. Graham, Donald E. Knuth, Oren Patashnik.	1994	510
Book 20	Fundamentals of Physics, (6ᵗʰ edition) by David Halliday	2001	530