

A matter of context: An investigation into the representation of bias in social tags and Library of Congress Subject Headings

Chloe Gisela Jane Dash

A dissertation submitted to Aberystwyth University in partial fulfillment of the requirements for the degree of MSc under Alternative Regulations

Department of Information Studies

Aberystwyth University

2015

Abstract

This study explores the representation of bias in social tags and Library of Congress Subject Headings, with a particular focus on the motivations of the layperson (the tagger) and the expert (the cataloguer).

A mixed methodological approach was adopted. A framework for measuring bias was defined and constructed and this was applied via a simple coding scheme to a total of 500 social tags from LibraryThing and 175 Library of Congress Subject Headings from the Library of Congress online catalogue. These were harvested from a sample of 50 popular feminist fiction titles.

The analysis demonstrated that, although there were a higher proportion of unbiased social tags than unbiased LCSH, issues of bias were found in both systems. The two systems displayed very distinct issues of bias, given the differing motivations of the tagger (personal) and the cataloguer (to allow subject access).

The research demonstrated the idea that the concepts of bias and interpretation are inseparable; and (regardless of system and language), one cannot interpret anything without applying personal, cultural and leaned biases based on a particular worldview.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed (candidate)

Date

STATEMENT 1

This work is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed (candidate)

Date

STATEMENT 2

I hereby give consent for my work, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date

Table of contents

Abstract	ii
List of tables	vii
List of figures	viii
List of abbreviations	ix
Acknowledgments	x
1. Introduction	1
1.1 Background to the research.....	1
1.2 Social tagging.....	2
1.3 Library of Congress Subject Headings	2
1.4 Research Purpose.....	3
1.5 Research aims and objectives.....	4
1.6 Research structure.....	5
2. Literature review	7
2.1 Literature search.....	7
2.2 Bias and LCSH.....	8
2.3 LCSH and social tags.....	12
2.4 Conclusion.....	18
3. Methodology	19
3.1 Introduction.....	19
3.2 Approaches to the research.....	19
3.2.1 Mixed method approach.....	19
3.3 Research methods.....	20
3.3.1 Literature review.....	21

3.3.2 Pilot study.....	21
3.3.3 Data selection.....	22
3.3.3.1 Sampling.....	23
3.3.4 Data organization and presentation.....	24
3.3.5 Data analysis.....	25
3.3.5.1 Coding.....	25
3.3.5.1.1 Defining and measuring bias.....	26
3.3.5.1.2 Coding manual.....	27
3.4 Limitations of this approach.....	31
3.5 Ethics.....	31
4. Results.....	32
4.1 Introduction.....	32
4.2 The representation of bias through the application of social tags and LCSH.....	32
4.2.1 Social tags: Categories of bias.....	32
4.2.2 LCSH: Categories of bias.....	33
4.3 Analysis of the ways in which social tags and LCSH present bias.....	34
4.3.1 Tags and headings based on stereotypes.....	34
4.3.2 Tags and headings that treat the object as an exception.....	35
4.3.3. Tags and headings that are comprised of offensive terminology.....	36
4.4 Geographical and racial references: a matter of context.....	37
4.4.1 The removal of geographical/racial social tags and LCSH.....	38
4.5 The representation of bias through the application of non-geographical/racial social tags and LCSH.....	40
4.5.1 Non-geographical/racial social tags: Categories of bias.....	40
4.5.2 Non-geographical/racial LCSH: Categories of bias.....	41

4.6 Analysis of the ways in which non-geographical/racial social tags and LCSH present bias.....	41
4.6.1 Non-geographical/racial tags and headings based on stereotypes	41
4.6.2 Non-geographical/racial tags and headings that treat the object as an exception.....	42
4.6.3 Non-geographical/racial tags and headings that are comprised of offensive terminology.....	43
5. Discussion.....	44
5.1 Introduction.....	44
5.2 Statement of results.....	44
5.2.1 Issues of bias in social tags.....	44
5.2.2 Issues of bias in LCSH.....	47
5.2.3 Comparisons between the two systems.....	48
5.3 Recommendations for further research	49
6. Conclusion.....	51
6.1 Introduction.....	51
6.2 The layperson and the expert: approaches to bias	51
6.2.1 The layperson.....	51
6.2.2 The expert.....	52
6.3 There is no neutrality	54
7. Bibliography.....	55
Appendices.....	59
Appendix A: Titles used in pilot study.....	59
Appendix B: Titles used in main study.....	60

List of tables

- 1. Outline of coding system used to organize headings and tags.....28
- 2. LibraryThing tags and LCSH based on stereotypes.....35
- 3. LibraryThing tags and LCSH that treat the object as an exception.....36
- 4. LibraryThing tags and LCSH comprised of offensive terminology.....37
- 5.1 Geographical and racial LibraryThing tags removed from the study.....38
- 5.2 Geographical and racial LCSH removed from the study.....39
- 6. Non-geographical/racial LibraryThing tags and LCSH based on stereotypes.....42
- 7. Non-geographical/racial LibraryThing tags and LCSH that treat the object as an exception.....43
- 8. Non-geographical/racial LibraryThing tags and LCSH that are comprised of offensive terminology.....43

List of figures

1. Pie chart showing the distribution of categories of bias in LibraryThing tags.....	33
2. Pie chart showing the distribution of categories of bias in LCSH.....	33
3. Pie chart showing the distribution of categories of bias in non-geographical/racial LibraryThing tags.....	40
4. Pie chart showing the distribution of forms of bias in non-geographical/racial LCSH.....	41

List of abbreviations

EX	Object as exception
LC	Library of Congress
LCSH	Library of Congress Subject Headings
LGBT	Lesbian, Gay, Bisexual and Trans
LISA	Library and Information Science Abstracts
OF	Offensive terminology
ST	Based on stereotypes
UN	Unbiased

Acknowledgements

With thanks to Dr Pauline Rafferty for her support, guidance and advice as a supervisor throughout the research process. Additionally, the rest of the staff in the Department of Information Studies have been valuable in their teaching and the commitment they have to distance learning students.

I am also grateful to both LibraryThing and the Library of Congress for allowing me permissions to conduct this research.

Furthermore, I would like to thank my family, friends and partner for their continuous support, patience and encouragement.

1. Introduction

1.1 Background to the research

Developments in Web 2.0 technologies over the past decade have seen an influx of systems, tools and communities that empower the everyday user with the ability to apply their own interpretation to a variety of media in the public domain. Social tagging (also known as user tagging or collaborative tagging) is just one aspect of the Web 2.0 revolution. It allows users to “tag” specific content by applying their own keywords to describe websites, images or other media.

Given that social tagging is a method of organizing information, it is not surprising that the concept has been a subject of growing interest in Library and Information Science and its related fields. Several studies have attempted to compare social tagging with controlled vocabularies such as Library of Congress Subject Headings (LCSH) to assess the viability of a tagging system in library catalogues (Kipp, 2007; Rolla, 2009; Yi and Chan, 2009); however, little research has been carried out on the semantic impact that these tags have on users attempting to locate information in a library context.

Allowing the untrained user free rein to index and categorize library resources using an uncontrolled vocabulary contradicts with many of the traditional principles of librarianship. Yet, the most widely used controlled vocabulary in libraries, LCSH, has been long criticized as a flawed system filled with inconsistencies, prejudices and bias. This presents questions as to how each system deals with its subjects in terms of objectivity and applying meaning. This

dissertation seeks to compare LCSH and social tags in terms of bias; with the purpose investigating whether the social tagger – with their uncontrolled vocabulary, lack of guidance and personal motivation – make them more or less prone to bias than the cataloguer; whose role is to facilitate subject access under the constraints of a controlled vocabulary.

1.2 Social tagging

An ever-growing number of social tagging applications exist online, and tagging takes place in various contexts and is applied to various mediums across these applications. Despite the varying contexts, studies have led to similar conclusions with regard to the advantages. Central to this is their use of natural language, which – unlike artificial languages such as LCSH – is potentially more user-friendly and familiar to searchers (Mathes, 2004); this, along with the lack of any need for formal training (Peterson, 2006), allows for descriptions to take a more fluid, current, and natural form (Kroski, 2007).

The key disadvantages of social tagging focus on the lack of orthographic control (Kroski, 2007), its ambiguity of meaning (Yi, 2010), and a lack of polysemy and synonymy (Mattes, 2004).

1.3 Library of Congress Subject Headings

Developed and maintained by the Library of Congress for over a century, LCSH is a list of authorized subject headings used widely in library catalogues around the

world to identify the subject of a work and to collate similar works on a particular subject.

Although other controlled vocabularies exist, LCSH is arguably the most comprehensive general subject heading list available. Its bibliographic records are commercially available other to libraries, and its strong organizational support ensures regular updating, making it arguably the most widely used subject heading list worldwide (Jones et al., 2012).

Studies have shown that LCSH has strong advantages in terms of its ridged structure (Mann, 2003), its rich vocabulary (O'Neill and Chan, 2003), and its homograph and synonym control (Yi, 2010). However, in an environment where users are accustomed to keyword searches via search engines, many commentators question the value of the complicated pre-coordinated subject strings that make up LCSH (Rolla, 2009).

Furthermore, despite its advantages, the system has faced strong and ongoing criticism with regard to its biased and inaccurate and inconsistent subject headings (Strottman, 2007), its lack of representation of marginalized groups (Olson, 2000), and the LC's long delays to update outdated terminology (Fischer, 2005).

1.4 Research Purpose

The purpose of the research is to explore the relationship between social tagging and LCSH in terms of their representation of bias. Any system that involves the

interpretation of information is prone to bias (Olson, 2000), so there is no question that both systems will inevitably contain biased terminology; what this study focuses on, therefore, is *how* that terminology is presented.

It does so by performing an analysis of the tags applied to fifty titles in the social-bookmarking site LibraryThing, and comparing these with the LCSH for the titles on the LC's online catalogue. The selected titles pertain to feminist fiction, given that this area is seen as particularly problematic in LCSH (Berman, 1993; Fischer, 2005; Rogers, 1993) and would therefore demonstrate clearly any signs of bias in both LCSH and the social tags.

Given that many libraries are implementing social tagging capacities to their online catalogues, the overarching purpose of this research is to provide insight into how the two systems present biases, prejudices and personal views.

1.5 Research aims and objectives

Further to the research purpose, the aim of the research is an investigation into the relationships between the "expert" (i.e. the cataloguer) and the "layperson" (i.e. the tagger) in their presentation of bias. The following research objectives have subsequently been defined:

1. To devise and construct an explicit definition of bias that can be applied to a workable, verifiable and rigorous framework which can easily be applied to both social tags and LCSH.

2. To systematically apply this framework to the tags and LCSH harvested from LibraryThing and the LC online catalogue.
3. To assess the relationships and differences between how the two systems present bias.
4. To explore and assess the validity of this approach, and to recommend further areas for research following this study.

Objective 1 is addressed in the study's Literature Review and Methodology chapter; the second objective is met within the Results chapter; the Discussion chapter addresses objective 3; and the dissertation's Discussion and Conclusion address the final objective.

1.6 Research structure

The dissertation begins with a review of the literature. This introduces and analyzes the viewpoints and findings of researchers on the topics central to this work.

The following chapter presents and justifies the methodology chosen for the study in terms of the approach to sampling, the construction of a framework to measure instances of bias, and the tag and LCSH analysis itself.

The results of the analysis are presented in the fourth chapter; and the assessments are explored, analyzed and discussed in the fifth chapter - recommendations for further research are also suggested. Finally, the conclusion

brings together the study's results into a wider discussion, summarizing the key trends of the findings. It also assesses the validity of the chosen approach.

2. Literature review

Given the breadth of the proposed research topic, it should be noted that the following literature review is by no means extensive. In order to remain relevant to the research question, the following points will be examined:

- Bias and Library of Congress Subject Headings (LCSH)
- Bias and social tagging/folksonomies
- Comparisons between LCSH and social tagging/folksonomies

2.1 Literature search

The key words highlighted above formed the basis of three separate search strategies conducted on the Library and Information Science Abstracts (LISA) database:

- **Bias in LCSH:** ("Subject heading schemes" OR "library of congress subject headings") AND ("Bias") **12 results.**
- **Bias in social tagging:** ((Social tag*) OR (folksonom*)) AND ("Bias") **3 results.**
- **LCSH vs social tagging:** ("Library of Congress Subject Headings" OR "Subject heading schemes") AND ((Social tag*) OR (folksonom*)) **12 results.**

In terms of limitations, these initial searches included very specific search terms selected intentionally in order not to stray too far from the research question.

However, this means that papers discussing topics similar (and possibly relevant) topics may not have been included in the results.

Only LISA was searched, and although the contents of the database are extensive and relevant; as a source, it is by no means exhaustive – and therefore it must be accepted that further literature may be available, particularly in terms of books or book chapters, which are not indexed as comprehensively in LISA.

Furthermore, the concepts of social tagging and folksonomies are fairly recent, therefore all of the relevant literature is recent; and although many authors make good attempts, the subject is yet to be explored fully due to its currency.

However, although the number of results for each search was relatively low, further relevant literature was obtained through the process of snowballing, thereby addressing some of the limitations of the search.

2.2 Bias and LCSH

It is generally accepted in the literature that bias is inevitable in all forms of indexing and classification (Olson, 2000; Knowlton, 2005); and that due to its 19th century American origin, LCSH has a significant bias towards the language, culture and perceptions of its founders (Olson, 2000). There is a comparatively large amount of literature on the subject, which, as Fischer (2005) points out, suggests that librarians are increasingly concerned with the language of LCSH (p. 77)

A sort of ‘chicken or the egg’ idea comes up repeatedly in the literature on bias and LCSH. Some claim that, given that the scheme is based on literary warrant, it is biased because the literature on which the headings are created is biased

(given that LC is the US's legal deposit library) (Olson, 2000), while others argue that the slow-to-adapt nature of the scheme means that bias-filled headings are forced upon bibliographic records covering an array of culturally sensitive subjects, thereby passing on these north American 19th century prejudices to readers around the world (Rogers, 1993).

Olson, although accepting that the scheme is far from perfect, takes the view that LCSH is purely a descriptive reflection of the world it is attempting to document:

To view LCSH as a simple hegemonic tool of dominance would be simplistic. It seeks to be a universally applicable vocabulary treating topics with the neutrality of equality. Yet it is under constant revision as not only the universe it represents changes, but also as the concepts of neutrality and equality change. (Olson, 2000, p. 66)

However, it is this constant revision – or lack thereof – that angers some commentators. Being such a universally accepted scheme, adopted by libraries worldwide, the LC's response to the most glaringly obvious biases has been painfully slow due to the huge administrative undertaking involved (Rogers, 1993, p. 181). But some commentators feel that change is taking place too slowly. According to Knowlton (2005), these outdated headings “stigmatize certain groups of people with inaccurate or demeaning labels, and create the impression that certain points of view are normal and others are unusual” (p. 125).

Knowlton suggests that the literary warrant upon which LCSH is based is a catalyst for its biases, suggesting that 'a more neutral set of terms' should be adopted:

By utilizing the language and perspective of a particular group of readers, rather than seeking a more neutral set of terms, LCSH can make materials hard to find for other users, stigmatize certain groups of people with inaccurate or demeaning labels, and create the impression that certain points of view are normal and others unusual. (Knowlton, 2005, p. 125)

He does not, however, make any further suggestion on how to go about creating 'a more neutral set of terms' – and it could be said that this viewpoint is somewhat simplified. To create such a vast body of terms without the use of literary warrant would be a close to impossible task. We must also not forget that LCSH are used to catalogue literature, and not society itself, and therefore literary warrant is an obvious method of selecting headings.

There is also some evidence to suggest that the corrections made to LCSH could in fact be dangerous – Rogers points out that although the Library of Congress has implemented some relatively drastic revisions on headings concerning women, the remaining forms of bias in LCSH are subtle, and therefore “difficult to root out” (1993, p. 190) – and thus presenting the public with a mirage of neutrality, when, in fact, this is not the case – nor is it possible to create an entirely neutral scheme (Olson, 2000).

Some commentators, like Strottman (2007), argue the importance of correcting these biases “Given the current efforts to expand the influence of LCSH internationally, it is also crucially important to recognize and correct biased and inaccurate subject headings” (p. 61).

Furthermore, Rogers (1993) is of the view that these biases should be corrected in order to break a vicious cycle of prejudice:

Although using headings that reflect various popular prejudices might give users easier access, it would also encourage those prejudices and instil them in young reader. (p. 196)

Rogers also argues that the library has a responsibility to be fair to the people referred to in its catalogue in the same way that librarians have a responsibility to be fair to the users they deal with (p. 196). This presents a wider question surrounding the role of the librarian – it is to facilitate access to material as widely as possible, or to take a moral high ground?

Although dramatic changes have indeed taken place in an attempt to combat some of the more glaringly obvious biases (Rogers, 1993, p. 182), Olson makes it clear in her 2000 paper that to remove the bias from LCSH would be impossible:

LCSH has the power to create meaning whether that power is used consciously or not. It cannot be neutral because there is no neutrality or universal meaning – no

“primordial unity.” Therefore, it should be used with a consciousness of that power.
(p. 66)

Some commentators argue that simply accepting the bias is easier and better than correcting it, given the impossibility of a ‘perfect’ system:

Even though the library field is grappling with the inadequacies of LCSH, it is apparent that nothing better or more comprehensive exists and the vocabulary of the list is very rich. (Fischer, 2005, pp. 64-65)

This is further echoed by Olson, who points out that:

By choosing to follow the convenience of a singular public and the canon of literary warrant we are introducing a bias towards the mainstream status quo that is just as much a bias as any professional judgment we are likely to employ. (Olson, 2000, p. 65)

It is therefore accepted that the assignment task is by nature subjective, no matter how objective the guidelines (Yi, 2010, p. 1659).

2.3 LCSH and social tags

A number of studies have sought to compare LCSH and social tags, with the majority finding that the two resources are “distinct and respond differently in similarity measures” (Yi, 2010, p. 1669).

Hecker et al.. (2007) and Kipp and Campbell (2006) both came to similar conclusions, in that they found that the key difference between social tags and LCSH is that tags are social driven and are often quite distinct from professionally assigned indexing terms.

Despite these differences, an analysis of popular tags and subject terms by Yi and Chan (2009) found that 61 percent of social assigned tags can be found directly in LCSH, with a further 10 percent of the remaining tags having potential matches. However, Lu (2010) points out that users and experts tend to employ these shared terms with differing motivations and contexts (p. 766).

Rolla (2009) concludes that the fact that the users of LibraryThing assign tags to books representing concepts not brought out by LCSH indicates that cataloguers, by following the LC guidelines, may omit concepts that are important to users. (p. 178) He emphasizes the different roles the two resources play:

Thus, approximately three-quarters of the time, catalogers and readers agree on at least part of what each book is about, even if the tags and subject headings express the content of the book differently. (Rolla, 2009, pp. 178-179)

This suggests that both tools fulfill their purposes, but that these purposes differ.

It is also worth noting here that the differing roles of the two systems is highlighted by the prevalence of subjective and personal social tags such as 'to read', 'read in 2009' and 'unread' are among the top frequent tags (Lu et al.,

2010, p. 777). Yoon (2012) also notes that it is interesting that users select and assign terms for two reasons: to remind themselves of the material and also to make their tags open to and viewable by the public (p. 924).

According to Kipp (2007) “These tags show that users view classification as a holistic process closely tied to themselves and their work”, which differs to LCSH as “all LCSH terms are used to describe the subjects of the works whereas many tags are used for purposes other than subject description” (Lu et al., p. 772).

Although no study looking specifically at bias and social tags was found in the literature, a number of commentators point out the effects that such a person-driven system could have on other users.

Rolla highlights that LC’s attempt to avoid polemical topics and maintain an objective stance in their subject heading often reveals subtle biases that are not present in social tags:

In the example of the book *A Savage War of Peace*, the only LCSH in the bibliographic record is “Algeria—History—Revolution, 1954–1962.” This subject heading does not explicitly mention France and its involvement in the war, while conversely more of LibraryThing’s tags for this book cited France than Algeria. The library-supplied subject heading, then, subtly erases the anticolonial nature of the war. (Rolla, 2009, p. 181)

This then suggests that, based on Rolla's study, that social tags are less biased in nature. However, this conflicts with Mathes' (2004) viewpoint that:

In collaborative tagging, selecting and determining tags are most likely influenced by personal preferences, interpretations, and prejudices. Thus social tags share the problems of ambiguity in meaning and lack of any kind of semantic control, such as polysemy and synonymy, inherent to uncontrolled vocabularies. (Mathes, 2004)

Furthermore, while many commentators express concern about the fact that LCSH is based on the literary warrant of LC's collections, and therefore has an inherent US bias, the literature on social tags represents similar concerns given the higher usage of social media in the Western world.

Despite the fact that, as Lu et al. argue "tags are created by a large number of users with different backgrounds" (2010, p. 769), Rolla points out that "If a library serves a more general population overall, it does need to be aware that social tags may come predominantly from specific populations or communities" (2009, p. 176). This presents the idea that, given that the application of social tags is optional, only certain members of a library community will make the effort to apply tags themselves – therefore inadvertently excluding other user groups and allowing for bias.

Rolla also comments on the fact that LibaryThing's users have an advantage over cataloguers: they have probably read the book before tagging it on the website and are therefore able to base their subject analysis on the book as a whole. This differs from the cataloguer's role, who:

can also be constrained by the fact that they are trying to assign subject headings that will be meaningful for a large group of unknown and potentially diverse end users, and they may not know what subjects in a book will be most important to all potential readers. In LibraryThing, the cataloger and the end user are the same. (Rolla, 2009, p. 181)

Yoon (2012) also echoes this:

Social tags on documents in a social bookmarking site reflect a user's views of an information object, which can augment the content description and provide more effective representation of information. (Yoon, 2012 p. 923)

However, Yi's (2009) findings that the majority of social tags show correspondence to their LCSH counterparts imply that knowledge of the content has little relation to the tags or subject headings applied. Lu et al. (2010) also finds that, according to data analysis of the terms in titles that are used as tags and LCSH, users are more likely than experts to annotate books with terms from the titles (p. 777).

A major criticism of LCSH was the delay in which it reacts to change. However, social tags - as Spiteri (2006) has pointed out - can adapt better and more quickly to changing terminologies and to new fields of study than LCSH or any controlled vocabulary can (p. 77). Therefore social tags presumably have the ability to shrug off dated biases and prejudices in a way that is impossible for LCSH.

Furthermore, Lu et al. (2010) highlight the fact that Social tagging is done in a totally uncontrolled environment:

Social taggers do not need to master specific metadata standards or indexing rules for tagging. They apply their own verbal descriptors to resources that interest them. For this reason, the entrance threshold of metadata creation is lowered and a great number of resources can be tagged within a short time. Unlike professionals who create metadata, taggers annotate resources from multiple individual perspectives. (Lu, 2010, p. 764)

Once again, Yoon's study supports this:

Underlying social tagging is the aggregation of individual users' activity of freely describing and assigning labels without any formats or rules regarding how to generate them. The multiple semantics in social tags creatively define the content at multiple levels of perspective and facilitate dynamic searching beyond that which is possible with the traditional information representation that uses a rigid and unitary language model. (Yoon, 2012, p. 924)

The mention of multiple individual perspectives by both of these authors again highlights the differing purposes of the two systems, and raises the question as to whether one is more appropriate than the other in terms of objectivity and bias.

2.4 Conclusion

It is accepted that the application of any subject term to a bibliographic record is a subjective task, but a comparison of the way in the two distinct roles of LCSH and social tags portray the inherent biases and prejudices of those applying terms has yet to be fully explored.

The two systems have differing purposes – social tags being personal and social driven, while LCSH seeks to remain objective and structured with the intention of allowing subject access to a resource.

The review of the literature has raised a number of questions that invite further research – the central one being: Do the personal motivations and the uncontrolled vocabulary of social taggers mean that their personal biases, prejudices and worldviews present themselves differently from those of the cataloguers?

3. Methodology

3.1 Introduction

The following chapter outlines and justifies the approaches chosen for the study. It describes the methods used and the criteria for sampling. It also outlines the structure for the coding scheme, with particular emphasis on the definition on bias and the methods for quantifying this. Furthermore, this chapter briefly outlines limitations to the study and highlights ethical implications of such a piece of research.

3.2 Approaches to the research

3.2.1 Mixed method approach

It is accepted that research strategies can be broadly defined as either quantitative or qualitative: “the former uses numerical representations to quantify occurrences, while the latter uses words to present anecdotal descriptions” (Gorman and Clayton, 2005, p. 10). Put simply, quantitative research deals with numbers, and qualitative research deals with words. Each approach has its flaws: the experimental nature of quantitative research leads to an inevitable level of artificiality (Bryman, 2012, p. 79), while the subjectivity of qualitative research implicates credibility and transferability (Pickard, 2007, p. 20).

However, according to Bryman (2012), “research methods are much more free-floating than is sometimes supposed” (p. 614); and the notion that research must be either quantitative *or* qualitative is criticized by commentators such as

Gorman and Clayton (2005), who argue “painting such clear distinctions can be misleading, if not downright inaccurate” (p. 12) and Feyerand, (1975, p. 296) in Pickard (2007, p. 13), who states that “both methodologies have their limitations and the only “rule” that survives is “anything goes””.

Based on these assessments, this study takes a mixed methods approach. A large amount of data was produced, necessitating the need content analysis in the form of coding in order to organize the data.

Although content analysis through coding achieves a more objective and systematic set of results (Bryman, 2012, p. 289), it is not without limitations. In particular, a coding manual will inevitably entail some interpretation on the part of the coder (Bryman, 2012, p. 306); especially given the nature of the subject of bias. Furthermore, according to Bryman “It is difficult to ascertain the answers to ‘why?’ questions through content analysis” (2012, p. 307).

Therefore, in an attempt to counteract these limitations, the results were analyzed qualitatively. This is because the subject of bias is, by nature, a philosophical one, which lends itself better to written discussion rather than statistics and graphs.

3.3 Research methods

This piece of research utilized the following research methods:

- Literature review
- Pilot Study

- Data Selection and Collection (via bibliography sampling)
- Data Organization and Presentation
- Tag and LCSH Evaluation (via content analysis and coding)

3.3.1 Literature review

The literature review was conducted in order to ascertain the following criteria:

- To clarify research aims
- To provide the depth and breadth of subject knowledge necessary
- To form the theoretical framework for empirical investigation
- To contribute to research design

(Adapted from Pickard, 2007, p. 26)

The literature review was structured, dealing with the key aspects of the research question under separate headings. The use of the abstracting service LISA revealed a number of key studies on the subjects of social tagging, LCSH and the existence and impact of bias within these systems. Further literature was obtained through the process of 'snowballing' via the bibliographies of the key studies.

3.3.2 Pilot study

Ten titles selected from an online bibliography of multicultural fiction from the website GoodReads (<https://www.goodreads.com/>) as the basis for a pilot study (see Appendix A). The sample for the main study consists of a bibliography of

feminist fiction; therefore, given that the literature suggests that multicultural fiction is similarly problematic in terms of biased terminology as feminist literature (Berman, 1992; Clack, 1994), the topic was considered to be relevant enough to the study without replicating it. The pilot study was conducted in order to assess the following:

1. The data collection methods
2. The quality of the coding scheme
3. The appropriateness of the research question

The pilot study confirmed points 2 and 3. However certain aspects of data collection were problematic. The sheer number of LibraryThing tags attached to each record meant that the decision was taken to include only the top ten tags (in terms of number of users applying each tag).

Furthermore, it was observed that LibraryThing tags frequently described genre, rather than subject. This was a concern given that it is difficult to compare genre tags directly with LCSH. However, it was sometimes difficult to distinguish genre and subject tags in LibraryThing (e.g. Lesbian/Lesbian fiction) so it was therefore decided not to exclude these tags, but to acknowledge the issue in the study's Discussion chapter.

3.3.3 Data selection

LibraryThing (<https://www.librarything.com>) is a social networking site that allows users to catalogue and tag their book collections. Described as “Facebook for books”, it is one of the largest networks of this type with over 1,800,000

registered users (LibraryThing, 2015). This means that it was an obvious choice in order to ensure a high volume of social tags. It has also been recognized in a number of papers as an important tool for the study of social tagging in practice (Yi, 2010; Rolla, 2009).

The LCSH were taken from the Library of Congress online catalogue. This is justified given that the Library of Congress is responsible for LCSH.

Although the data utilized in this study was openly available in the public domain, permission was granted from both LibraryThing and the Library of Congress prior to conducting the study.

3.3.3.1 Sampling

It was considered important to use a sample that would clearly demonstrate any signs of bias in both subject headings and LCSH. Alongside religious literature (Knowlton, 2005; Wilk et al., 2001) and multicultural works (Berman, 1992; Clack, 1994), literature relating to women is perceived to be highly problematic in terms of LCSH (Fischer, 2005; Berman, 1993; Rogers, 1993), yet a direct comparison between the way in which subjects relating to women are represented through LCSH and LibraryThing tags has yet to be explored. It was therefore decided that a bibliography sample of feminist literature would be selected.

The sample was selected from a bibliography of 598 popular feminist fiction books from the website GoodReads. A systematic sample was selected of the most popular titles. The aim of selecting popular works was to ensure that the

more widely read books were selected, which is likely to result in richer and more varied social tags.

The following criteria were then applied to the bibliography:

- The title should be an individual piece of literature rather than a series of books
- The title must be fiction
- At least one edition of the title should be present on both the Library of Congress catalogue and LibraryThing.
- At least one tag and one subject heading should be attached to the title in both of the sources

The bibliography was assessed against the criteria until fifty titles were chosen.

If a title did not meet the sampling frame, it was excluded from the sample.

3.3.4 Data organization and presentation

Due to the large amounts of data, a systematic approach was necessary in terms of its organization. The sample bibliography was listed in a Microsoft Word document. A Microsoft Excel document was created with two sheets: one titled 'LibraryThing tags' – for the extracted tags, and one titled 'LCSH' – for the extracted subject headings.

In the extraction of LibraryThing tags, the following types of tag were omitted, given their lack of relevance to the research question:

- Personal tags (e.g. to read, read, book club)

- Author names as tags
- The tag 'fiction', given that the sample consists only of fiction material and the tag is deemed irrelevant for the purposes of this study
- Physical descriptions (e.g. picture book, ebook, hardback)
- Date or period of the work's publication

After filtering the tags for the above criteria, the ten most inputted tags for each title were inputted into the 'LibraryThing tags' sheet.

The LCSH for each title were inputted into the 'LCSH' sheet. Again, 'fiction' subheadings were omitted for the reasons outlined above.

3.3.5 Data analysis

Content analysis, as “an approach [...] that seeks to quantify content in terms of predetermined categories in a systematic and replicable manner” (Bryman, 2012, p. 291), was employed to analyze the data. The aim of using data analysis was to be able to convert the principally philosophical notion of bias into data that can be analyzed quantitatively and systematically.

3.3.5.1 Coding

Bryman describes coding as “a crucial stage of the process of doing a content analysis” (2012, p. 298), and he places a great deal of emphasis on the quality of the coding manual to ensure consistency and reliability (Bryman, 2012, p. 303). Therefore, in order to ensure a reliable coding scheme, it was deemed necessary

to seek to define the various aspects of bias that might occur in the process of tagging and applying subject headings.

3.3.5.1.1 Defining and measuring bias

In their meta-analysis of research and reported experience relating to subject access for marginalized and topics, Olson and Schegl (2001) cite Jessica Milstead and Doris Hargrett Clack's 1979 model for measuring and defining bias. Their study analyzed subject headings and classification numbers to determine:

(1) If the subject analysis is prejudiced; i.e., if its slant is negative; (2) if the vocabulary is objective; i.e., are terms still in use that today are considered pejorative? (3) if the subject analysis is, to the best of the observer's ability to determine from available literature, offensive to the affected group; and (4) if the subject analysis provides access via the terms the searchers at whom the work is directed would be likely to use; specifically, whether books obviously aimed at improving understanding of a group could be accessed through the name of the group or another term associated with it.

(Harris & Clack, 1979, pp. 376-77 in Olson and Schegl, 2001, pp. 64-65)

Olson and Schlegl adapt this model slightly to include the following categories:

- treatment of the topic as an exception
- ghettoization of the topic
- omission of the topic
- inappropriate structure of the standard

- biased terminology

(Olson and Schlegl, 2001, pp. 64-65)

Maggio discussed this in her book *Talking About People: A Guide to Fair and Accurate Language*, published in 1997. She lists six ways in which biased language refers to people:

- It leaves out certain individuals or groups.
- It makes unwarranted assumptions.
- It calls individuals and groups by names or labels that they did not choose for themselves.
- It is based on stereotypes that imply that all lesbians/Chinese/women/people with disabilities/men/teenagers are alike.
- It treats groups in nonparallel ways in the same context.
- It categorizes people when it is unnecessary to do so and when this is not done in similar cases.

(Maggio, 1997, p.2)

3.3.5.1.2 Coding manual

A coding manual was devised based on a simplified combination of the three definitions outlined in this chapter.

Categories were selected based on their suitability for the study's research methods. Points such as 'omission of the topic' were discounted, for example, due to the fact that it was not possible to have a good enough knowledge of the

literary content of the entire sample to be certain whether key topics were intentionally omitted. Furthermore, although Olson and Schlegl’s category of ‘inappropriate structure of the standard’ can be applied to LCSH, social tags generally (although not always) appear as individual words, meaning that this analysis cannot be applied to them.

Table 1: Outline of coding system used to organize headings and tags

Category	Code
Based on stereotypes	ST
Object as exception	EX
Offensive terminology	OF
Unbiased	UN

In order to apply the categories systematically, the following guidance was applied:

i. Based on stereotypes

Maggio’s statement that biased language is “based on stereotypes that imply that all lesbians/Chinese/women/people with disabilities/men/teenagers are alike” (1997, p. 2) is also related to Olson and Schlegl’s notion of ‘ghettoization of a subject’: “the problem of gathering and then isolating a topic, rather than integrating it” (2001, p. 67). Thus, this code can be applied to any terminology that infers a stereotype (e.g. Gays, Teenagers, Mexican Americans).

Although it could be argued that this terminology is merely descriptive rather than stereotypical, the issue here is related to the how the tag or heading is

phrased: for example, the stereotypical term 'Mothers' could be replaced by 'Motherhood' or 'Mothering' in order to eliminate this bias of targeting a specific group of people.

ii. Treating the topic as an exception

Olson and Schlegl take the treatment of the topic as an exception to mean that:

The topic was represented outside some accepted norm. Examples [include] the now obsolete LCSH headings for "women as...", such as "women as physicians," that seem to express astonishment that such anomalous creatures should exist. (2001, p.67)

More broadly, treating the topic as an exception can also be applied to tags and subject headings such as 'Gay teenagers' or 'male friendship' given that there is an implication that homosexual teenagers are distinct from other teenagers, or that male friendship differs from other forms of friendship.

It should also be noted that the vast majority of tags and headings that treat the topic as an exception are also based on stereotypes (e.g. Gay teenagers).

However, for the purposes of this study, the two dimensions were treated as separate categories.

iii. Offensive terminology

Although it is sometimes difficult to know what may or may not be offensive to a particular individual or group, Milstead and Clack state that "if the subject

analysis is, *to the best of the observer's ability to determine from available literature, offensive to the affected group*" (1979, p. 376, emphasis added), and Maggio simply states that offensive language is the naming of "individuals and groups by names or labels that they did not choose for themselves" (1997, p. 2). Moreover, Olson and Schlegl point out the offensiveness of naming people as an adjective:

An ongoing discussion with the American Library Association regarding the subject headings for people in poverty is the suggestion from Sanford Berman that naming people only with an adjective, *Poor*, is not considerate. (2001, p. 67)

The code can be applied to tags or headings that may cause offense to some members of a group, but not others. An example being 'queer' – a term many homosexuals choose define their own culture and lifestyle, yet the history of the term as an insult still has the capability to cause offence, therefore contradicting Maggio's statement.

Furthermore, Sanford Berman's suggestion that naming people only with an adjective (e.g. *Poor*) is inconsiderate and therefore offensive (Berman, 1993 in Olson and Schlegl, 2001, p. 69).

iv. Unbiased

This code can be applied to all other headings or tags that do not meet any of the above criteria.

3.4 Limitations of this approach

This approach sought to analyze tags and headings independently of the records attached to them; this meant that potential findings in terms of a direct comparison of specific records were not investigated. This was unavoidable given the length constraints of the study.

Furthermore, although care was taken to ensure the coding scheme was systematic and transparent, by its very nature, the topic of bias will require some interpretation on the part of the coder. This was overcome by the creation of a clear manual and a policy of objectivity.

3.5 Ethics

Aberystwyth University's policies for Ethical Practice in Research and the Department of Information Studies Ethics Police were adhered to during the process of this research. Permission was obtained from both LibraryThing and the Library of Congress for the extraction and analysis of tags and headings.

4. Results

4.1 Introduction

Assessing the ways in which the presentation of bias appears through the application of social tags and LCSH requires implementing a somewhat rigid definition of bias as presented in the methodology of this study. This is arguably an unrealistic approach to what is considered a predominantly ambiguous subject; however this approach ensures consistent and systematic results.

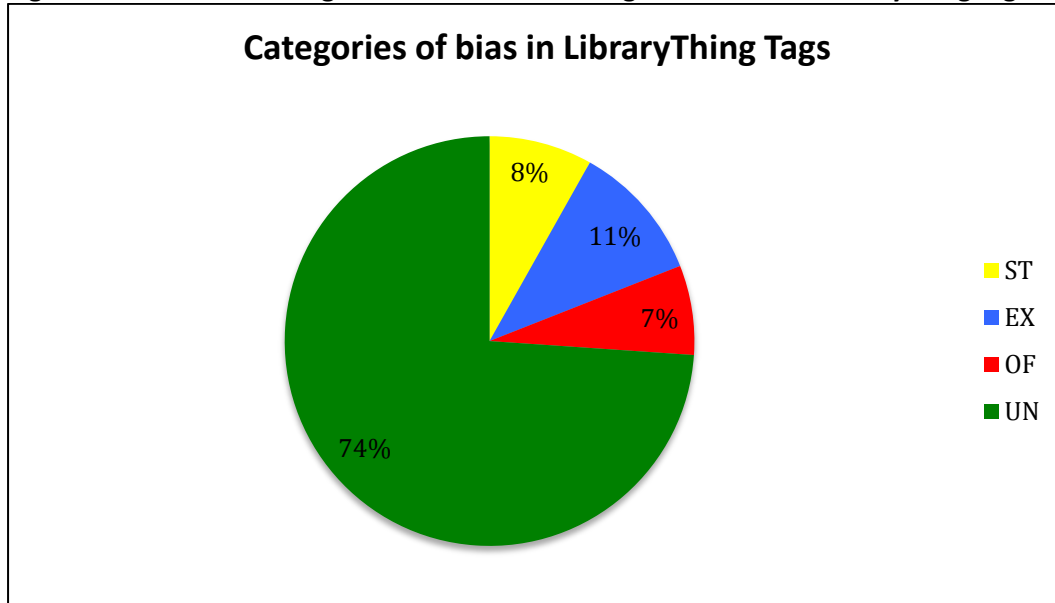
4.2 The representation of bias through application of social tags and LCSH

Across the fifty records assessed, 183 disparate social tags and 130 LCSH were recorded. Many terms appeared in more than one record (39 percent of social tags and 18 percent LCSH). This demonstrates a clear trend towards a significantly higher number of social tags being applied to records.

4.2.1 Social tags: Categories of bias

The results show that 74 percent of LibraryThing tags were considered unbiased (UN); 11 percent treated the object as an exception (EX); 8 percent were based on stereotypes (ST); and 7 percent comprised of offensive terminology (OF).

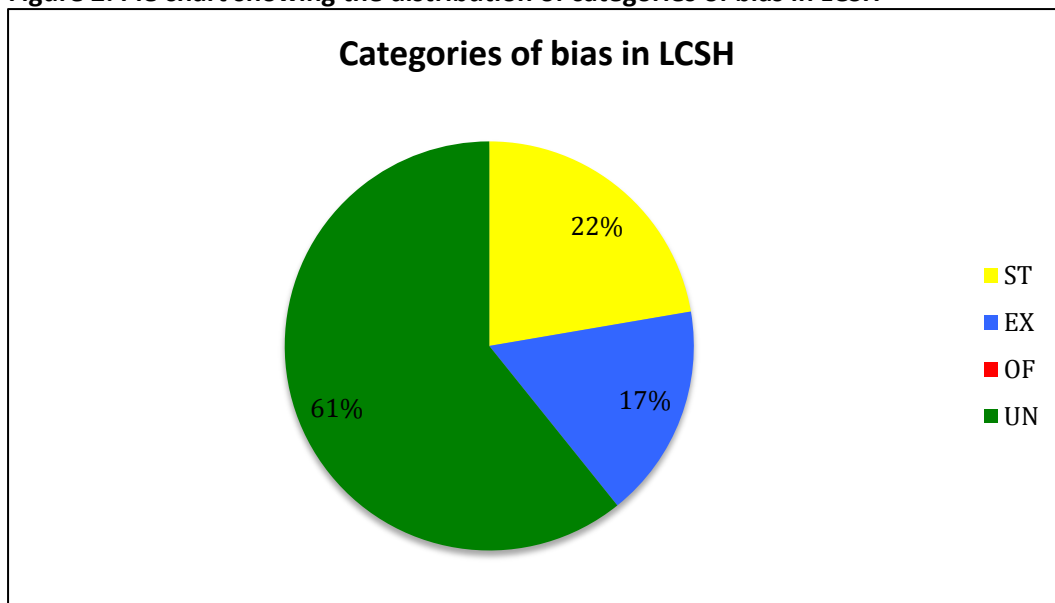
Figure 1: Pie chart showing the distribution of categories of bias in LibraryThing tags



4.2.2 LCSH: Categories of bias

The results show that 61 percent of LCSH were considered unbiased (UN); 22 percent were based on stereotypes; 17 percent treated the object as an exception (EX); and none of the LCSH analyzed were considered to comprise of offensive terminology (OF).

Figure 2: Pie chart showing the distribution of categories of bias in LCSH



4.3 Analysis of the ways in which social tags and LCSH present bias

4.3.1 Tags and headings based on stereotypes

15 LibraryThing tags (8 percent of the tags analyzed) and 29 LCSH (22 percent of the headings analyzed) were considered as being based on stereotypes. Of the 44 tags and headings analyzed, eight identical or interchangeable terms (e.g. suffragists/suffragettes) (18 percent) appear in both locations.

Table 2: LibraryThing tags and LCSH based on stereotypes

LibraryThing Tags	Number of occurrences	LCSH	Number of occurrences
women	31	Lesbians.	5
lgbt	4	Sisters.	4
sisters	3	African Americans.	3
lesbians	2	Fathers and daughters.	3
Taliban	1	Girls.	3
suffragettes	1	Women.	3
mothers and daughters	1	Americans--Congo (Democratic Republic).	1
maids	1	Beekeepers.	1
immigrants	1	Feminists.	1
grandmother	1	Gangs.	1
governess	1	Governesses.	1
gangs	1	Household employees.	1
elderly	1	Missionaries.	1
African Americans	1	Mothers and daughters.	1
		Mothers--Death.	1
		Orphans.	1
		Sisters--Death.	1
		Slaveholders.	1
		Slaves.	1
		Suffragists.	1
		Suicide victims.	1
		Transsexuals.	1
		Widowers.	1
		Widows.	1
		Women--Alabama.	1
		Women--Canada.	1
		Women--Louisiana.	1
		Women--New York (State)--New York.	1
		Women--Suffrage.	1

4.3.2 Tags and headings that treat the object as an exception

20 LibraryThing tags (11 percent of the tags analyzed) and 22 LCSH (17 percent of the headings analyzed) were considered to treat the object as an exception. Of

the 42 tags and headings analyzed, one identical term (teenage girls) appears in both locations.

Table 3: LibraryThing tags and LCSH that treat the object as an exception

LibraryThing Tags	Number of occurrences	LCSH	Number of occurrences
American literature	17	African American women.	7
Canadian literature	7	Teenage girls.	3
African-American Literature	7	Female friendship.	2
Canadian fiction	6	Married people.	2
British literature	6	Married women.	2
English literature	5	Mentally ill women.	2
Lesbian Fiction	2	Middle-aged women.	2
young adult	1	Young women.	2
teenage girls	1	Abused wives.	1
Southern Literature	1	Adult child sexual abuse victims.	1
southern fiction	1	African American families.	1
girl gangs	1	English--Scotland.	1
feminist SF	1	Married women--Psychology.	1
feminist fiction	1	Older women.	1
Christian literature	1	Poor women.	1
children's literature	1	Women college students.	1
children's	1	Women in the Bible.	1
American fiction	1	Women murderers.	1
African American Literature	1	Women novelists.	1
African American Fiction	1	Women painters.	1
		Women slaves.	1
		Working class women.	1

4.3.3 Tags and headings that are comprised of offensive terminology

13 LibraryThing tags (7 percent of the tags analyzed) and 0 LCSH (0 percent of the headings analyzed) were considered to comprise of offensive terminology.

Table 4: LibraryThing tags and LCSH comprised of offensive terminology

LibraryThing tags	Number of occurrences	LCSH	Number of occurrences
American	20		
feminist	8		
African American	8		
lesbian	7		
Canadian	7		
British	7		
southern	5		
English	5		
queer	4		
pagan	1		
Irish	1		
gay	1		
Celtic	1		

4.4 Geographical and racial references: a matter of context

It was observed that a significant proportion (20 percent) of the LibraryThing tags considered as bias referred to geographic location or race, (compared with only 4 percent of LCSH). These tags frequently refer to the literature of a specific country or race (e.g. African American literature, British literature), and therefore were categorized as treating the object as exception (i.e. isolating a certain body of literature from literature as a whole). Many tags of this type are ambiguous as to whether they refer to the literature or people (e.g. British); and although this trait can also be observed with other tags (e.g. feminist), it was felt that given the disproportionate number of tags in this category, further investigation into other areas of the findings was required.

4.4.1 The removal of geographical/racial social tags and LCSH

Geographical and racial tags were removed from the results and the findings are reported below.

Table 5.1: Geographical and racial LibraryThing tags removed from the study

Geographical/racial LibraryThing tags	Number of occurrences
African Americans	1
American literature	17
African-American Literature	7
Canadian literature	7
British literature	6
Canadian fiction	6
English literature	5
African American Fiction	1
African American Literature	1
American fiction	1
southern fiction	1
Southern Literature	1
American	20
African American	8
British	7
Canadian	7
English	5
southern	5
Celtic	1
Irish	1
England	8
Canada	7
South	5
New York	3
USA	3
Africa	2
Alabama	2
American South	2
London	2
Ohio	2
South Carolina	2
Toronto	2
Afghanistan	1
America	1
Florida	1
Ireland	1
Louisiana	1
Middle East	1
Mississippi	1
New Orleans	1
New York City	1
Quebec	1
Texas	1
the south	1

Table 5.2: Geographical and racial LCSH removed from the study

Geographical/racial LCSH	Number of occurrences
African Americans.	3
Americans--Congo (Democratic Republic).	1
Women--Alabama.	1
Women--Canada.	1
Women--Louisiana.	1
Women--New York (State)--New York.	1
African American women.	7
African American families.	1
English--Scotland.	1
England.	5
Ohio.	3
Southern States.	3
New York (N.Y.).	2
Afghanistan.	1
Alabama.	1
Appalachian Region, Southern.	1
Congo (Democratic Republic).	1
Canada.	1
Georgia.	1
Greenville County (S.C.).	1
Family life--New England	1
Ireland--Antiquities.	1
Ireland--Church history--To 1172.	1
London (England).	1
Los Angeles (Calif.).	1
Michigan.	1
New England--History--19th century.	1
New Orleans (La.).	1
New York (State).	1
Québec (Province).	1
Romans--Ireland.	1
Skye, Island of (Scotland).	1
Southern States--Social life and customs.	1
Toronto (Ont.).	1

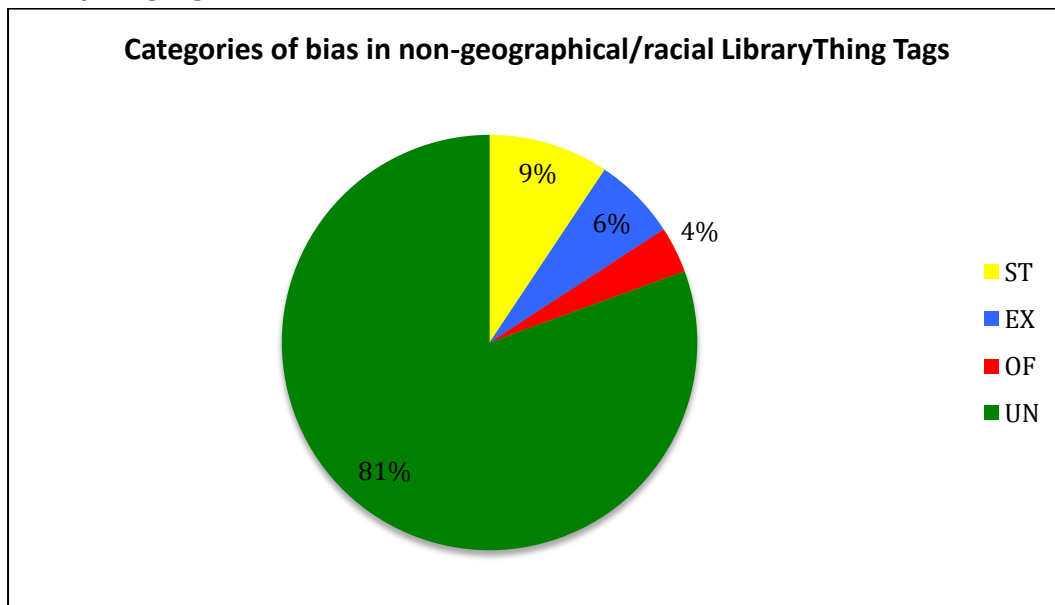
4.5 The representation of bias through application of non-geographic/racial social tags and LCSH

The removal of geographic and racial tags and headings left disparate 139 social tags and 96 LCSH.

4.5.1 Non-geographical/racial social tags: Categories of bias

The results show that 81 percent of non-geographical/racial LibraryThing tags were considered unbiased (UN); 6 percent treated the object as an exception (EX); 9 percent were based on stereotypes (ST); and 4 percent comprised of offensive terminology (OF)

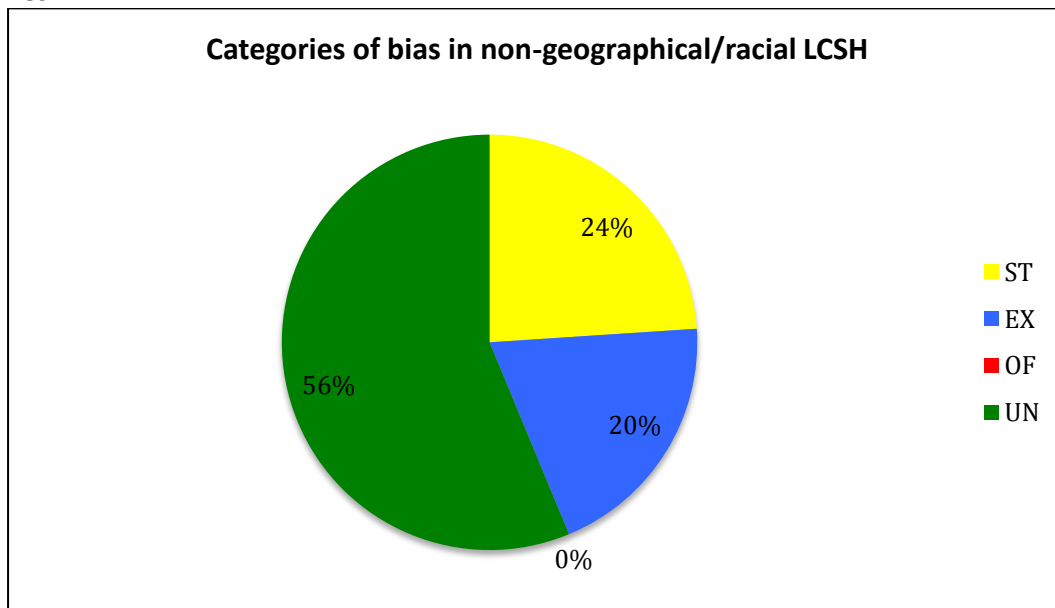
Figure 3: Pie chart showing the distribution of forms of bias in non-geographical/racial LibraryThing tags



4.5.2 Non-geographical/racial LCSH: Categories of bias

The results show that 56 percent of non-geographical/racial LCSH tags were considered unbiased (UN); 20 percent treated the object as an exception (EX); 24 percent were based on stereotypes (ST); and none of the LCSH analysed were considered to comprise of offensive terminology (OF).

Figure 4: Pie chart showing the distribution of forms of bias in non-geographical/racial LCSH



4.6 Analysis of the ways in which non-geographical/racial social tags and LCSH present bias

4.6.1 Non-geographical/racial tags and headings based on stereotypes

13 LibraryThing tags (9 percent of the tags analyzed) and 23 LCSH (24 percent of the headings analyzed) were considered as being based on stereotypes. Of the 36 tags and headings analyzed, seven identical or interchangeable terms (e.g. suffragists/suffragettes) (19 percent) appear in both locations.

Table 6: Non-geographical/racial LibraryThing tags and LCSH based on stereotypes

LibraryThing tags	Number of occurrences	LCSH	Number of occurrences
women	31	Lesbians.	5
lgbt	4	Sisters.	4
sisters	3	Fathers and daughters.	3
lesbians	2	Girls.	3
Taliban	1	Women.	3
sufferalettes	1	Beekeepers.	1
mothers and daughters	1	Feminists.	1
maids	1	Gangs.	1
immigrants	1	Governesses.	1
grandmother	1	Household employees.	1
governess	1	Missionaries.	1
gangs	1	Mothers and daughters.	1
elderly	1	Mothers--Death.	1
		Orphans.	1
		Sisters--Death.	1
		Slaveholders.	1
		Slaves.	1
		Suffragists.	1
		Suicide victims.	1
		Transsexuals.	1
		Widowers.	1
		Widows.	1
		Women--Suffrage.	1

4.6.2 Non-geographical/racial tags and headings that treat the object as an exception

Nine LibraryThing tags (6 percent) and 19 LCSH (20 percent) were considered to treat the object as an exception. Of the 28 tags and headings analyzed, one identical or term (teenage girls) appears in both locations.

Table 7: Non-geographical/racial LibraryThing tags and LCSH that treat the object as an exception

LibraryThing tags	Number of occurrences	LCSH	Number of occurrences
Lesbian Fiction	2	Teenage girls.	3
young adult	1	Female friendship.	2
teenage girls	1	Married people.	2
girl gangs	1	Married women.	2
feminist SF	1	Mentally ill women.	2
feminist fiction	1	Middle-aged women.	2
Christian literaure	1	Young women.	2
children's literature	1	Abused wives.	1
children's	1	Adult child sexual abuse victims.	1
		Married women-- Psychology.	1
		Older women.	1
		Poor women.	1
		Women college students.	1
		Women in the Bible.	1
		Women murderers.	1
		Women novelists.	1
		Women painters.	1
		Women slaves.	1
		Working class women.	1

4.6.3 Non-racial/geographical tags and headings that are comprised of offensive terminology

5 LibraryThing tags (4 percent) and 0 LCSH (0 percent) were considered to comprise of offensive terminology.

Table 8: Non-geographical/racial LibraryThing tags and LCSH that are comprised of offensive terminology

LibraryThing tags	Number of occurrences	LCSH	Number of occurrences
feminist	8		
lesbian	7		
queer	4		
pagan	1		
gay	1		

5. Discussion

5.1 Introduction

This study was conducted in order to investigate the relationships between social tags and LCSH in terms of their representation of bias. The aim of the study was to discover whether the ‘expert’ (i.e. the cataloguer) is more or less prone to bias than the layperson (i.e. the LibraryThing user).

5.2 Statement of results

The results showed that, although there were a higher proportion of unbiased social tags than unbiased LCSH, there are issues of bias in both systems. This result was expected, given that it is accepted that where interpretation is required, bias is inevitable (Olson, 2000). However, both systems differ in the way that this bias is presented, and therefore the ‘layperson’ and the ‘expert’ have differing approaches in how they display their biases.

5.2.1 Issues of bias in social tags

The key finding in terms of how bias is presented in social tags was that, unlike LCSH, offensive terminology was found to exist. Although this was a comparatively small proportion (7 percent of tags in total), the complete lack of offensive terminology in LCSH makes this result a significant one.

It is difficult to draw conclusions on the motivations of taggers surrounding the offensive terminology. However, analysis of the terms used suggests that the issue is one of context, rather than prejudice.

Rorissa's 2008 observation that Flickr tags are "richer in their semantic context than index terms, which are at times devoid of context" (p. 1745) contradicts with the findings of this study. Many of the social tags in this piece of research were considered biased due to their lack of context. For example, the social tag "gay" was considered offensive given Sanford Berman's suggestion that naming people only with an adjective is not considerate (Olson and Schlegl, 2001, p. 67). However, it is likely that the intention of the tagger was to highlight that the work was considered to be gay literature, or featured a gay protagonist. However, without this context, the tag still has the potential to be considered offensive.

Furthermore, the use of potentially offensive tags such as 'queer' presents an interesting notion that social tagging systems represent a form of democracy - a point also made by Spiteri and Tarulli in their 2012 study. Through this democratic process, users are taking ownership of these tags and thereby reappropriating terminology that would otherwise be considered offensive if placed in the context of a single authority (or 'expert') applying the term.

In terms of other forms of bias in social tags, it was found that a number of tags also existed under the categories of being based on stereotypes and treating the object as an exception (8 percent and 11 percent of tags in total respectively).

Furthermore, it was found that when geographic/racial social tags were removed, the proportion of tags that treat the object as an exception dropped from 11 percent to 6 percent (with tags based on stereotypes remaining

relatively unchanged). This suggests particular issues of bias (or the interpretation thereof) in these areas. An explanation for this this shift in terminology that treats the object as an exception when geographic/racial tags were removed could, once again, be a case of context.

Knowlton (1995) describes issues of geographical biases in LCSH. However, the findings of this study demonstrate that social tags are even more problematic in terms of their use of geographic and racial tags.

The findings of this study also highlight a tendency for taggers to categorize literature by country of origin, which is also observed in Rolla's 2009 study. The issues with regard to context are apparent here: tags such as 'American', 'British' and 'African American' were common, and were ambiguous as to whether they referred to literature, people, or culture – thus meaning that the tagger's motivations (and therefore biases) were unclear.

Under the same category, it was noted that the tags such as 'American literature', 'Canadian literature' and 'African-American Literature' were frequently applied. Under the definitions of bias outlined in the methodology for this study, these were considered to treat the object as an exception (due to the fact that they isolate a certain nation's literature from literature as a whole), and therefore the shift in the types of bias seen when geographic/racial tags were removed is explained by this tendency for taggers to 'pigeonhole' literature by its country of origin. (It should be noted here, that, although this finding could also infer a bias

towards Western nations, the fact that the vast majority of works in the sample were of Western origin, means that conclusions cannot be made based on this).

It was also observed that this 'pigeonholing' also took place by genre or type of literature and not just by country of origin. Biased terms such as 'lesbian fiction' and 'feminist' appeared in multiple records, as did unbiased terms such as 'historical fiction', 'classic'. This suggests that that the role of the tagger is not simply to provide subject access in terms of content (as is the role of the cataloguer in their application of headings), but also to categorize the genre or type of literature.

5.2.2 Issues of bias in LCSH

Biased LCSH were found to be split fairly equally between stereotypical terminology and headings that treat the object as an exception, with only a slightly higher proportion of headings under the stereotypical category. This fairly even distribution was seen in both the LCSH with geographic/racial headings included (22 percent stereotypical terminology and 17 percent treating the object as an exception) and when geographic/racial headings were removed (24 percent stereotypical terminology and 20 percent treating the object as an exception).

It is generally accepted that, that, given the fact that the purpose of LCSH is to categorize subject matter, biased headings in terms of stereotypical terminology and treating the object as an exception are not only unavoidable, but also necessary if the system is to serve its purpose in terms of providing subject

access. However, this study also revealed that more problematic forms of bias highlighted in previous studies also appear in a number of records.

The finding that none of the headings analyzed contained offensive terminology shows that LCSH has indeed moved on in certain problematic areas (Berman, 1993); however, given the choice of sample, gender bias was particularly apparent in the findings of this study. This is something that Rogers highlighted in her 1993 paper over twenty years ago, thus demonstrating that little appears to have changed since then. In particular, this study has highlighted a number of problematic headings that isolate women (and therefore treat them as an exception) are still in use (e.g. 'Mentally ill women' and 'Women murderers') where headings describing men in the same way do not exist (e.g. there is no 'Mentally ill men' heading).

This highlights the LC's painfully slow approach to change as highlighted in Fischer's 2005 review of the critical views of LCSH, who summarizes that most commentators on the subject cite the slow changes as result of long delays by LC to add new terms to the list as being most problematic area for the system (p. 94).

5.2.3 Comparisons between the two systems

Other studies have indicated social tags and LCSH bear a number of similarities in terms of how experts and users share common terms when annotating records (Lu et al. 2010) with Yi and Chan (2009) finding that around 61 percent of social assigned tags can be directly found in LCSH, with a further 10 percent of

the remaining tags having potential matches. However, the findings of this study show that, of the 99 headings and tags considered as being biased, only nine terms were found to be identical or interchangeable in both systems (9 percent); indicating that, in terms of how the two systems present bias, the differences are more distinct.

The lack of identical or interchangeable biased terminology in both systems indicates a contradiction in Olson's (2000) suggestions that through the application of LCSH "we are introducing a bias toward the mainstream status quo" (p.65) and Rogers' (1993) view that "using [LCSH] that reflect various popular prejudices [...] might also encourage those prejudices and instill them in young readers" (p. 196).

Furthermore, the issue of geographical and racial terminology demonstrates a clear difference in the way in which LibraryThing users categorize literature. This suggests an inclination by taggers towards 'pigeonholing' literature into its country of origin, genre, or type of literature, rather than using the system to tag the subject matter covered in the work.

5.3 Recommendations for further research

This piece of research represents a starting point for further investigation on the subjects of bias within LCSH and social tags. However, the limitations on this study meant that a number of aspects of the research could not be explored fully.

The results of this study highlighted the importance of context with regard to the application of social tags. Therefore, further insight into the intentions of taggers could be gained from a more in-depth investigation of their motivations in the application of tags to a record, as well as why members of a particular community choose not to contribute to user tags (thereby allowing for dominant user groups and worldviews); and- in the case of there being a social tagging functionality within a library catalogue - whether taggers are influenced by the LCSH already present.

6. Conclusion

6.1 Introduction

Both systems contain biased terminology, however the lack of overlap in biased terms between the two systems suggests strongly that ‘laypersons’ and ‘experts’ differ in their application of bias. The following two sections will explore these differences in an attempt to answer the research question.

6.2 The layperson and the expert: approaches to bias

6.2.1 The layperson

According to the FAQs on the LibraryThing website, the purpose of the tagging capacity is for the user to categorize their own books according to how they think of them “not how some library official does” (LibraryThing, 2015) to enable them to search for and locate books in their own collections. Thus, the motivation of the tagger is a personal one, albeit in a public capacity.

This personal context means that there is no need – from the user’s point of view – for context. A LibraryThing user uses the tags as a form of personal shorthand. They know, for example, that their tag ‘American’ refers to the American novels in their collection, and not to American people or culture, or whatever else ‘American’ might be perceived to mean by another user.

Yet the fact that these tags are in the public domain means that they are open to interpretation by others, meaning that they can be perceived as biased, or even

offensive, to other users. However, according to Postigo (2011), Web 2.0 information, by its very nature, should always remain open to reinterpretation by knowledge of the crowd, and it is this openness to interpretation that differentiates it Web 1.0.

LibraryThing users know that the tags are not assigned by experts; and they are aware that tags are most likely influenced by other user's personal preferences, interpretations, and prejudices. This gives the social tagging community a democratic nature, meaning that the users own the tags – not “some library official” (LibraryThing, 2015)– and the fact that LibraryThing displays tags by popularity means that the top tags for the vast majority of records will have been applied by several users, therefore contributing to this democracy.

This is arguably part of the beauty of a folksonomy; and it requires tags to be viewed from a very different perspective. This, therefore, is a key difference between the presentation of bias in social tags and LCSH, and means that the two should be viewed as two separate non-interchangeable systems within this context.

6.2.2 The expert

The primary purpose of LCSH is to allow the user access to the records they require through its controlled vocabulary. In order to fulfill this purpose, it could be said that bias is not only unavoidable (as we have ascertained several times during the course of this study), but also necessary: In order to provide a formal description of the subject content of any bibliographic unit some form interpretation is required, which in turn allows for inevitable biases, both in

terms of stereotypical language and – in order to describe the book in the most precise terms possible – treating the object as an exception.

However, although it is a fairly stable system, the issues regarding terminology in LCSH, and its lack of any clear rules concerning the application of terms, mean that it is considered problematic in terms of inconsistent, outdated, and biased headings.

Problematic headings continue to exist (given the slow pace of change to the system); and, if we accept that central to the purpose of LCSH is to describe a work in the most precise terms possible, then headings such as ‘mentally ill women’ are considered correct, despite the lack of ‘mentally ill men’ counterpart.

Furthermore, LCSH’s preference for direct entry allows for headings that isolate people or cultures. For example, in the case of the heading ‘Women murderers’, ‘Murderers – Women’ would arguably be more preferable, both in terms of bias and subject access. Therefore, the issues of bias within LCSH stem not from the experts applying the headings, but are deeply (and arguably irreversibly) imbedded in the system itself.

The role of the cataloguer is therefore a very different one to that of a tagger. They are not assigning headings for their own use – rather, it is their role to utilize what is accepted to be a far-from-perfect system, to attempt to ensure best possible subject access to as wider social group as possible. This inevitably means that using headings that reflect various popular prejudices, as this

(whether acceptable or not) ensures easier access to the majority of the users that the system was intended for: users of the LC. No system is neutral; therefore any attempt to provide subject access to a social group is inevitably going to isolate other social groups. The problem lies; therefore, in the fact that, unlike social tags, LCSH appears to the layperson to be objective and unbiased; when – like social tags, or any other system used to categorize information, bias is inevitable.

6.3 There is no neutrality

Applying tags, headings, or any other form of categorization to information creates meaning, and therefore bias, whether consciously or not. This brings into the question the validity of the methodology of this study. Although it was necessary in order to glean useful results, the defining of bias and its categorization into ‘pigeonholes’ (no matter how systematically and scholarly this was conducted), seems somewhat contradictory given the subject matter: biased categories being applied to biased headings and tags. However, flawed this may appear, it reinforces the point that the concepts of bias and interpretation are inseparable. Regardless of one’s motivations, training, background knowledge, language, or system, one cannot interpret anything without (whether consciously or not) applying personal, cultural and learned biases based on a particular worldview.

7. Bibliography

Berman, Sanford (1992). Things Are Seldom What They Seem: Finding Multicultural Materials in Library Catalogs. In Berman (ed.) *Alternative Library Literature, 1990/1991*. Phoenix, A.Z.: Oryx Press. 132-136.

Berman, Sanford. (1993). *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Rev. ed., Jefferson, N.C.: McFarland & Co.

Bryman, A. (2012). *Social Research Methods*. 4th ed. Oxford: Oxford University Press.

Clack, Doris H. (1994). Subject Access to African American Studies Resources in Online Catalogs: Issues and Answers. *Cataloging & Classification Quarterly*. 19.2: 49-66.

Feryerand, P. (1975). *Against Method*. London: New Left Review.

Fischer, K S. (2005). Critical views of LCSH, 1990-2001: the third bibliographic essay. *Cataloging & Classification Quarterly*. 41.1: 63-109.

Gorman, G. E., & Clayton, P. (2005). *Qualitative research for the information professional*. 2nd ed. London: Facet Publishing.

Harris, Jessica L. Milstead, and Doris Hargrett Clack. (1979). Treatment of people and peoples in subject analysis. *Library Resources & Technical Services*. 23.Fall: 374-90.

Hecker, M., Muhlbacher, S., Wolff, C. (2007). Tagging tagging: Analysing user keywords in scientific bibliography management systems. *Digital Library of Information Science Technology*. 9.2. Retrieved 11th April 2014 from: <http://journals.tdl.org/jodi/index.php/jodi/article/view/246>.

Jones H. P., Tedd, L., Wheatley, A. (2012). *Module DSM7820: Information Organisation and Retrieval*. 7th ed. Aberystwyth: Open Learning Unit, Department of Information Studies, University of Wales, Aberystwyth.

Kipp, M.E.I (2007). @toread and cool: tagging for time, task and emotion. *Proceedings for the 8th Information Architecture Summit*. Las Vegas: Association for Information Science and Technology.

Kipp, M.E.I., Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Proceedings of the American Society for Information Science and Technology*. 43: 1-18. Medford: Information Today.

Knowlton, S A. (2005). Three decades since Prejudices and Antipathies: a study of changes in the Library of Congress Subject Headings. *Cataloging & Classification Quarterly*. 40.2: 123-145.

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

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*. 6.6: 763-779.

Maggio, Rosalie. (1997). *Talking About People: A Guide to Fair and Accurate Language*. 3rd ed. Westport, CT: Greenwood Press

Mann, T. (2003). Why LC Subject Headings Are More Important Than Ever. *American Libraries*. 34.9: 52-54

Mathes, A. (2004). *Folksonomies—Cooperative classification and communication through shared metadata*. Retrieved 11th April 2014 from: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.

O'Neill, E. T., Chan, L. M. (2003). FAST (Faceted Application of Subject Terminology): A simplified vocabulary based on the Library of Congress Subject Headings. *IFLA Journal*. 29.4: 336-342.

Olson, H A., Schlegl, R. (2001). Standardization, objectivity, and user focus: a meta-analysis of subject access critiques. *Cataloging and Classification Quarterly*. 32.2: 61-80.

Olson, H.A. (2000). Difference, culture and change: the untapped potential of LCSH. *Cataloging and Classification Quarterly*. 29.1/2: 53-71.

Peterson, E. (2006). Beneath the metadata: some philosophical problems with folksonomy. *D-Lib*. 12:11. Retrieved 15th August 2014 from: <http://www.dlib.org/dlib/november06/peterson/11peterson.html>

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

Postigo, H. (2011). *Questioning the Web 2.0 Discourse: Social Roles, Production, Values, and the Case of the Human Rights Portal*. *The Information Society*, 27:181-193.

Rogers, M N. (1993). Are we on equal terms yet? Subject headings concerning women in LCSH, 1975-1991. *Library Resources and Technical Services*. 37.2: 181-196.

Rolla, P J. (2009). User Tags Versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections? *Library Resources & Technical Services*. 53.3: 174-184.

- Rorissa, A. (2008). User-generated descriptions of individual images versus labels of groups and image: A comparison using basic level theory. *Information Processing & Management*. 44.5: 1741-1753.
- Spiteri, L.F. (2006). The Use of Folksonomies in Public Library Catalogues. *Serials Librarian*. 51.2: 75-89.
- Spiteri, L.F., Tarulli, L. (2012). Social Discovery Systems in Public Libraries: If We Build Them, Will They Come? *Library Trends*. 61.1: 132-147.
- Strottman, T.A. (2007). Some of our fifty are missing: Library of Congress Subject Headings for Southwestern cultures and history. *Cataloging & Classification Quarterly*. 45.2: 41-64
- Wilk, David, Shlomo Rotenberg, Sarah Schacham, Gita Hoffman, Shifra Liebman (2001). Problems in the Use of Library of Congress Subject Headings as the Basis for Hebrew Subject Headings in the Bar-Ilan University Library. *International Cataloguing and Bibliographic Control*. 30.3: 43-6.
- Yi, K. (2010). A semantic similarity approach to predicting Library of Congress subject headings for social tags. *American Society for Information Science and Technology*, 61.8: 1658-1672.
- Yi, K., Chan, L. M. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*. 65.6: 872-900.
- Yoon, K. (2012). Conceptual Syntagmatic Associations in User Tagging. *Journal of the American Society for Information Science and Technology*. 63.5: 923-935.

Appendices

Appendix A: Titles used in pilot study

From GoodReads' list of popular multicultural fiction books. Retrieved 4th August, 2014 from: <https://www.goodreads.com/shelf/show/multicultural-fiction>

1. See, Lisa. *Snow Flower and the Secret Fan*
2. Martel, Yann. *Life of Pi*
3. Verghese, Abraham. *Cutting for Stone*
4. Hosseini, Khaled. *A Thousand Splendid Suns*
5. Lahiri, Jhumpa. *Interpreter of Maladies*
6. Esquivel, Laura. *Like Water for Chocolate*
7. Alexie, Sherman. *The Absolutely True Diary of a Part-Time Indian*
8. Kwok, Jean. *Girl in Translation*
9. Lin, Grace. *Where the Mountain Meets the Moon*
10. Ford, Jamie. *Hotel on the Corner of Bitter and Sweet*

Appendix B: Titles used in main study

From GoodReads' list of popular feminist fiction books. Retrieved 4th August, 2014 from: <https://www.goodreads.com/shelf/show/feminist-fiction>

1. Atwood, Margaret. *The Handmaid's Tale*
2. Walker, Alice. *The Color Purple*
3. Plath, Sylvia. *The Bell Jar*
4. Chopin, Kate. *The Awakening*
5. French, Marilyn. *The Women's Room*
6. Atwood, Margaret. *Cat's Eye*
7. Atwood, Margaret. *Alias Grace*
8. Neale Hurston, Zora. *Their Eyes Were Watching God*
9. Morrison, Toni. *Beloved*
10. Perkins Gilman, Charlotte. *The Yellow Wallpaper*
11. Bronte, Charlotte. *Jane Eyre*
12. Flagg, Fannie. *Fried Green Tomatoes at the Whistle Stop Cafe*
13. Stockett, Kathryn. *The Help*
14. Cunningham, Michael. *The Hours*
15. Allison, Dorothy. *Bastard Out of Carolina*
16. Woolf, Virginia. *Orlando*
17. Alcott, Louisa May. *Little Women*
18. Woolf, Virginia. *Mrs. Dalloway*
19. Woolf, Virginia. *To the Lighthouse*
20. Morrison, Toni. *The Bluest Eye*
21. Atwood, Margaret. *The Blind Assassin*
22. Diamant, Anita. *The Red Tent*
23. Winterson, Jeanette. *Written on the Body*
24. Atwood, Margaret. *The Edible Woman*
25. Kingsolver, Barbara. *The Poisonwood Bible*
26. Allison, Dorothy. *Cavedweller*
27. Atwood, Margaret. *The Penelopiad*
28. Winterson, Jeanette. *Oranges Are Not the Only Fruit*

29. Allison, Dorothy. *Trash*
30. Kingsolver, Barbara. *Prodigal Summer*
31. Perkins Gilman, Charlotte. *Herland*
32. Jordan, Hillary. *When She Woke*
33. Monk Kidd, Sue. *The Secret Life of Bees*
34. Walker, Alice. *The Temple Of My Familiar*
35. Oates, Joyce Carol. *Foxfire: Confessions of a Girl Gang*
36. Atwood, Margaret. *Surfacing*
37. Morrison, Toni. *Sula*
38. Brown, Rita Mae. *Rubyfruit Jungle*
39. Butler, Octavia E. *Kindred*
40. Piercy, Marge. *Sex Wars: A Novel of Gilded Age New York*
41. Lee, Harper. *To Kill a Mockingbird*
42. Tepper, Sheri S. *Gibbon's Decline and Fall*
43. Atwood, Margaret. *The Robber Bride*
44. Moon, Elizabeth. *Remnant Population*
45. Walbert, Kate. *A Short History of Women*
46. Hosseini, Khaled. *A Thousand Splendid Suns*
47. Austin, Jane. *Pride and Prejudice*
48. Donoghue, Emma. *The Sealed Letter*
49. Terrell, Heather. *Brigid of Kildare*
50. Morrison, Toni. *Song of Solomon*