LibraryThing tags and Library of Congress Subject Headings: a comparison of Science Fiction and Fantasy works

by

Nicholas Carman

Submitted to the School of Information Management,
Victoria University of Wellington
in partial fulfilment of the requirements for the degree of
Master of Library and Information Studies

June 2009

Acknowledgements

I owe thanks to my supervisor Alastair Smith for the many meetings and the high quality of the feedback he gave me.

Thanks are also owed to Deborah Laurs and Dennis Dawson at the Student Learning Support service who gave me help with the writing and interpreting the statistics.

Contents

Abstract

Introduction

This study examines the extent to which LibraryThing tags match their equivalent Library of Congress subject headings and looks at whether they offer any additional information about the subject matter of the books to which they are applied.

Method

This study has a largely quantitative methodology with some qualitative aspects. The researcher harvested tags from ten books in the genres of science fiction and fantasy. The tags were then classified into categories created by the researcher and examined using descriptive statistics inside Excel.

Findings

The most frequently used tags were those that matched the Library of Congress subject headings, but there were a significant number of non-matching tags that offered useful additional information about the books in the sample.

Conclusion

Library of Congress subject headings mostly identify the basic genres that the books in the sample belonged to, but added little additional information. Integrating tagging into library OPACs would create more opportunities for library users to find books in which they are interested.

Keywords

LibraryThing, tagging, LCSH, science fiction, fantasy.

Introduction

Access to bibliographic information has traditionally been through controlled vocabularies such as LCSH (Library of Congress Subject Headings), which meant that library users conducted searches using the terms created by the cataloguers. Library patrons often find these vocabularies hard to use and have tended to prefer the simplicity of Google searches (Rethlefsen, 2007).

The rise of websites such as LibraryThing has seen ordinary users creating their own terms for categorizing and accessing information. While the Library of Congress prefers terms such as *'Fantasy—Juvenile literature'* for books like 'The Subtle Knife', the users have chosen terms such as *'Steampunk'* and *'sci-fi'*.

There has been a corresponding move in the Library and Information field towards offering new services based on this new technology to better serve users and to base these services on 'local' knowledge and a better understanding of users' needs (Weaver, 2007, p. 579).

This study looks at what can be learned from the tags used by LibraryThing users and how they might be used to improve cataloguing practice, and OPAC interfaces.

What is LibraryThing?

LibraryThing (www.librarything.com) is a website that allows its users to catalogue their books online. LibraryThing users have a variety of ways to sort their book collections including by adding descriptive tags to each book which can then be shared with other users ("LibraryThing| Catalogue your books online," 2008). If a tag is used by more than one user to describe a particular book in their collection it becomes "popular", and the more people that use that tag (for the same book) the more popular it becomes. The tags are presented in tag clouds which vary the size of the tags according to the ranking of the tags (the popular tags are displayed in larger text than the unpopular ones). These tags constitute a folksonomy, which is a user-created cataloguing system which relies on tag popularity as a measure of value (Kroski, 2007, p. 94).

LibraryThing allows its users to access a great deal of information about these user-created tags. While the default option for any given book is to only show the top 30-50 tags, it is possible to request that all of the tags should be displayed, and also to give a count of the number of times a tag has been used. This study aims to harvest and use this readily available information.

Research Question

To what extent do LibraryThing book tags match their equivalent Library of Congress subject headings and do they offer any additional information about the subject matter of the books?

Sub questions

1. To what extent and how often do LibraryThing tags represent the same concepts as Library of Congress subject headings?
2. How often do LibraryThing tags represent concepts not included in the Library of Congress subject headings (and vice-versa)?

2a. What sorts of concepts are represented in the LibraryThing tags that are not included in neither the Library of Congress subject headings nor the OCLC records?

3. Are the Library of Congress subject heading concepts highly ranked in the LibraryThing tag lists?

Research Statement

The aim of this research was to determine how closely LibraryThing tags match Library of Congress subject headings and how much additional subject information the LibraryThing tags provide.

Purpose Statement

This research was intended to provide feedback to the designers of future folksonomy schemes to help improve the designs of their schemas. It may also offer useful lessons and insights to improve library cataloguing practice such as the ability to add new descriptors to existing classification schemes or the advisability of adding a tag layer to library catalogues.

Limitations

This was a small scale project limited by the time and resources available to the researcher. This means that the sample was from necessity quite small and that caution must be used when applying this study's conclusions to folksonomies as a whole.

This study was intended to be a quantitative study; however the need for the researcher to assign research codes to tags from LibraryThing introduced a qualitative element.

Delimitations

To make this research feasible, given the time and resource constraints, it focussed on tags from the genres of fantasy and science fiction in LibraryThing. These genres were chosen because the researcher has read widely within them which made the interpretation and coding of the LibraryThing tags much easier.

This study only took the top 30-50 tags for each book in LibraryThing (those made available by default to LibraryThing users). This study assumes that most casual users would not bother to examine all of the tags available for any given book, and that the top 30-50 tags constitute the most significant access points.

Quantitative methodology

The methodology of this research was quantitative in nature; it harvested tags from LibraryThing and divided them into nominal variables which were counted and subjected to basic statistical analysis (Pickard, 2007, p. 7).

Data was harvested at the beginning of this research for analysis, and collected at the end to describe the results; which fits with quantitative methodology (Sogunro, 2002, p. 5).

Quantitative methodologies use statistical analysis which can be checked and confirmed by other researchers (Sogunro, 2002, p. 5). This approach closely fits the content analysis method intended to be used in this research.

Quantitative methodology is better suited to dealing with smaller numbers of variables (Sogunro, 2002, p. 5). This study has tight limitations and de-limitations, which make it well suited to the quantitative methodology.

Literature Review

Why study folksonomies?

Folksonomies have been a topic of interest to the Library and Information profession because they are at the heart of a 21st century problem; how does one find and organize information on the Internet?

The number of web pages and resources on the internet is simply enormous and cataloguing them all using the classic taxonomies and cataloguing methods is impossible (Kroski, 2007, p. 101). Rather than wait for librarians to catch up internet users have taken it upon themselves to organize the internet, and according to Kroski they are succeeding (Kroski, 2007, p. 91). However the degree to which they have succeeded is open to debate and opinions vary from author to author.

The internet has been described as an environment with no rules which favours self-organizing systems where the whole is greater than the sum; folksonomies are a prominent example of such systems (Kipp & Campbell, 2006, p. 1).

The value of folksonomies as tools for organizing information has been hotly disputed. They have many qualities that make them useful including an ability to scale up that is unmatched by conventional classification systems (Kipp & Campbell, 2006, p. 2). A folksonomy can become massive, with large numbers of people contributing, compared to a limited number of cataloguers.

Folksonomies also boast a high degree of granularity (users can add a lot of detailed information), and are largely maintained by their users; and therefore cheap to run. Folksonomies are also noted for their high degree of currency (reflecting user concerns) (Kroski, 2007, p. 97).

The shortcomings of folksonomies versus cataloguing taxonomies are neatly described by Peter Morville:

"When it comes to findability, their inability to handle equivalence, hierarchy, and other semantic relationships causes them to fail miserably at any significant scale"(Morville, 2005, p. 139).

These areas of weaknesses for folksonomies are where most controlled vocabularies excel. However Morville is of the opinion that it is not necessary to choose between controlled vocabularies and folksonomies.

Pace Layering

Morville suggests that it is possible to combine these systems by using a concept called Pace Layering (Morville, 2005, p. 141). Pace layering is a theory that proposes that aspects of society change at different rates. When applied to information management the folksonomies serve as a fast moving top layer which has high currency and reflects user concerns, while controlled vocabularies serve as the solid ordered foundation (Morville, 2005, p. 141).

Morville's view is shared by Mika, who views folksonomies as "lightweight semantic structures" where meaning develops over time (Mika, 2007, p. IX). Mika attempts to demonstrate the viability of extracting information structure from folksonomies (Mika, 2007, p. 203). This would need to be done on a regular basis as the supporting community changes over time.

Folksonomies add value

There are already examples of how folksonomies have been used to add valuable metadata to archival collections that have already been catalogued. The Library of Congress in the United States made a number of its photographic collections available on Flickr in 2008. They were delighted by the spontaneous emergence of user-created categories such as 'Rosie the Riveter' from a photo archive from the 1930-1940's time period (Oates, 2008).

In New Zealand the War Art Online website has also allowed users to create tags to help describe their photographic collection:

"It was recognised that archival description of the artworks would not be able to cover many features of the artworks. Therefore we were keen to enable the public to add comments that would fill out the description of the artworks or provide information on the subject matter or the artists" (Jennings, 2008, p. 6).

Folksonomies are already in use

While folksonomies are not an ideal system for information retrieval they are already being put to use by librarians helping users locate resources on the internet. Rethlefsen quotes a school librarian from Australia: “Librarians are so careful about what is bought for the print collection, but then we watch our googlers race around the internet among unedited and ill-founded and repetitive single-page resources. [Using del.icio.us] is our attempt to select suitable material, to post it without delay, and to provide access points and comments on content” (Rethlefson, 2007b). Therefore it makes sense to take a closer look at folksonomies to find out how we can improve them.

Tagging in library catalogues

There has already been some discussion of the impetus towards the use of tagging systems in library catalogues, as well as the practical implications of this trend.

Steele talks about the growing expectation from library users that they will be able to interact with the catalogue, not just passively receive information delivered by the cataloguers. He observes that many catalogues do not as yet allow this, and also require the user to be familiar with thesauri such as LCSH to search successfully (Steele, 2009, p. 68).

Steele discusses the ways in which tags might complement LCSH; for example a user could search the tags and if too many or too few results are found use LCSH to broaden or narrow their search. Steele also points out that while tagging provides more terms, browsing subject headings will give the most complete list of materials that a library holds on a particular topic (Steele, 2009, p. 72). These observations fit in with Morville’s ‘pace layering’ concept; tags compensate for the weaknesses of LC subject headings and vice-versa.

Steele says that LCSH wins in terms of longevity; it has been around far longer than tagging and covers a large number of resources that are unlikely to ever be tagged. He also makes the point that an ever-changing body of tags could be unhelpful for older and larger static collections (Steele, 2009, p. 72). However tags could still play a useful role for these collections when one considers that users needs and interests change over time.

LibraryThing for Libraries

The creators of LibraryThing have developed LibraryThing for Libraries (LTFL) which can be used to provide tags for library OPACs. LTFL works as on overlay on the OPAC that queries LibraryThing data by ISBN as the user selects a specific record. The tags provide the user with tags related to the book and suggests other similar books in the library's collection (Westcott, Chappell, & Lebel, 2009, p. 78).

LTFL was viewed as a low cost, low risk way of implementing next generation catalogue concepts at Claremont University in the United States (Westcott et al., 2009, p. 79).

The library found LTFL useful when searching for foreign language publications which lacked LC subject headings, and finding further related items via the tags (and also general browsing). An example cited concerned a student who was writing on themes of abuse in fiction. The student was able to generate a reading list using the tag *'The Color Purple'* (Westcott et al., 2009, p. 80). Anecdotes such as these demonstrate the usefulness of tags in regard to readers' advisory services.

The authors noted that LTFL was extraordinarily easy to use and implement and draws on a huge pool of book lover's tags and suggestions. This has saved the library from having to create their own tags from scratch (Westcott et al., 2009, p. 80).

Mendes makes the point that leveraging metadata from third parties is not a new issue for most libraries, as library records are often created by third party vendors (Mendes, Quinonez-Skinner, & Skaggs, 2009, p. 32).

Furthermore the LTFL development team is apparently looking at integrating LC subject headings into the tag clouds, as well as LCCN numbers (Westcott et al., 2009, pp. 81-82). The LTFL development team has decided that LC subject headings and tags would make a worthwhile combination.

Mendes et al examine the implementation of LTFL at the Oviatt Library at California State University, Northridge (Mendes et al., 2009, p. 30).

Mendes selected twenty-one books from the library catalogue and compared their LibraryThing tags against their equivalent LC subject headings (Mendes et al., 2009, p. 37).

Mendes et al found that the number of LC headings varied from book to book, but that on average there were more tags than there were headings. They also found that for every book discovered through LCSH four more could be found via the LibraryThing tags, although the authors noted that they had not tried to judge the relevance of these additional books. Mendes conceded that a larger study would be required to confirm these findings (Mendes et al., 2009, p. 38). Nor did their study attempt to judge the relevance of the LibraryThing tags in their sample.

LTFL is a relatively new development and these articles are only preliminary studies. Nobody has yet made a detailed study of the LTFL service, and exactly how the tags and LC subject headings interact, and how LibraryThing tags might add value for library users.

Tag clouds

Since this research extracted data from tag clouds it is worth taking a closer look at them as a search tool, and also as a way of presenting data to users.

According to Sinclair and Cardew-Hall there is some debate as to the usefulness of tag clouds, and it has been suggested that they are skewed towards popular tags (Sinclair & Cardew-Hall, 2008, p. 17).

These authors undertook a combined qualitative and quantitative study of tag clouds and found their strengths were similar to tagging in that they were good for browsing, but also that they provided a visual summary of the database, and that it took less effort to scan a tag cloud than to come up with a search query (Sinclair & Cardew-Hall, 2008, p. 27). Also a tag cloud could help a searcher formulate a search query by suggesting synonyms.

However tag clouds are limited by text size; if a tag is too small the tag will be unreadable, and likewise if it is too large. Therefore it isn't realistic for a tag cloud to show all the tags on any one item. According to Tim Spalding (the founder of LibraryThing) a tag cloud should "convey some of the peaks and troughs" of the available tags but only occupy its allotted area on the webpage and be legible (Rethlefson, 2007a).

This weakness was confirmed in a study by Sinclair and Cardew. They found that tag clouds were poor at specific searching and more importantly they obscure around half of the data from the user (Sinclair & Cardew-Hall, 2008, p. 28). This is especially true of LibraryThing book tag clouds as by default only the top 30-50 tags are displayed (Rethlefson, 2007a).

Methodologies used to investigate folksonomies

Studies of folksonomies can be roughly divided into two approaches; those that chose to study tags harvested from working folksonomies on the internet, and those that studied tags created by volunteers. The research methods range from purely quantitative, to qualitative, plus mixed methods.

Guy and Tonkin

Guy and Tonkin performed what they called a "brief, simple and relatively unscientific" investigation of a sample of tags taken from Del.icio.us and Flickr as part of an exploratory study of folksonomies (Guy & Tonkin, 2006, p. 10).

Guy and Tonkin found that quality of the tags they had harvested was quite variable, with up to 40% of the Del.icio.us tags, and 28% of the Flickr tags being misspelt. They also found problems with the use of plural word-forms, and symbols such as # (Guy & Tonkin, 2006, p. 5).

Guy and Tonkin felt that these quality problems could be controlled by providing software generated check-lists, tag suggestions, and spelling guidance to folksonomy users (Guy & Tonkin, 2006, p. 7).

Guy and Tonkin's research was (as they pointed out) not that rigorous and was exploratory in nature. They conceded that trying to control user tagging might very well cause folksonomies to lose their information value (Guy & Tonkin, 2006, p. 11).

Guy and Tonkin's approach can be contrasted by that taken by Weaver, who asked a group of public library users to tag a selection of library books as part of a reader's advisory survey (Weaver, 2007, p. 579).

Lin et al

Lin's study looked at whether social classification and traditional content indexing represent content in different ways (Lin, Beaudoin, Bui, & Desai, 2006, p. 3).

Lin tried to answer this question by comparing MeSH (Medical Subject Heading Terms) terms and Connotea tags, plus the results of software-based indexing of the titles. Connotea (http://www.connotea.org/) is an online service allowing the user to create and share citations (Lin et al., 2006, p. 4).

Lin's data showed a small overlap (11%) between the MeSH headings and the Connotea tags, but a greater overlap between the tags and the title-based indexing (Lin et al., 2006, p. 6).

Lin observed that rather than try to describe the entire document, tag users appeared to focus their tags on the aspects of the article of most interest to them (Lin et al., 2006, p. 6).

Lin also suggested that the greater match between the tags and the title-based indexing indicated that users were choosing title based terms to describe articles and that this was consistent with the principle of least effort (Lin et al., 2006, p. 7).

Both Lin and Bruce's studies (see below) looked at narrowly focused academic controlled vocabularies and citation systems which were made for more serious pursuits. It is harder to compare these systems to LCSH and LibraryThing which have a far broader focus. For example, LibraryThing users are likely to create a wider range of tags other than title based tags (unlike CiteUlike or Connotea users).

Weaver

Weaver compared the tags produced by his readers against the tags for the same books from LibraryThing and observed that the tags from the reader's survey were more precise, and produced more useful information (Weaver, 2007, p. 587). Weaver felt that this was due to the way that LibraryThing favours frequently used terms over infrequently used terms. He contended that it was the infrequently used terms that are the most informative (Weaver, 2007, p. 587).

However Weaver observed that both the LibraryThing tags and the tags produced by his survey participants were better than the subject headings used in public libraries (Weaver, 2007, p. 588).

Weaver's research involved a small sample of tags, and a small group of survey participants. The high quality and precision of his tags could be put down to the controlled nature of his experiment, and the well-educated library patrons who took part.

Wetterstrom

Wetterstrom's research involved giving research participants a selection of books from the National Library to tag. The tags were then compared to the books LC subject headings (Wetterstrom, 2007, p. 18).

Wetterstrom found that while there were relatively few matches to the LC subject headings the tags generally complemented the LC headings by providing additional points of access. He concluded that the tags and the LC subject headings could co-exist with the tags serving as a bridge between the two (Wetterstrom, 2007, p. 33).

Wetterstrom suggested that a narrower study of a specific subject with specific users might give a clearer picture of the collaborative value of tags (Wetterstrom, 2007, p. 34).

Bruce

Robert Bruce's study attempted to determine the extent that ERIC subject headings matched the tags produced by users of CiteUlike. ERIC is an education based online index, while CiteUlike (http://www.citeulike.org/) is an online service allowing users to share citations (Bruce, 2008, p. 2).

Bruce's study took the entire CiteUlike tag metadata and used a Perl program to look for exact matches in the corresponding ERIC subject headings (Bruce, 2008, p. 3).

Bruce’s study did not find many matches and he concluded that CiteUlike users were making use of different language than the creators of ERIC. Bruce felt that tags might serve to supplement controlled vocabularies by providing users a means for personal organization (Bruce, 2008, p. 4).

However the analysis was performed by software which only searched for exact matches and Bruce conceded that a more detailed semantic study could turn up more matches. ERIC is also a far smaller vocabulary than LCSH, and written for a different audience.

Conclusion

This research drew a great deal of inspiration from Wetterstrom’s work. LibraryThing provided both specific subjects (genres), plus also a specific group of users. This research also used LC subject headings as a yardstick to measure book tags.

However there is also value in studying tags that have been created ‘in the wild’ as opposed those that are created just for the purposes of research. Harvesting LibraryThing tags helped to provide a realistic idea of how tags are being used in practice.

Theoretical Framework – Zipf’s Power Law & Principle of Least Effort

Zipf’s Principle of Least Effort and Zipf’s Power Law help to explain why the information value of tags in folksonomies and in particular in LibraryThing is so variable.

Zipf contends “that the entire behaviour of an individual is at all times motivated by the urge to minimize effort” (Zipf, 1965, p. 3). According to Zipf an individual will attempt find a solution to any given problem that involves the least amount of exertion on his part (Zipf, 1965, p. 1). Zipf points out that attempting to minimize the work done in the present often results in more work having to be done on future problems (Zipf, 1965, pp. 5-6).

According to Munk & Mork the creation of Folksonomies is best explained by Zipf’s Power Law which they describe as “a law of nature for complex systems” (Munk & Mork, 2007, p. 19). By this Munk & Mork mean that a small subset of tags is heavily used while the rest are only used infrequently. This phenomenon has been referred to as the 'long tail'.

Munk & Mork attribute this to user laziness and a fixation on basic categories, which seems logical given the vast number of resources available for internet users to tag. Following Zipf's principle of least effort the tag creator will use as little effort as possible to create a tag, while anybody who uses that tag to search the folksonomy will try to squeeze out as much meaning as possible (Munk & Mork, 2007, p. 25).

This laziness manifests as a tendency for tag creators to copy other people's tags or only use the most superficial and basic tags. The tags that do not conform to Zipf's law are often created by people who are more knowledgeable about the thing they are tagging (Munk & Mork, 2007, p. 29). They are more likely to create specific and descriptive tags (Munk & Mork, 2007, p. 29).

Study Design

Choice of sampling model

There were many sampling models to choose from; however there are four models that might have been relevant to this research:

- Random samples:

  Random samples involve listing all of the relevant material that could be sampled from a population and making a selection via a dice roll, roulette wheel etc (Krippendorff, 1980, p. 66). This would have required choosing books at random from the LibraryThing website.

- Stratified samples:

  Stratified sampling identifies strata or sub-groupings in the population, then makes a random sample in each strata (Krippendorff, 1980, p. 66). A possible method would have been to choose ten fiction genres (for example romance, science fiction, and thriller) and then pick a random book from each genre.

- Systematic samples:

Systematic sampling picks items at pre-arranged intervals in a population (for example every tenth item) (Krippendorff, 1980, p. 67). This might have been done by selecting every tenth item from the top books column on the LibraryThing zeitgeist page.

- Cluster samples:

Cluster sampling involves the researcher picking out items which exhibit natural borders or boundaries.

Cluster sampling was chosen over stratified, systematic, and random sampling because of the small size of the sample and the limited time and resources available to the researcher. It was felt that using any of the latter sampling models would result in a small dis-jointed sample of distantly related items from which little meaning could be drawn.

The 'clusters' chosen in this instance were the genres of science fiction and fantasy which the researcher has great familiarity with. The researcher felt that he would be better able to judge the information value of tags from familiar genres. The sample included the tag clouds from ten books; five will be science fiction and five fantasy. The books were chosen by searching for the tags 'science fiction' and 'fantasy' and then picking five titles at random under each tag.

Data Collection

Choosing the units of analysis

The tag data collected came from the tag clouds of the LibraryThing website. There were two types of units collected; syntactical and physical. The syntactical units included the tags, the Library of Congress subject headings, and the OCLC records the subject headings were sourced from. The 'physical' units included the books the tags have been applied to (the context from which the tags are taken).

Data collection and recording

The data collection technique itself was straightforward. The researcher navigated to the specific page in LibraryThing for each book, and clicked on the link that requested the number of users for each tag be displayed.

All of the tags including the user numbers were harvested by use of the copy and paste function into a text editor which stripped the hyperlinks from the text and converted it into a single font all of the same size. Each tag and its accompanying number were spaced out onto their own lines in the plain-text document and tag delimited. This allowed all the tag data to be copied and pasted into an Excel spreadsheet.

In addition a screen shot of the page for each book on LibraryThing was taken by requesting the web browser to print the page as a PDF (via software called CutePDF). These PDF files were saved on the researcher's computer. Each book's tag data was recorded in its own spreadsheet tab prior to analysis.

Library of Congress subject headings for each book were harvested from the WorldCat on OCLC database and recorded on their own spreadsheet tab. A screen shot of each title on the OCLC database was also saved via the CutePDF software.

Both the OCLC and LibraryThing records are subject to change over time due to alterations made by cataloguers and users. PDF copies were made to ensure that the researcher could always refer back to the records as they existed during the data collection process.

Analysis

Coding of tags

Once the data collection was complete the researcher required a tool to make comparisons between the LibraryThing tags and the Library of Congress (LC) subject headings. To that end this study borrowed and adapted the coding used by Wetterstrom in his 2007 research project (Wetterstrom, 2007, pp. 18-19).

Each tag was assigned a code which put it into one of five broad categories designated by the letters A-E. The tags were then given an additional two-letter code (e.g. AT) which put it into a further narrower category. This coding is described below:

A. Match with LC Subject Headings:

- The LibraryThing tag is an exact match for the LC subject headings (A--A)

B. Non-subject Match with OCLC record:

The LibraryThing tag provided information that matched with information given by the OCLC database record and not matched by the LC subject headings.

- Author/Title information (B--AT)
    - The tag gives the name of the author or the name of the book; e.g. *Alistair Reynolds, House of Suns*.
- Publishing details (B--PD)
    - The tag provides publishing information about the book such as a date of publication, publisher, series number etc.
- Format Information (B--FI)
    - The tag refers to the format of the book e.g. *paper back*.

C. Partial match with LC Subject Headings:

The LibraryThing tag was a close but not an exact match with a LC subject heading.

- Cross reference (C--CR)
    - The tag matches a LC subject heading which is cross-referenced to the book's subject heading found in WorldCat-OCLC.
        - Cross references were identified via the 2007 edition of the Library of Congress subject headings manual (*Library of Congress subject headings*, 2007).
- Spelling variation (C--SV)
    - The spelling of the tag differs from the subject heading; e.g. it is the plural of the subject heading etc.
- Tags appear in sub-division (C--SD)
    - The tag appears as a part of one of the LC subject headings, or is inverted compared to the subject heading.

- Related term or different point of view (C--RT)
    - The tag is a synonym (not popular language) of a LC subject heading.
- Popular language (C--PL)
    - The tag is a more popular or less formal version of the LC subject headings, for example acronyms, contractions, phrases, and unconventional combinations of terms.

D. No match with either LC Subject Headings or OCLC record:
The LibraryThing tag gives additional information that is not supplied by the LC subject heading or the OCLC record.

- Plot details (D--PLo)
    - The tag is descriptive of the plot; e.g. *robots, paranoia, dystopia*.
- Format Information (D--FI)
    - The tag refers to the variety of formats the book can be found in (and not mentioned on the OCLC record) e.g. *softcover, hardback, movie* etc.
- Literary Genre Information (D--LG)
    - The tag offers additional detail about the book's literary genre; e.g. *space opera, classic, epic*.
- Place and character names (D--PC)
    - The tag is the name of a character or location (for example a city) within the book.
- Other non-specific info (D--BT)
    - The tag describes book data that doesn't easily fit into any one category, e.g. *Hugo Award Nominee, banned*.
- User specific (D--US)
    - The tag is specific to a particular user e.g. *mustread, wishlist* etc.
- Unclear (D--UC)
    - The tag is cryptic and difficult to interpret.

E. No match between tag and LC subject heading but concept appears in another part of the OCLC record (E--E)

Coding in practice

Each LibraryThing book tag was compared to the LibraryThing subject headings and assigned a code as detailed above.

Observer checking

As the coding was in essence a qualitative exercise a second observer was used to double-check a sample of the classifications made by the researcher.

Pilot study

A pilot study was conducted by the researcher to confirm the feasibility of the sampling method, sample size and the analysis techniques.

Analysis inside Excel

Once the data had been put into Excel and its coding was confirmed by the second observer the researcher used Excel to generate some basic statistics.

Analysis of the Results

Sub-questions 1 & 2

1. To what extent and how often do LibraryThing tags represent the same concepts as Library of Congress subject headings?
2. How often do LibraryThing tags represent concepts not included in the Library of Congress subject headings (and vice-versa)?

It was considered desirable to have a single figure for each book (and the sample as a whole) which could be used to demonstrate how often LibraryThing users' tags replicated concepts in the LC subject headings. This was labelled as the 'LCSH match score'.

Two columns were added to the data collection spreadsheets; 'Code Score' and 'Weighted Score'.

The 'Code Score' column was created by assigning a nominal value to the code categories A-E. Category 'A' was considered to be a exact match with the LC subject headings and was counted as 100%, while category 'C' was deemed a partial match and was assigned a value of 75%. Categories 'B', 'D', and 'E' did not match the LC subject headings are were given a value of 0%.

The 'Weighted Score' was created by multiplying the 'Tag Count' column with the 'Code Score' column.

For example if the tag *'myth'* had a 'tag count' of 36 (been used 36 times by LibraryThing users) and was an exact match with the LC subject headings its weighted score would be 3600. If *'myth'* was only a partial match it would be given a score of 2700.

Once a 'Code Score' and a 'Weighted Score' had been generated for every tag, both of these columns were totalled. The final step was to divide the total 'weighted score' column by total 'tag count' column. This generated a percentage figure which could be used to help answer both sub-questions 1 and 2. This figure was intended to show what percentage of the total tag count (for each book) consisted of exact or partial matches with the LC subject headings. These figures were then totalled to give an overall percentage across the sample. See Figure 1 below for an example of this calculation.

Tags	Tag count	Code	Code Score	Weighted Score
Fantasy fiction	25	A--A	100.00%	2500.00
british	63	C--RT	75.00%	4725.00
Unread	147	D--US	0.00%	0.00
Total	235			7225.00
Average code score=	31%			

Figure 1 Example of relevance weighting test

Data from sub-question 3 (the 'Group Count' column) was used to calculate the overall percentages of tags that were directly equivalent to the LC subject headings (Exact Match), those that were partly equivalent (Partial Match), and those that were completely unrelated (No Match).

In addition the lists of LC headings were checked and a count made of instances where a LC subject heading had no equivalent to any of the collected tags.

Sub-question 2a

> What sorts of concepts are represented in the LibraryThing tags that are not included in either the Library of Congress subject headings or the OCLC records?

The 'D' code categories (those which matched neither the LC subject headings, nor the OCLC records) were counted and sub-totalled for each book. Each book's 'D' code sub-totals were then totalled on a separate spreadsheet to give an overall count of the types of tags found.

In addition the researcher undertook qualitative assessment of the 'D' coded tags. The researcher looked for any themes in the tags for each 'D' code category that revealed an interesting aspect of the book not covered by the LC Subject Headings.

The researcher attempted to judge the relevance of the tags to the books they were applied to, and also to develop a sense of what things LibraryThing users thought were important in each category.

Individual tags were also investigated on the LibraryThing website to further explore their meaning, and to see where what other books a user might find by clicking on them.

Sub-question 3

> Are the Library of Congress subject heading concepts highly ranked in the LibraryThing tag lists?

The aim of this sub-question was to discover the average number of times a tag was used for all of the three match categories. This showed how often LibraryThing users made use of concepts also used by the LC subject headings compared to the alternatives in the 'No Match' category.

This was calculated by adding a further column to the spreadsheets which was labelled 'Match'. The 'Match' column was populated by classifying the code categories A-E. Category 'A' was classified an 'Exact Match' (with the LC subject headings), while category 'C' was classified as a 'Partial Match'. Every other code category was classified as a 'No Match'.

The ‘Tag Count’ column was totalled, and then Excel formulas were used to calculate the amount of tag counts for ‘Exact Match’, ‘Partial Match’, and ‘No Match’. These figures were then calculated as percentages of the tag count total.

The next step required calculating the total number of matches (the number of occurrences of matches) and then the sub-totals for all of the match categories (under the heading of ‘Group Count’).

The final output of this test was calculated by dividing the tag counts for each category (Exact, Partial and No Match) by their corresponding ‘Group Count’.

It is worth noting the possibility that more LC subject heading concepts were used, but were lowly ranked in LibraryThing and were not included in the sample.

Results and Discussion

Note on terminology

In the following discussion defining the terms ‘use’ and ‘occurrence’ seems appropriate.

‘Use’ refers to the number of times a tag was used by LibraryThing users. For example in Figure 2 below the tag adventure was used 30 times.

‘Occurrence’ refers to the number of tags in the sample. In Figure 2 below there are three occurrences (*‘adventure’*, *‘america’*, and *‘american’*).

Book	**Author**	
American Gods	Neil Gaiman	
Tags	**Tag count**	**Code**
adventure	30	D--Plo
america	148	D--PC
american	49	D--SD
Total	227	

Figure 2

Tag codes by occurrence

The three tag classes that appeared most often were the ‘Literary Genre’ tags (D—LG) with 93 occurrences (21% of total occurrences), the ‘Plot Details’ tags

(D—PLO) with 89 occurrences (21%), and 'User specific' tags (D—US) with 61 occurrences (14%).

The three tag classes that appeared least often were 'Cross referenced from LCSH' tags (C—CR) with 1 occurrence (0.2% of all total occurrences), 'Spelling variation' tags (C—SV) with 2 occurrences (0.5%), and 'Format Information from OCLC' tags (B—FI) with 5 occurrences (1.1%). To see the full range of statistics for this criteria see Figure 3 below.

The most frequently occurring tag classes are those that do not match the LC subject headings. This could indicate that there are a number of facets for the books in the sample that are not covered by the LC subject headings.

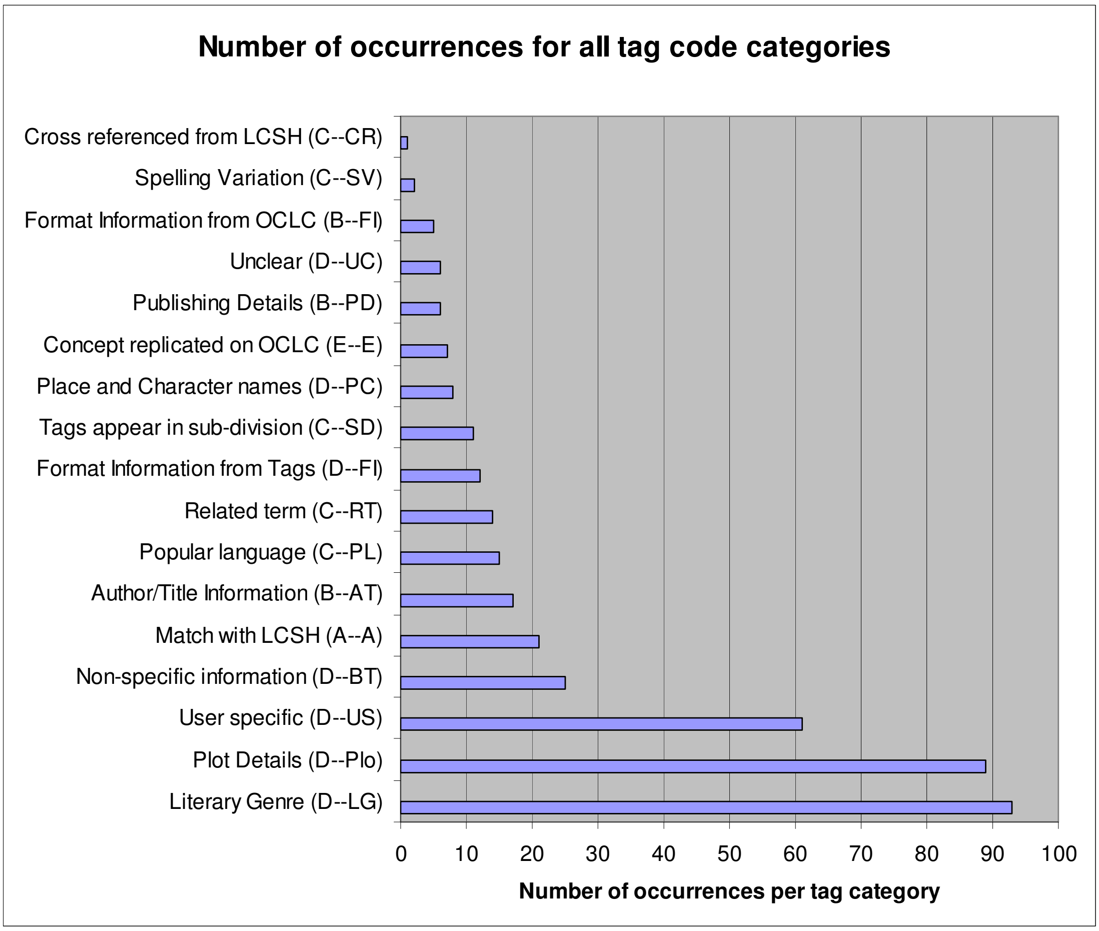

Figure 3

Tag codes by use

The three tag classes that were used most often were the 'Match with LCSH' tags (A—A) with 19,350 uses (29.89% of total uses), the 'Related term' tags (C—RT) with 13,659 uses (21.10%), and ‘Literary Genre’ tags (D—LG) with 7,237 uses (11.18%).

The three tag classes that were used least often were the ‘Cross referenced from LCSH’ tags (C—CR) with 47 uses (0.07% of total occurrences), ‘Publishing Details’ tags (B—PD) with 161 uses (0.25%), and ‘Popular language’ tags with 184 uses (0.28%). To see the full range of statistics for this criteria see Figure 4 below.

A significant number of the tags used are in full or partial agreement with the LC subject headings. However, while taggers use LCSH type tags extensively, the relatively high number of uses for ‘Literary Genre’ tags (11.18%) plus the high number of occurrences in this tag class demonstrate that LibraryThing users are adding many non-LC subject heading tags.

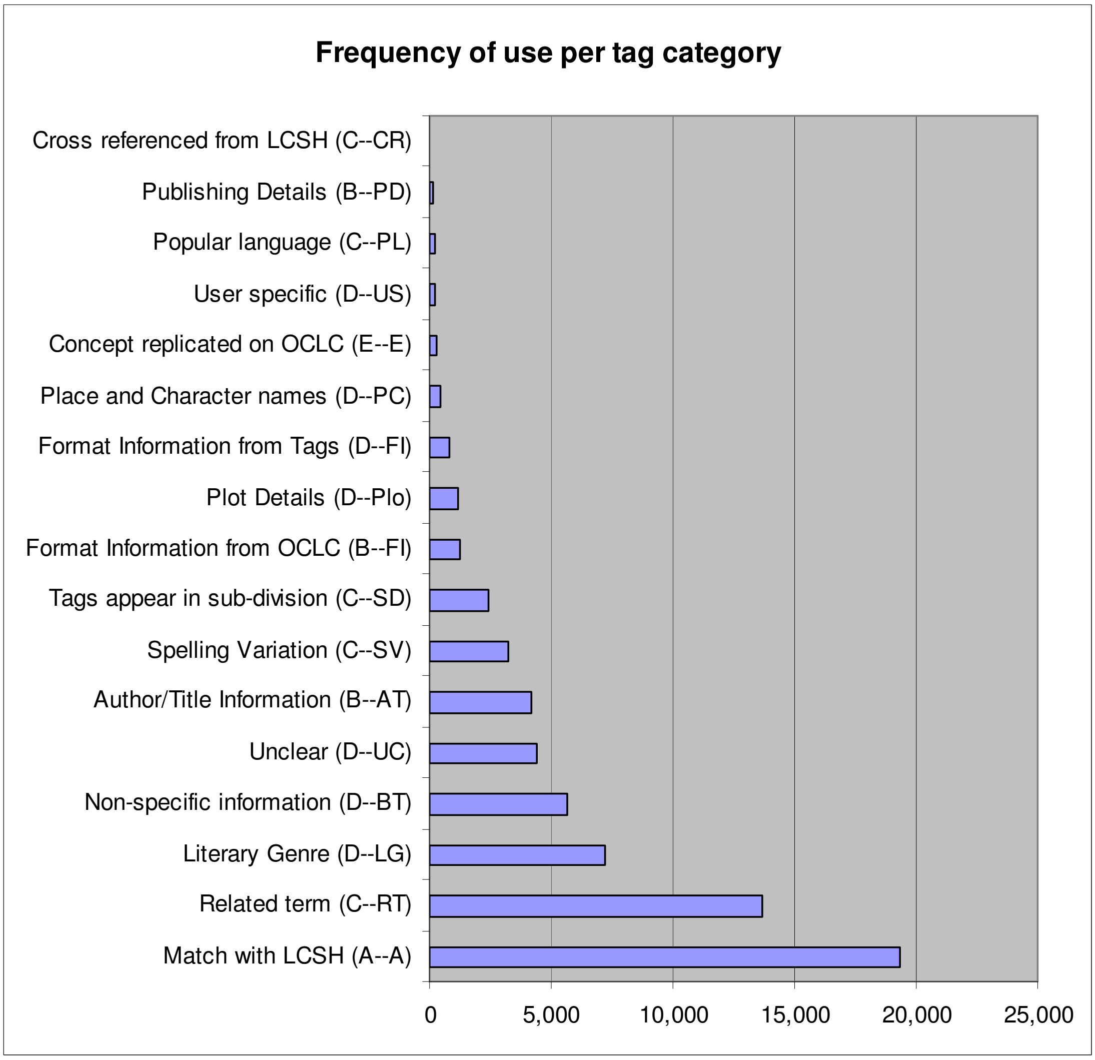

Figure 4

Sub-questions 1 & 2: Commonalities between LibraryThing tags and LC subject headings

"To what extent and how often do LibraryThing tags represent the same concepts as Library of Congress subject headings?"

"How often do LibraryThing tags represent concepts not included in the Library of Congress subject headings (and vice-versa)?"

The total count for the use of all tags was 64,734; 52% of these matched the LC subject headings using the LCSH match score. The data obtained through the LCSH match score showed that LibraryThing users' replication of concepts in the LC subject headings was in the range of 40-60% for each book. These percentages were similar for all the books in the sample, indicating that the replication of LC subject heading concepts was consistent for each book.

These percentages were reversed to ascertain how often the LibraryThing tags did not match the LC subject headings. Overall there was a 48% score for tags that did not match the LC subject headings. The data obtained through the LCSH match score also yielded results in the 40-60% range for each book.

In summary about half of the tags used in LibraryThing represented the same or similar concepts as the LC subject headings. However the LCSH match score represents at best an arbitrary measure of the use of LC subject headings in the sample, although it does indicate a significant overlap between the LC subject headings and the LibraryThing tags.

In addition a count by match occurrence shows that only a small proportion of tags actually matched the LC subject headings. See figures 8 and 9 in the appendix for the data from the LCSH match score.

Total tag matches with LC subject headings

Dividing the tags for the sample into the categories of Exact, Partial, and No Match and then counting the number of occurrences, showed that there were 27 exact matches (or 6% of the sample) for the LC subject headings, and 52 partial matches (12%). The remaining 358 tags (82%) did not match the LC subject headings at all. This is shown in Figure 5 below.

The total tag match count shows that Exact and Partial matches make up 18% of all occurrences in the sample, however they account for 52% of total uses in the LCSH match score. This difference between the LCSH match score and the total tag matches can be explained by the large numbers of LibraryThing users who favoured simple terms such as *'fiction'*, *'science fiction'*, and *'fantasy'* (11,098, 5,148, and 16,246 uses over the whole sample respectively). The LCSH match score was skewed by the high number of uses of these simple terms. For example, in 'American Gods' the term *'myth'* (a partial match) was used 65 times, while *'fiction'* (an exact match) was used 1,455 times.

The tendency of LibraryThing users to favour these simple terms seems to reflect Zipf's Principle of Least Effort as discussed by Munk and Mork. The simple categories are the first that comes to mind for many users (Munk & Mork, 2007, p. 29).

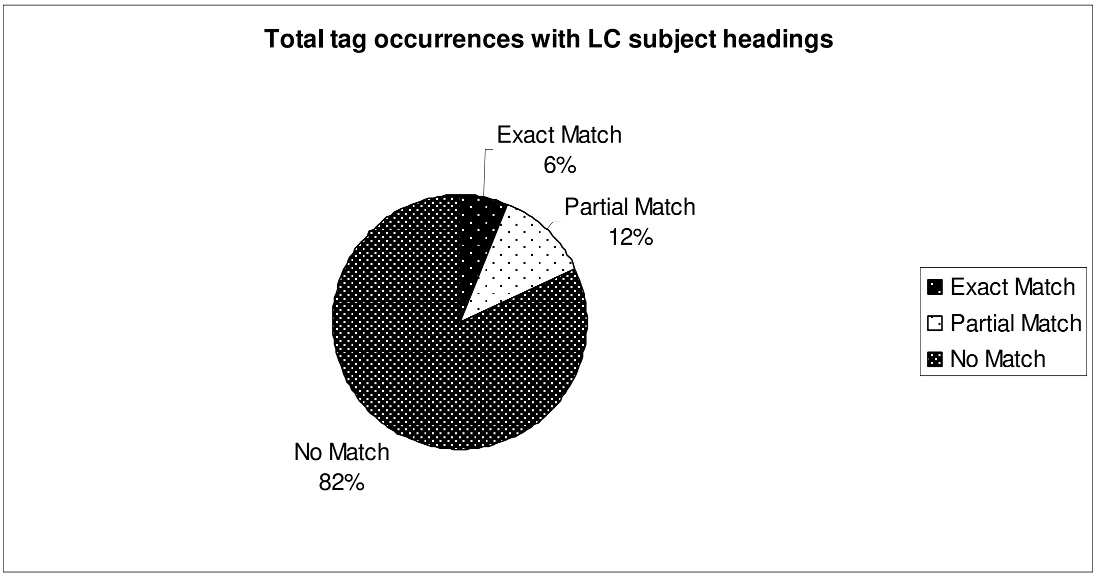

Figure 5

Sub-question 2: LC subject heading concepts not present in LibraryThing tags

The total number of all of the subject headings in the sample was 50, 37 of which (74%) had LibraryThing equivalents. The remaining 13 subject headings (26%) did not have any LibraryThing equivalents.

Five of the ten books had subject headings with no match in the LibraryThing tags. The number of subject headings (both matched and unmatched) for each title varied from a maximum of 14 for 'Harry Potter and the

Order of the Phoenix' to a minimum of 1 for the 'Left Hand of Darkness'. This variability in the number of subject headings is similar to what Mendes found in his recent study (Mendes et al., 2009, p. 38).

The LC subject headings with no LibraryThing equivalents generally cover valid facets of the books to which they are applied, but some seem very minor facets. For example the subject headings *'Widowers'* and *'Bodyguards'* in the book 'American Gods' only apply to the protagonist Shadow, and are not necessarily memorable aspects for most readers. However a check on 'American Gods' on the LibraryThing website found both tags were present but were not included in the sample.

The remainder of the LC subject headings are place and character names such as *'Wiggin, Peter (Fictitious character)'* from 'Ender's Game', and *'York (England)'* from 'Jonathon Strange and Mr Norrell'.

York is the setting for much of the plot for 'Jonathon Strange and Mr Norrell'; it seems a relatively minor facet.

Peter Wiggin is an important character from 'Ender's Game', and at first glance it seems surprising that he was omitted from the LibraryThing tags.

However, a check on 'Ender's Game' and 'Jonathon Strange and Mr Norrell' on the LibraryThing website found that there were tags for Peter Wiggin and York but they had not been used often enough to be picked up in the sample for this study.

Given the overall low use of the 'Place and Character names' (D—PC) it would appear that LibraryThing users consider them to be minor aspects of the books.

A small proportion of the LC subject headings represent concepts not used by the LibraryThing tags. The unmatched concepts generally cover minor facets of the books in the sample.

Nevertheless the sample taken in this study is very small and these results require further investigation to confirm these findings.

Sub-question 2a: LibraryThing concepts not present in LC subject headings

"What sorts of concepts are represented in the LibraryThing tags that are not included in neither the Library of Congress subject headings nor the OCLC records?"

There were 316 tag occurrences which did not match with either the LC subject headings or the OCLC record in total. The 'D-coded' tags occurrence averaged around 32 tags for each book. The range across the sample was consistent with a minimum of 22 occurrences for 'Harry Potter and the Order of the Phoenix' and a maximum of 36 for 'American Gods'. For further details see the discussion and Figure 6 below. Figure 6 represents the numbers and types of tags which did not match the LC subject headings.

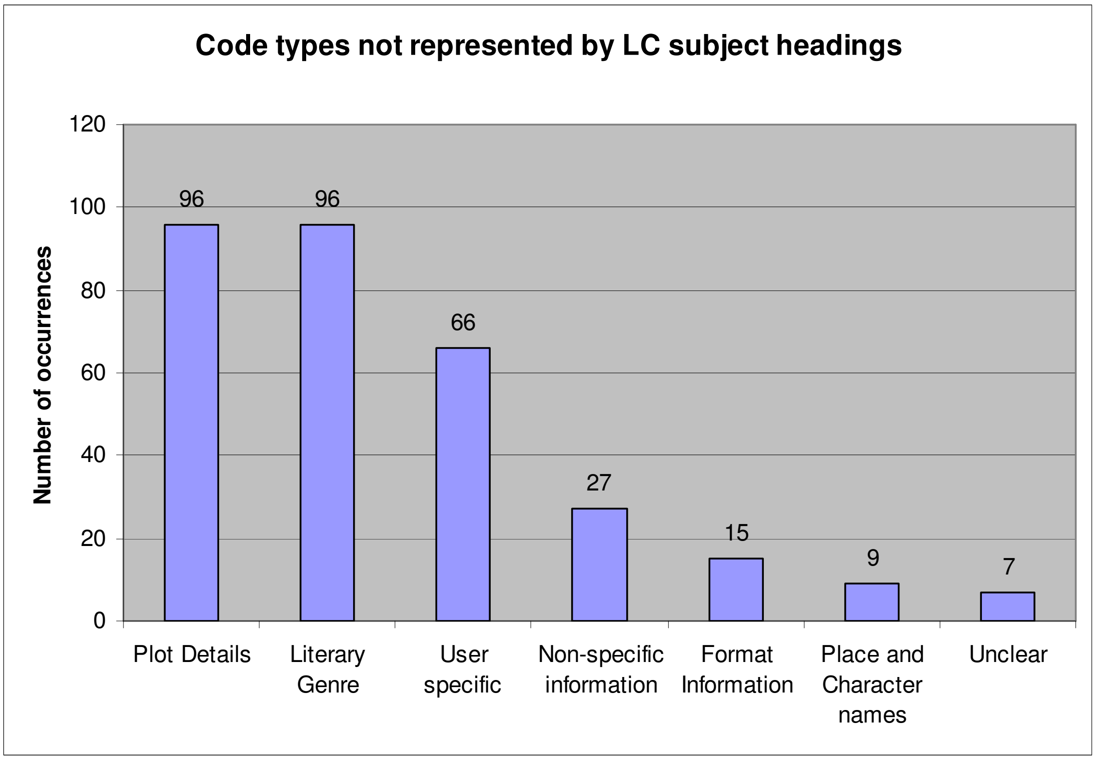

Figure 6

Sub-question 2a –Analysis of 'D' code tags

The objective of this analysis was to identify tag concepts that were not covered by the LC subject headings or the OCLC records. These were terms that were not used by the cataloguers but are important to LibraryThing users.

It is important to note that LibraryThing users were probably not trying to 'catalogue' these books. Most users were probably trying to jot down the aspects of the book that they found most memorable or were of interest to them, although some might later use their tags to retrieve their books.

It is probable that there are equivalents for some of these tags in the LC subject headings, (some examples are cited below) so the question arises of why they are not used.

Also the numbers of LC subject headings themselves vary inside the sample. For example 'Harry Potter and the Order of the Phoenix' has fourteen separate subject headings, while 'The Left Hand of Darkness' has only one. Considering that cataloguers have historically only used one or two subject headings per book this variation is surprising.

Plot details (D--PLo)

This was the largest D-code category with 96 occurrences (30% of the total), equal ranked with the 'Literary Genre' category. Compared to 'Literary Genre' and the other categories, its tags were more varied and with fewer synonyms.

The numbers of plot detail tags varied from book to book inside the category; ranging from a minimum of seven tags for 'Jonathon Strange' to a maximum of sixteen for 'Snow Crash'. This variation is expected with a small sample. One further explanation for Snow Crash's popularity might be that it has a cult following on LibraryThing and is therefore described in finer detail by its fans.

The relevance of the plot tags varied from tag to tag; for example the *'norse mythology'* tag described quite a small aspect of the plot of 'American Gods'. Norse mythology in particular does not play a large role in this novel; however this tag was used by 27 LibraryThing users. This tag links to a list of books in LibraryThing concerning Norse mythology in general (for example 'Poetic Edda' by Carolyne Larrington).

A tag of greater relevance is *'coming of age'* which is used in the books 'Ender's Game' and 'The Subtle Knife'. The protagonists of both novels are precocious children that are forced by circumstances to grow up quickly.

In LibraryThing this tag leads the user to other novels with similar themes such as the 'Catcher in the Rye' by J.D. Salinger and 'Great Expectations' by Charles Dickens. Tags with themes such as *'coming of age'* could be very useful to libraries that provide a reader's advisory service, or even in academic libraries (as Westcott's anecdote about the student using the tag *'The Color Purple'* to find books with themes of abuse demonstrates) (Westcott et al., 2009, p. 80).

Of more interest in this category is where users appear to agree over the plot of a book by using synonyms to describe it. For example in 'The Subtle Knife' the tags *'Alternate Universe', 'alternate worlds', 'parallel worlds'* are each used 31, 21, and 25 times. 'The Subtle Knife' and the other books in Phillip Pullman's 'His Dark Materials' series are set in parallel worlds to our own, each with slightly different histories. No obvious equivalent for this concept could be found in the LC subject headings.

These tags link to other books in LibraryThing where history has been re-imagined into a different pattern by the author, for example 'Fatherland' by Robert Harris (a book in which Nazi Germany never fell), or 'Black Powder War' by Naomi Novik who re-wrote the Napoleonic Wars to include dragons.

There are other examples where synonyms seem to indicate a degree of user agreement about the plot. In particular, the synonyms *'language'* and *'linguistics'* in 'Snow Crash' (22 and 25 uses each), and *'gender'* and *'gender roles'* in 'The Left Hand of Darkness' (135 and 10 uses each). The plot of 'Snow Crash' features a computer virus that attacks the victims' ability to recognise languages, while 'The Left Hand of Darkness' features an alien planet whose population has no gender most of the time.

These tags were judged to be important by LibraryThing users and their absence in the LC subject headings restricts users' abilities to find similar titles.

Literary Genre Information (D--LG)

This was the largest D-code category with 96 occurrences (30% of the total), equal ranked with the 'Plot details' category. However in comparison to 'Plot details' its tags were less varied and had more synonyms.

The numbers of literary genre tags varied from book to book inside the category; ranging from a minimum of three types for 'Harry Potter and the Order of the Phoenix' to a maximum of fifteen for 'The Subtle Knife'. 'Harry Potter and the Order of the Phoenix' had few literary genre tags because it was extremely well-described by both the LC subject headings and the OCLC record. 'The Subtle Knife' had the most tags but many of them were synonyms.

The relevance of some tags in this category was not immediately obvious to the researcher. For example some users had tagged the fantasy novels 'American Gods' and 'Jonathon Strange' as science fiction. Likewise the science fiction novel 'Stranger in a Strange Land' was tagged as fantasy. Some users fudged this gap by using the tag *'sff'* which stands for 'science fiction/fantasy'. This illustrates different understandings amongst LibraryThing users of the terms science fiction and fantasy. There is no single subject heading in LCSH which covers this concept.

Some literary genre tags gave very general classifications for books (for example *'science fiction', 'fantasy'*), while others are far more specific.

'Steampunk' (used to tag 'The Subtle Knife') is a genre of fantasy with either a Victorian-era setting or technology that is distinctively Victorian-level (for example steam engines and Sir Charles Babbage's Difference Engine). In LibraryThing this tag links to lists of other novels with similar themes, for example 'The Diamond Age' by Neal Stephenson.

'Dark Fantasy' provides a more detailed literary classification for 'American Gods'. The term covers fiction writers who use a mixture of fantasy and horror. In LibraryThing this tag links to lists of titles such as 'Queen of the Damned' by Anne Rice, and 'The Drawing of the Three' by Stephen King.

Other tags in this category are used to provide a very general date context for the books, for example *'20th century'*.

Some of the tag concepts (such as *'Steampunk'* and *'Dark Fantasy'*) are genuine literary sub-genres that could be used to collocate similar titles in a library catalogue. Allowing tags such as these into the library catalogue would make it easier for users to find the books that they want.

User specific (D--US)

This category was the third most common of the D-tags with 66 occurrences (21% of the total), however in comparison to the 'plot' and 'literary genre' tags it displayed little variety. The tags used were mostly synonyms relating to individual users' concerns and interests.

These interests included ownership status (*'own'*), whether they had read, or intended to read a book, (*'read', 'tbr'*) whether they liked it or not, (*'favorite'*) and noting if they had a signed or first edition copy of a book (*'signed', 'First Edition'*).

User specific tags embody Guy's assessment of one of the problems with folksonomies: "they are trying to serve two masters at once; the personal collection, and the collective collection" (Guy & Tonkin, 2006, p. 10). While these tags are too personal to be of use in a library context, they may be quite helpful to LibraryThing users in a social sense as they allow users to compare their collections. It is possible that these tags could serve a similar function if their use was allowed in library OPACs. For example library reading groups could share tags such as *'read'*, *'unread'*, amongst themselves.

Other non-specific info (D--BT)

The 'Other non-specific info' category made up 9% of the 'D'-coded tags with 27 occurrences. This smaller category contains a number of interesting tags that generate interesting search results inside LibraryThing.

Tags such as *'hugo award'*, and *'nebula'* refer to books that have won the Hugo and Nebula prizes for science fiction (such as 'Ender's Game' and 'Jonathon Strange'). Searching via these tags in LibraryThing generates a list of other books that have also won these prizes.

A search under the tag *'1001 books'* gives the user a list of books tagged by users as included in Peter Boxall's book '1001 Books You Must Read Before You Die'.

The tag *'green dragon'* refers to one of the many discussion groups on LibraryThing. The Green Dragon group are interested in Tolkien, fantasy, science fiction and mythology.

A search under the tag *'Inklings'* calls up a list of books written by authors associated with the literary group known as the Inklings. This group most famously included C.S. Lewis and J.R.R. Tolkien. This tag is available as a subject heading in the Library of Congress database.

While some of these tags may not be suitable for inclusion in the LC subject headings, they would be interesting and useful to some readers. These tags could easily co-exist with the LC subject headings as Wetterstrom has suggested (Wetterstrom, 2007, p. 33).

Format Information (D--FI)

The 'Format Information' category makes up 5% of the 'D'-coded tags with 15 occurrences. There was little variety inside this category with five distinct tag types recorded.

These tags mostly served as personal notes made by users indicating the format of a particular item in their library or the format in which they 'read' a particular item. Some are useful only to their creators, while others might have a wider audience. For example a tag such as *'hardback'* would probably only be useful inside the collections of a few users and would not produce useful search results for a wider audience. However, searching for books in rarer formats using tags such as *'audiobook'* could be more helpful. In hindsight these tags could have been classified as 'user specific' (D—US).

These tags do not fit in with the LC subject headings and more properly belong to the catalogue record. They possibly do serve a function inside LibraryThing, allowing users to compare their collections. For libraries with access to OCLC (or an equivalent) 'format information' tags are not that useful.

Place and character names (D--PC)

The 'Place and character name' category occurred 9 times (3% of the 'D'-coded tags); they are infrequently used at best.

The place names used are mostly real-world places, for example *'Oxford'*. In LibraryThing this tag links to a list of books set in England and Oxford, for example 'Brideshead Revisited' by Evelyn Waugh, and Pride and Prejudice by Jane Austen. It is interesting to note that many of the books listed under these tags are from the fantasy genre. This might be indicative of the membership of LibraryThing.

The remaining tags are the names of distinctive characters and fictional places such as *'hiro protagonist'* from the novel 'Snow Crash', and *'ekumen'* which is the name of Ursula K. Le Guin's fictional interstellar association. The first of these tags is useless for collocating other books in LibraryThing as the character only appears in one book, but a search under *'ekumen'* brings up all of Le Guin's books set in the 'Ekumen'.

It is possible to have a subject heading for imaginary places in LCSH; for instance the Silmarillion was given the subject heading Middle Earth (Imaginary place), but this was not done for Ursula Le Guin's books. This is an example of inconsistent cataloguing practice inside LCSH.

Unclear (D--UC)

The 'Unclear' category was the smallest (only 7 occurrences), making up 2% of the 'D'-coded tags. This low number can be partly attributed to the researcher's knowledge of the science fiction and fantasy genres. They were mostly date tags whose meaning can only be guessed at. Presumably they referred to when a LibraryThing user read a particular book as none of the dates used matched the dates of publication for any of the books.

While searching under these date tags seems useless, they could be of interest to users who want to know what was popular during any given year; the LibraryThing equivalent of a 'best sellers' list. They could also be of use for research into the reading habits of LibraryThing users.

The remainder (*'Collection', 'stories'*) are tags that do not fit easily into any of the categories created by the researcher. In hindsight they should have been put into the 'Other non-specific info' category.

Summary for sub-question 2a; concepts not present in LC subject headings

A large proportion of the concepts not included in the LC subject headings are potentially very useful for collocating other books of interest to LibraryThing users. They would be quite useful in libraries that run readers' advisory services as they would allow the librarians and their patrons to "search using a collection of terms that reflect taste and desired experiences" (Weaver, 2007, p. 586).

The large numbers of 'User specific' tags demonstrate a desire amongst LibraryThing users to track their own reading habits, and this should be taken into account when designing future OPAC systems.

The smaller but significant number of 'non-specific' tags show that LibraryThing users are coming up with their own unique categories to cover their specific interests. While it is unlikely that these tags could be fitted into LCSH, their presence shows the usefulness of allowing tagging in the library catalogue. This echoes Morville's assertion that controlled vocabularies such as LCSH and folksonomies such as LibraryThing should be allowed to co-exist and complement each other (Morville, 2005, p. 139).

Sub-question 3: The ranking of LC subject headings inside LibraryThing

"Are the Library of Congress subject heading concepts highly ranked in the LibraryThing tag lists?"

Tags which exactly matched the LC subject headings were used on average 717 times (62%) over the whole sample. Partially matched tags were used on average 370 times (or 32%), while tags which did not match the LC subject headings were used an average 73 times (6%). Figure 7 gives an overview of the results.

The ranges of uses for the tags inside each match category were varied. For example in the 'Exact Match' category the maximum was 2788 uses for *'fiction'* in 'Harry Potter and the Order of the Phoenix', while the minimum was 25 uses for *'Fantasy fiction'* in 'The Silmarillion'.

In the 'Partial Match' category the maximum was 4273 uses for *'Fantasy'* in 'Harry Potter and the Order of the Phoenix', while the minimum was 14 uses for *'sff'* in 'Snow Crash'.

'No Match' had a maximum count of 725 uses for *'magic'* in 'Jonathon Strange and Mr Norrell', and a minimum of 6 uses for *'Movie'* in 'The Martian Chronicles'. Overall tags in the categories of 'Exact' and 'Partial' matches had higher uses than 'No Match' tags.

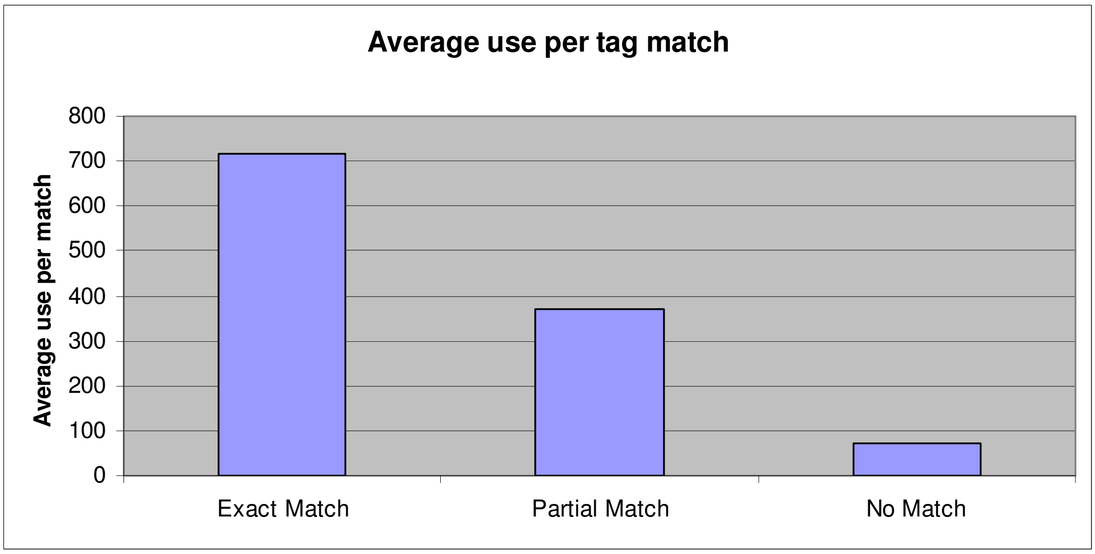

Figure 7

As with the LCSH match score the average use of Exact and Partial matches were skewed by LibraryThing users' preference for basic genre categories such as *'fiction'*, *'fantasy'*, and *'science fiction'* (11,098, 5,148, and 16,246 uses over the whole sample respectively).

Looking at the un-averaged matches gives a somewhat different picture. Tags which exactly matched the LC subject headings were used 19350 times (30%) over the whole sample. Partially matched tags were used on average 19259 times (or 30%), while tags which did not match the LC subject headings were used an average 26125 times (40%). The 'Exact Match' category no longer dominates, and the 'No Match' category is larger.

The un-averaged tag matches also give some notion of the relative popularity of these books inside LibraryThing. 'Harry Potter and the Order of the Phoenix' had an overall use of 17,429; by far the most popular book in the sample. 'The Left Hand of Darkness' had the lowest overall use at 2,941.

LC subject heading concepts are highly ranked even when taking into account LibraryThing users' tendency to make heavy use of basic categories. It appears that cataloguers using LCSH are good at identifying the basic categories to which most books belong, but are less effective with more specific detail.

Summary, conclusion, and suggestions for further research

Summary

The most frequently occurring tag classes were those that did not match the LC subject headings ('Literary Genre', 'Plot details' and 'User specific' tags) which made up 55% of all occurrences.

The two most frequently used tag classes were those that matched or closely matched the LC subject headings ('Match with LCSH' and 'Related term' which made up 51% of all tag uses). The third most used tag class was 'Literary Genre' at 11% which did not match any of the LC subject headings.

The extent to which LibraryThing tags matched the LC subject headings was calculated at 52%; however this percentage was exaggerated by LibraryThing users' preference for using the most obvious terms. Exact and partial matches only made up 18% of all tag occurrences in the sample. Nevertheless there was an overlap between the LibraryThing tags and the LC subject headings. Less than half (48%) of the LibraryThing tags used represented concepts not included in the LC subject headings.

A small proportion of the LC subject headings (26%) did not match any of the concepts used in the LibraryThing tags. The unmatched concepts generally covered minor facets of the books in the sample.

The majority of the concepts (60%) represented by the LibraryThing tags and not covered by the LC subject headings dealt with plot details, or provided additional classification details for the books they were applied to. A smaller number (21%) of tags applied personal tags to the books.

LibraryThing tags that were exact and partial matches for the LC subject headings were ranked highly by use (on average 62% and 32%). However the average was skewed by LibraryThing users' preference for basic genre categories.

Conclusion

The heavy use of tags that match with LCSH shows that cataloguers are good at identifying the basic categories to which most books belong, but are less effective with more specific details. These finer details constitute information access points that are neglected in the LC subject headings. This implies that cataloguers should either widen the number and scope of their headings or perhaps allow user tagging. Making use of services such as LibraryThing for Libraries would be a reasonable alternative.

Allowing users to create tags in library OPACs might be a good way to provide these alternate access points. Although LibraryThing users' tags often echoed the LC subject headings many also provided additional details about the books to which they were applied. The frequent occurrence of user specific tags implies that some users like to track their own reading habits and communicate with other LibraryThing users. These preferences should be taken into account when designing future OPACs.

In terms of information retrieval, tags which matched the LC subject headings were good for locating books by broader themes and genres but non-matching tags identified a wider range of books more likely to suit a diverse range of reader interests. For libraries offering readers advisory services these tags would increase the chances of library staff and patrons finding books that fit their specific needs and tastes.

Suggestions for future research

It could be useful to compare the results of this study with a larger sample from LibraryThing looking at other genres besides science fiction and fantasy. A larger sample could be used to confirm whether LC subject headings are as heavily used in other genres, and also investigate tagging patterns across LibraryThing.

Another interesting line of research would be to compare the tags from LibraryThing to tags created in library catalogues to see if there are many personal tags (or even if they are allowed in the catalogue; LibraryThing for Libraries filters them out), and if library taggers also make heavy use of basic categories for their tags.

Bibliography

Bruce, R. (2008). Descriptor and folksonomy concurrence in education related scholarly research [Electronic Version]. *Webology*, 5(3), 1-5. Retrieved 21st April, 2009 accessed from http://www.webology.ir/2008/v5n3/a59.html.

Guy, M., & Tonkin, E. (2006). Folksonomies: tidying up tags [Electronic Version]. *D-Lib Magazine*, 12(1). Retrieved 25th August, 2008 accessed from http://www.dlib.org/dlib/january06/guy/01guy.html.

Jennings, G. (2008). War Art Online - Presentation for ARANZ 2008. Archives New Zealand.

Kipp, M. E. I., & Campbell, D. G. (2006). *Patterns and inconsistencies in collaborative*. Paper presented at the ASIS & T Annual Meeting. Retrieved 11th Dec, 2008, from http://dlist.sir.arizona.edu/1704/01/KippCampbellASIST.pdf.

Krippendorff, K. (1980). *Content Analysis: an Introduction to its Methodology*. London: Sage Publications.

Kroski, E. (2007). Folksonomies and user-based tagging. In N. Courtney (Ed.), *Library 2.0 and beyond: innovative technologies and tomorrow's user* (pp. 91-103). Westport, Connecticut: Libraries Unlimited.

Library of Congress subject headings. (2007). (30th ed. Vol. 1-5). Washington, D.C. USA: Cataloguing Distribution Service, Library of Congress.

LibraryThing| Catalogue your books online. (2008). Retrieved 19th Oct, 2008, from http://www.librarything.com/about

Lin, X., Beaudoin, J. E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification [Electronic Version]. *Proceedings of the 17th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research* 17, 1-19. Retrieved 21st April, 2009 accessed from http://dlist.sir.arizona.edu/1790/01/lin.pdf.

Mendes, L. H., Quinonez-Skinner, J., & Skaggs, D. (2009). Subjecting the catalog to tagging [Electronic Version]. *Library Hi Tech*, 27(1), 30-41. Retrieved 21st April, 2009 accessed from Emerald Insight.

Mika, P. (2007). *Social Networks and the Semantic Web*. New York: Springer.

Morville, P. (2005). *Ambient Findability*. Sebastopol, California: O'Reilly.

Munk, T. B., & Mork, K. (2007). Folksonomy, the power law and the significance of the least effort. *Knowledge Organization, 34*(1), 17-33.

Oates, G. (2008). *The Commons on Flickr*. Paper presented at the Museums and the Web 2008. Retrieved 11th Dec, 2008, from http://www.archimuse.com/mw2008/papers/oates/oates.html.

Pickard, A. J. (2007). *Research Methods in Information*. London: Facet Publishing.

Rethlefsen, M. L. (2007). Tags help make libraries del.icio.us [Electronic Version]. *Library Journal*, 132(15), 26-28. Retrieved 14th August, 2008 accessed from ProQuest 5000.

Rethlefson, M. L. (2007a). Chief Thingamabrarian [Electronic Version]. *Library Journal*, 132(1). Retrieved 14th August, 2008 accessed from Proquest 5000.

Rethlefson, M. L. (2007b). Tags help make libraries del.icio.us [Electronic Version]. *Library Journal*, 132(15), 26-28. Retrieved 14th August, 2008 accessed from ProQuest 5000.

Sinclair, J., & Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? [Electronic Version]. *Journal of Information Science*, 34(1), 15-29. Retrieved 14th August, 2008 accessed from SAGE Journals Online.

Sogunro, O. A. (2002). Selecting a quantitative or qualitative methodology: an experience. *Educational Research Quarterly, 26*(1), 3-10.

Steele, T. (2009). The new cooperative cataloguing [Electronic Version]. *Library Hi Tech*, 27(1), 68-77. Retrieved 21st April, 2009 accessed from Emerald Insight.

Weaver, M. (2007). Contextual metadata: faceted schemas in virtual library communities [Electronic Version]. *Library Hi Tech*, 25(4), 579-594. Retrieved 25th July, 2008 accessed from Emerald Insight.

Westcott, J., Chappell, A., & Lebel, C. (2009). LibraryThing for libraries at Claremont [Electronic Version]. *Library Hi Tech*, 27(1), 78-81. Retrieved 21st April, 2009 accessed from Emerald Insight.

Wetterstrom, H. M. (2007). *The complementarity of tags and LCSH: a tagging experiment and investigation into added value in a New Zealand library context.* Victoria University, Wellington.

Zipf, G. K. (1965). *Human Behaviour and the Principle of Least Effort*. New York: Hafner Publishing.

Appendix

Book Titles	Tag count Totals	Total Weighted Scores
American Gods	7496	357750
Ender's Game	5969	335525
Harry Potter and the Order of the Phoenix	17429	926825
Jonathon Strange and Mr Norrell	7538	375825
The Silmarillion	7028	377025
Snow Crash	3924	211175
Stranger in a Strange Land	3369	208975
The Left Hand of Darkness	2941	150075
The Martian Chronicles	3143	178600
The Subtle Knife	5897	257650
TOTALS	64734	3379425

Figure 8

Book Titles	Sub-question 1	Sub-question 2
	Match with LC Subject Headings	No match with LC Subject Headings
American Gods	47.23%	52.77%
Ender's Game	56.21%	43.79%
Harry Potter and the Order of the Phoenix	53.17%	46.83%
Jonathon Strange and Mr Norrell	49.86%	50.14%
The Silmarillion	53.65%	46.35%
Snow Crash	53.82%	46.18%
Stranger in a Strange Land	62.03%	37.97%
The Left Hand of Darkness	51.03%	48.97%
The Martian Chronicles	56.82%	43.18%
The Subtle Knife	43.69%	56.31%
Overall Average	**52.20%**	**47.80%**

Figure 9

Word Count: 10,964