



PROJECT MUSE®

Characteristics and Effectiveness of Tags in Public Library  
Online Public Access Catalogues/Les caractéristiques et  
l'efficacité des balises dans les catalogues publics en  
ligne des bibliothèques publiques

Isola Ajiferuke, Jamie Goodfellow, Adeola Opesade

Canadian Journal of Information and Library Science, Volume 39,  
Number 3/4, September-December / septembre-décembre, pp. 258-278  
(Article)

Published by University of Toronto Press



➔ For additional information about this article

<https://muse.jhu.edu/article/598445>

# Characteristics and Effectiveness of Tags in Public Library Online Public Access Catalogues

# Les caractéristiques et l'efficacité des balises dans les catalogues publics en ligne des bibliothèques publiques

Isola Ajiferuke

Faculty of Information and Media Studies, University of Western Ontario  
iajiferu@uwo.ca

Jamie Goodfellow

Sheridan College Library

Jamie.goodfellow@sheridancollege.ca

Adeola Opeade

Africa Regional Centre for Information Science, University of Ibadan  
morecrown@gmail.com

**Abstract:** The purpose of this study was to investigate the characteristics and effectiveness of tags in public library online public access catalogues (OPACs). Three public libraries that have adopted BiblioCommons' OPAC system—Edmonton Public Library, Seattle Public Library, and Christchurch City Libraries—were selected for the study. In the OPAC of each of these libraries, fifty queries were searched using tags as well as keyword and subject as access point. The results of the study showed that a large number of items in public libraries are still not being tagged, while for those items that have been tagged, the tags were mostly made up of one or two words and were subject related. In terms of effectiveness, the precision level of a tag search was found acceptable and somewhat comparable to the precision levels of keyword and subject searches, but of the three access points, tags retrieved the fewest number of items.

**Keywords:** BiblioCommons, Christchurch City Libraries, Edmonton Public Library, online public access catalogues, Seattle Public Library, tagging

**Résumé :** Le but de cette étude était d'étudier les caractéristiques et l'efficacité des balises dans le catalogue public en ligne (OPAC) de bibliothèques publiques. Trois bibliothèques publiques ayant adopté le système d'OPAC de BiblioCommons, la bibliothèque publique d'Edmonton, la bibliothèque publique de Seattle, et les bibliothèques municipales de Christchurch, ont été choisies pour l'étude. Dans l'OPAC de chacune de ces bibliothèques, cinquante requêtes ont été lancées en utilisant aussi bien des balises que des mots-clés et des termes de sujet comme point d'accès. Les résultats de l'étude ont montré qu'un grand nombre d'ouvrages dans les bibliothèques publiques ne sont toujours pas balisés alors que pour les éléments qui ont été balisés, les balises sont majoritairement constituées d'un ou deux mots et ont

un lien avec le sujet. En termes d'efficacité, le niveau de précision d'une recherche par balise a été jugé acceptable et dans une certaine mesure comparable aux niveaux de précision obtenus dans la recherche par mots-clés et par sujet, mais des trois points d'accès, ce sont les balises qui ont extrait le plus petit nombre d'éléments.

Mots-clés : BiblioCommons, Bibliothèques municipales de Christchurch, Bibliothèque publique d'Edmonton, Catalogues publics en ligne, balisage

## Introduction

Tagging is one of the forms of implementation of Web 2.0 technologies in libraries (DeZelar-Tiedman 2011). Tagging, which is a form of free indexing, refers to users assigning tags (or words) of their own choice to documents, blog posts, or webpages that they have created or viewed (Hedden 2008). The main advantages of tagging include the use of people's own vocabulary and the fact that everyone has the opportunity to contribute and share tags (Spiteri 2007).

Popular academic applications of social tagging include its use in bookmarking academic articles on CiteULike (<http://www.citeulike.org/>) or BibSonomy (<http://www.bibsonomy.org/>) (Eckert, Hänger and Niemann 2009) and the personal online cataloging of books on LibraryThing (<https://www.librarything.com/>), Shelfari (<http://www.shelfari.com/>), BookBump (<http://www.bookbump.com/>), GoodReads (<http://www.goodreads.com/>), and BookJetty (<http://www.bookjetty.com/>). The relative popularity of LibraryThing has aided the arguments of some members of the library community that the use of tags in online public access catalogues (OPACs) would be beneficial (Spiteri 2006; Rethlefsen 2007; Rolla 2009; Thomas, Caudle and Schmitz 2009). Although a few libraries, such as Pennsylvania State University library, have individually incorporated the use of tags in their OPACS, the first major initiative for public libraries was launched by BiblioCommons.

BiblioCommons, a project of Knowledge Ontario and funded by the province of Ontario, Canada, developed an OPAC system called BiblioCore that includes tags as an access point. It was launched at Oakville Public Library in July 2008, and it has since been adopted by forty-one other public libraries in Canada (BiblioCommons 2014a), thirty-one public libraries in the United States (BiblioCommons 2014b), and three in Australia and New Zealand (BiblioCommons 2014c). BiblioCore not only allows users in a particular library to add tags to any of the items in that library, but it shares the added tags with other libraries in the consortium that have the same item in their collections.

Despite the wide adoption and application of tagging in many areas, it still has the same problems, such as ambiguities and lack of control of synonyms that is usually associated with the use of uncontrolled vocabularies (Spiteri 2007). This assertion has been corroborated by empirical studies (Ding et al. 2009; Spiteri 2009) looking at the various forms of tags that were assigned by users of some tagging systems. In addition, some other studies have found that users also often assign non-subject tags (Golder and Huberman 2006; Lawson 2009; Thomas, Caudle and Schmitz 2009; Lu, Park and Hu 2010; Kipp 2011a, 2011b).

So how does the assignment of varying forms of tags and non-subject tags affect the retrieval performance of online public access catalogue systems?

One earlier attempt to answer this question is the pilot study by Isola Ajiferuke and Jamie Goodfellow (2012). The current study extends the pilot study by collecting data from multiple libraries, making use of more contemporary queries, and expanding the number of queries searched. In addition, we will also investigate the characteristics of the tags that users have assigned. Specifically, the research questions that the current study will try to address are:

- How active are users in tagging the items in the public library collections?
- What kinds of tags are users assigning to the items in the public library collections?
- How effective is the use of tags in searching the public library collections?
- How well does a tag compare to a subject heading or a keyword as an access point in public library OPACs?

### **Literature Review**

Most of the studies investigating the use of tags have either examined the characteristics of the tags used or compared the tags with keywords or controlled vocabularies. In the first category, Louise Spiteri (2007) examined the structure and forms of tags used in Del.icio.us, Furl (another social bookmarking site), and Technorati. The author found that single-word terms constituted 93 percent of Del.icio.us tags, 76 percent of Furl tags, and 80 percent of Technorati tags; nouns accounted for 95 percent of Del.icio.us tags, 94 percent of Furl tags, and 97 percent of Technorati tags; and most of the tags represented things (76 percent in Del.icio.us, 82 percent in Furl, and 90 percent in Technorati), with activities forming a distant second (12 percent in Del.icio.us, 10 percent in Furl, and 4 percent in Technorati). Kerstin Bischoff et al. (2008) also analyzed the tags used in Del.icio.us along with those in Flickr and Last.fm. They found that more than 50 percent of the tags in Del.icio.us were topic related, while most of the tags in Las.fm corresponded to music genres with opinion/quality and author/owner being the second and third most used types of tags attached to music resources.

Another study that compared the characteristics of tags in Del.icio.us with those from other social tagging services was by Ying Ding et al. (2009). In their study, they compared tags in Del.icio.us with those from Flickr and YouTube over a three-year period—that is, from 2005 to 2007. The authors found social tagging activities to have increased dramatically between 2005 and 2007 in all three services, but while the use of topical tags dominated in Del.icio.us and YouTube, Flickr taggers used dates, locations, colours, and seasons to tag their photographs. Two other studies in this category examined only tags from one social tagging service each. Hao-Ren Ke and Ya-Ning Chen (2012) investigated CiteUlike, while Henk Voorbij (2012) examined LibraryThing. Using a data set of 4,215 tags attributed to 1,600 scholarly articles from fifteen library and information science journals in CiteUlike, Ke and Chen found that topic-related

tags accounted for 45.2 percent of the tags and title-related tags accounted for 43.77 percent, while content-related tags accounted for 6.53 percent. Voorbij took a random sample of 600 records from the catalogue of an academic library that had adopted LibraryThing and examined these records to determine whether they carried tags, while a random of 160 records with tags was taken to determine the nature of the tags. It was found that about one-third of the records had tags, 80 percent of the tags were subject terms, and 50 percent of the subject tags were covered by a keyword in the record.

In the second category of studies, Margaret Kipp (2011a, 2011b) has compared user tags, author keywords, and professional indexer-assigned descriptors used in articles indexed by CiteUlike for two different subject areas. The first study used library and information science articles for the comparisons, while the second study made use of biomedical articles. In each study, comparisons of the three types of indexing terms were based on seven categories: same, synonym, broader term, narrower term, related term, related, and not related. While the keywords/descriptors matches were the most common in both studies, the tags/keywords matches were more than the tags/descriptors matches for library and information science articles, while the reverse was the case for biomedical articles. The three most common types of relationships between the matched terms were same, related term, and related. The unrelated terms were found to be about time and task management, geographic or personal, specific details and qualifiers, generalities, emergent vocabulary, and other. Similarly, Daniel Lee and Titus Schleyer (2012) compared CiteUlike tags to medical subject headings (MeSH) terms for 231,388 biomedical papers indexed in MEDLINE. The authors found that, on the average, papers were annotated with 4.7 tags versus 12.2 MeSH terms. However, there was a low degree of overlap between tags and MeSH terms assigned per paper.

Other studies compared tags assigned to books with the Library of Congress subject headings (LCSH) for the same books. Karen Lawson (2009) looked at the number of subject headings as well as the tags assigned by users of LibraryThing and Amazon to 155 books selected from WorldCat. While the average of three subject headings were assigned to these books, thirty-four tags and twenty-nine tags were assigned on average by Amazon and LibraryThing users respectively. In total, 57 percent of the Amazon tags were found to be objective (that is, topical), while only 43 percent of the LibraryThing tags were objective. Paul Heymann and Hector Garcia-Molina (2009) took a sample of 309,071 LibraryThing works, and compared their assigned tags to the subject headings of the same works found in the Library of Congress MARCH records from the Internet Archive. The study analyzed only works found in both LibraryThing and the Library of Congress and only unique subject headings and tags that have been used to index at least ten works. Out of the 8,783 unique LCSH terms and 47,957 unique tags examined, the study found that 3,408 LCSH terms were exactly equivalent to a tag, while an additional 838 were almost exactly equivalent to a tag. However, the study also found that the sets of works annotated by corresponding LCSH terms and tags rarely intersected to a significant extent.

Caimei Lu, Jung-ran Park and Xiaohua Hu (2010), applying a similar methodology as Heymann and Garcia-Molina to collect their data, examined 8,562 books that were both included in LibraryThing and Library of Congress bibliographic records from the Internet archive and had at least twenty unique tags assigned to each of them. The results obtained were similar to Heymann and Garcia-Molina's and found that about 3,824 unique terms overlapped out of the 176,105 unique tags and 7,628 unique subject headings analyzed. The overlapping terms represented about 2.2 percent of the unique tags and about 50.1 percent of the unique subject headings. In addition, the study found that 7,276 books (85 percent) had at least one of its LCSH terms used by the users to tag the same books, while only 407 books (4.75 percent) had more than half of their LCSH terms used as tags.

In his own study, Peter Rolla (2009) compared the tags assigned to a sample of forty-five books in LibraryThing with the LCSH terms assigned to the same books in WorldCat. The study found that the subject headings and the user tags assigned to thirty-five books (75.6 percent) represented the same subject or concept, though often expressed in different terms. At the same time, it was found that for every book in the sample, the user tags contained subject terms or concepts that the subject headings did not express, while for twenty-five of the books (55.6 percent) subject headings brought out concepts or topics that the user tags did not.

In another study by Christine DeZelar-Tiedman (2011), the LCSH headings assigned to works by twentieth- and twenty-first-century English and American literary authors found in the University of Minnesota online catalogue were compared with tags assigned to the same works by LibraryThing users. For works having both LCSH headings and tags, 18.7 percent of the tags had exact or partial LCSH matches, while the remaining 81.3 percent of the tags had no LCSH match. The study also found that for records having tags but no LCSH headings attached, the tags were too broad to be useful for searching in a large academic library collection.

Finally, there was a study in this second category that compared tags assigned to non-bibliographic items by users with keywords assigned by professionals. Catherine Hall and Michael Zarro (2011) examined the free-text keywords attached to 720 records from the History subject collection of the ipl2 digital library with tags assigned to the same records in Del.icio.us. It was found that for 204 records (33 percent) there was a match between at least one tag and one keyword.

While many authors have speculated on the usefulness of tags in searching, only a couple of studies have actually empirically investigated their effectiveness in searching. Jason Morrison (2008) compared the search information retrieval performance of tags from three social bookmarking sites (Del.icio.us, Furl, and Reddit) against three search engines (Google, Microsoft Live, and AltaVista) and two directories (Yahoo and Open Directory Project). Thirty-four participants were asked to create three queries each, and for each query, the participant examined up to twenty results from each information retrieval system. Using a

cut-off at twenty for precision and the pooled method to obtain recall, the social bookmarking sites were found to have fared surprisingly well, though the search engines had the highest precision and recall, while the directories were more precise than the social bookmarking sites. In their own study, Kun Lu and Margaret Kipp (2014) compared the retrieval effectiveness of tags with that of author keywords. Using a test collection of 17,264 biomedical articles with tags assigned in CiteULike and author keywords available from PubMed Central, the authors found tags to be comparable to author keywords in terms of average precision but inferior in terms of recall. It should be noted that while Morrison's study examined the effectiveness of searching the web using tags and the Lu and Kipp's study examined the effectiveness of tags in searching for journal articles, our own study looks at the effectiveness of using tags to search primarily for books. Our study will therefore be complementing the two earlier studies.

### Methodology

For this study, we selected three libraries for the evaluation of the effectiveness of user tags. While seventy-six public libraries have adopted BiblioCore as their online public catalogue system, searching a large number of libraries has become obsolete as the OPACs of these libraries have the same interface and the same retrieval mechanism. In addition, the cooperative manner in which tags are shared among participating libraries in BiblioCore is such that tags assigned to a particular item in one library are the same for that item in other libraries. For example, the book entitled *The Devil in the White City* by Erik Larson had been assigned fifteen tags at the time of data collection for this study, and these fifteen tags were the same for over thirty libraries with the same book in their collections. Hence, using one public library with a large collection might be sufficient for the study, but we decided to select one from each of the three countries/continents that have adopted BiblioCore to account for country/continental differences in library collections. Edmonton Public Library, which has a very large collection, was selected from Canada, Seattle Public Library was selected from United States due to its large collection (it was also one of the first public libraries to adopt BiblioCore in the United States), and Christchurch City Libraries was selected from Australia/New Zealand again due to its large collection size.

To formulate queries for the study, we obtained sample reference questions submitted to the London Public Library and Toronto Public Library. We selected those questions that we felt could be satisfied by searching the OPAC and supplemented these with queries formulated in an information retrieval class at the University of Western Ontario to obtain a total of fifty queries, which is the minimum number recommended by Paul Clough and Mark Sanderson (2013) and used in the study by Lu and Kipp (2014). Two of the authors conducted the search for the study, and to ensure inter-searcher consistency, the two searchers worked together to break down each query into search syntax, format, and audience (see Appendix 1). For each of the fifty queries, a search was conducted by one of the searchers in each of the three public library OPACs that had been selected. Searches were conducted by access point (keyword, subject, and tag)

one at a time. We did not include author and title access points because they are mainly useful for known item searches. For each query, the retrieved set of items was narrowed down by format (that is, book, magazine or journal, DVD, music CD, audiobook CD, eBook, and so on) and/or audience (that is, children, teen, or adult), where applicable. For example, for the query “find books for adults about drug abuse,” “drug abuse” was used as the search term for each access point, but the retrieved set was then narrowed down by format to books and by audience to adults. The OPACs allow the results to be sorted by relevance, by the date acquired, by title, by author, and by published date, with sorting by relevance as the default. We made use of this default sorting for the final set and then examined the first thirty items for relevance. The number of items examined was limited to thirty since previous studies have shown that most users rarely view more than the top thirty documents retrieved in response to a query (Spink et al. 2001).

The two most popular effectiveness measures are precision and recall. However, the databases of the OPACs were too large for the determination of the number of relevant items in the database for each query, which is required for the calculation of recall. Therefore, precision was used as the sole effectiveness measure, and it was calculated as (number of relevant items examined)/(number of items examined). To estimate the proportion of items in the OPACs that have been tagged as well as to examine the characteristics of the tags that have been assigned to items, we selected the query that retrieved the highest number of items when searched as a keyword for each library but made sure that we did not select the same query for all of the libraries. For each query, we used the advanced search interface to search for items using the “Keyword Anywhere” field. It should be noted that while searching by keyword looks for the presence of a query term in an item’s full record (the full record in these OPACs includes the item’s title, alternative title, publisher, contents, language, statement of responsibility, and so on), searching by “Keyword Anywhere” looks for the presence of the query term in all access points, including subject and tag. This was done to increase the pool of items to examine.

The items retrieved were then sorted by publication date, and systematic random sampling was used to select 5 percent of the items retrieved. For each item retrieved, the title, format of the item, year of publication, and number of tags assigned were noted. For each item that had been tagged, the number of words for each tag was noted and the tag categorized into one of the seven Golder and Huberman (2006) categories (that is, “identifying what (or who) it is about,” “identifying what it is,” “identifying who owns it,” “refining categories,” “identifying qualities or characteristics,” “self reference,” and “task organizing”).

## Results

### *Characteristics of Tags*

The following queries were selected to examine the characteristics of tags: “parenting” was selected for Edmonton Public Library, “historical fiction” was



Table 1: Frequency distribution of number of tags per item

Number of tags	Parenting (Edmonton) Number of items	Historical fiction (Seattle) Number of items	Time travel (Christchurch) Number of items
0	269 (77.75%)	230 (64.43%)	155 (87.08%)
1	32 (9.25%)	49 (13.73%)	5 (2.81%)
2	10 (2.89%)	15 (4.20%)	5 (2.81%)
3	8 (2.31%)	15 (4.20%)	3 (1.69%)
4	3 (0.87%)	10 (2.80%)	3 (1.69%)
5	5 (1.45%)	38 (10.64%)	7 (3.93%)
6	2 (0.58%)	0 (0.00%)	0 (0.00%)
7	1 (0.29%)	0 (0.00%)	0 (0.00%)
8	2 (0.58%)	0 (0.00%)	0 (0.00%)
9	0 (0.00%)	0 (0.00%)	0 (0.00%)
10	1 (0.29%)	0 (0.00%)	0 (0.00%)
11	2 (0.58%)	0 (0.00%)	0 (0.00%)
12	2 (0.58%)	0 (0.00%)	0 (0.00%)
13	1 (0.29%)	0 (0.00%)	0 (0.00%)
14	0 (0.00%)	0 (0.00%)	0 (0.00%)
15	8 (2.31%)	0 (0.00%)	0 (0.00%)
Total	346	357	178
Mean number of tags	0.97	0.99	0.40
St. Dev.	2.87	1.17	1.18
Median number of tags	0.00	0.00	0.00

selected for Seattle Public Library, and “time travel” was selected for Christchurch City Libraries. The number of items examined after using systematic sampling to select 5 percent of the items retrieved searching the query using the “Keyword Anywhere” field are 346, 357, and 178 for Edmonton Public Library, Seattle Public Library, and Christchurch City Libraries respectively. Overall, the minimum number of tags per item was zero and the maximum was fifteen (see table 1). In fact, for each library, the percentage of items without a tag was at least 60 percent. In addition, the mean number of tags per item was less than one in each library.

The characteristics of the tags were then examined for any item that had a tag. For each library, the most common number of words per tag was one, and at least 70 percent of the tags had one or two words (see table 2). This finding implies that most users are assigning simple tags to the items. The largest number of words for a tag was six. When categorizing tags using the Golder and Huberman classification categories, we noted that the most used category was “identifying what (or who) it is about,” while the least used categories were “identifying who owns it” or “self-reference” (see table 3). There were also some instances of the tag being assigned for task organization and a few instances where we could not classify the tag into any of the seven categories, which occurred when the tag was just a letter (for example, “d”) or when we did not understand the word assigned (for example, ‘amigurumi’).

Table 2: Frequency distribution of number of words per tag

Number of words	Parenting (Edmonton) Number of tags	Historical fiction (Seattle) Number of tags	Time travel (Christchurch) Number of tags
1	203 (60.24%)	148 (41.93%)	39 (54.93%)
2	92 (27.30%)	119 (33.71%)	25 (35.21%)
3	26 (7.72%)	39 (11.05%)	3 (4.23%)
4	11 (3.26%)	28 (7.93%)	2 (2.82%)
5	3 (0.89%)	18 (5.10%)	1 (1.41%)
6	2 (0.59%)	1 (0.28%)	1 (1.41%)
Total	337	353	71
Mean number of words	1.59	2.01	1.65
Median number of words	1	2	1

Table 3: Classification of tags into Golder and Huberman categories

Category	Parenting (Edmonton) Number of tags	Historical fiction (Seattle) Number of tags	Time travel (Christchurch) Number of tags
Identifying what (or who) it is about	173 (51.34%)	242 (68.56%)	51 (71.83%)
Identifying what it is	26 (7.72%)	12 (3.40%)	7 (9.86%)
Identifying who owns it	0 (0.00%)	2 (0.57%)	0 (0.00%)
Refining categories	66 (19.58%)	15 (4.25%)	0 (0.00%)
Identifying qualities or characteristics	42 (12.46%)	40 (11.33%)	11 (15.49%)
Self-reference	10 (2.97%)	0 (0.00%)	0 (0.00%)
Task organizing	15 (4.45%)	41 (11.61%)	2 (2.82%)
Not sure	5 (1.48%)	1 (0.28%)	0 (0.00%)
Total	337	353	71

### *Search Effectiveness*

For the Edmonton Public Library, four of the queries did not retrieve any documents when the keyword was used as the access point, eight yielded nothing when the subject was used as the access point, and seventeen queries did not retrieve any documents when the tag was used as the access point (see table 4). The maximum number of documents retrieved for any query was 4,912 for the keyword, 2,994 for the subject, and 724 for the tag. The medians for the number of items retrieved were 103, 54, and 2.5 for keyword, subject, and tag respectively (the distribution of the number of items retrieved was skewed for each of the access points, hence, the most appropriate measure of the central tendency is the median.) These medians were found to be significantly different when we performed a Friedman test, which gave a chi-square value of 79.307 with a degree of freedom of two and a p-value of .000.

For Seattle Public Library, four of the queries did not retrieve any documents when the keyword was used as the access point, five yielded nothing with the subject as the access point, and eighteen queries did not retrieve any documents with the tag as the access point (see table 5). The maximum number of documents retrieved for any query was 11,094 for the keyword, 9,978 for

Table 4: Descriptive statistics of the number of items retrieved for the Edmonton Public Library

Access point	Number of queries that yielded no results	Maximum number of items retrieved	Mean number of items retrieved	Median number of items retrieved
Keyword	4	4,912	401.56	103
Subject	8	2,994	263.74	54
Tag	17	724	33.02	2.5

Table 5: Descriptive statistics of the number of items retrieved for the Seattle Public Library

Access point	Number of queries that yielded no results	Maximum number of items retrieved	Mean number of items retrieved	Median number of items retrieved
Keyword	4	11,094	920.2	107
Subject	5	9,978	609.24	59
Tag	18	1,350	52.82	3

Table 6: Descriptive statistics of the number of items retrieved for the Christchurch City Libraries

Access point	Number of queries that yielded no results	Maximum number of items retrieved	Mean number of items retrieved	Median number of items retrieved
Keyword	5	5,006	437.44	55
Subject	9	3,345	265.06	41
Tag	22	749	26.94	1.5

subject and 1,350 for tag. The medians for the number of items retrieved were 107, 59, and 3 for keyword, subject, and tag respectively. These medians were found to be significantly different when we performed a Friedman test, which gave a chi-square value of 89.585 with a degree of freedom of two and a p-value of .000.

Finally, for Christchurch City Libraries, five of the queries did not retrieve any documents when the keyword was used as the access point, nine yielded nothing with the subject as the access point, and twenty-two queries did not retrieve any documents with the tag as the access point (see table 6). The maximum number of documents retrieved for any query was 5,006 for a keyword, 3,345 for a subject, and 749 for a tag. The medians for the number of items retrieved were 55, 41, and 1.5 for keyword, subject, and tag respectively. These medians were also found to be significantly different when we performed a Friedman test, which gave a chi-square value of 84.182 with a degree of freedom of two and a p-value of .000.

In the case of the precision ratio, we could only obtain values for queries for which at least one item was retrieved. For all of the libraries, the distribution of the precision ratio was skewed for at least one of the access points. Hence, the most appropriate measure of central tendency is the median. For the Edmonton

Table 7: Descriptive statistics of the precision ratio for the Edmonton Public Library

Access point	Minimum	Maximum	Mean	Median
Keyword	0.0	1.0	0.6263	0.6833
Subject	0.0	1.0	0.6513	0.7238
Tag	0.0	1.0	0.5956	0.6667

Table 8: Descriptive statistics of the precision ratio for the Seattle Public Library

Access point	Minimum	Maximum	Mean	Median
Keyword	0.3158	1.0	0.8349	0.913
Subject	0.0	1.0	0.7308	0.8333
Tag	0.0	1.0	0.5482	0.6015

Table 9: Descriptive statistics of the precision ratio for the Christchurch City Libraries

Access point	Minimum	Maximum	Mean	Median
Keyword	0.0	1.0	0.8057	0.8667
Subject	0.0	1.0	0.7675	0.8667
Tag	0.0	1.0	0.5112	0.4667

Public Library, the median precision ratios were .6833, .7238, and .6667 for keyword, subject, and tag respectively (see table 7). A Kruskal–Wallis test showed that these medians were not significantly different (chi-square value = .521 with a degree of freedom of two and a p-value of .771). For the Seattle Public Library, the median precision ratios were .913, .8333, and .6015 for keyword, subject, and tag respectively (see table 8). A Kruskal–Wallis test showed that the medians for keyword and subject were significantly higher than they were for the tag (chi-square value = 7.127 with a degree of freedom of two and a p-value of .028). Finally, for the Christchurch City Libraries, the median precision ratios were .8667, .8667, and .4667 for keyword, subject, and tag respectively (see table 9). A Kruskal–Wallis test showed that the medians for keyword and subject were significantly higher than they were for the tag (chi-square value = 8.982 with a degree of freedom of 2, and a p-value of .011).

## Discussion

A large percentage of the items in these public libraries have not been tagged. Although the cooperative manner in which tags are shared among the libraries is meant to increase the likelihood of an item that is common to two or more libraries being tagged, many items found in multiple libraries still remain untagged—for example, the 2006 book *The Castle in the Forest* by Norman Mailer could be found in more than twenty of the libraries, yet it had not been tagged. Of course, if an item has not been tagged, then it would be difficult to find it via

the tag access point. Hence, the libraries might want to provide some sort of incentive to patrons to encourage them to tag library items.

The year of publication of an item (which may or may not be the same as the year of acquisition by a library) does not seem to influence the tagging of the item as the correlation between the number of tags and the year of publication for any of the libraries was found to be below 0.2. Also, while the formats of items in the libraries include book, ebook, audiobook CD, audiobook cassette, DVD, downloadable audiobook, and so on, the prevailing dominance of the book format made it unfeasible to determine whether certain formats get tagged more often than others. It is hoped that as the libraries acquire more items in non-book formats, the comparison may be possible. However, it was surprising to note that different formats of the same item often attracted a different number of tags. For example, at the time of data collection for this study, the number of tags for the book *The Devil in the White City* by Erik Larson was fifteen (true crime, Chicago, historical true crime, history, World's fair, architects, architecture, assassination, bpl non-fiction, city building, Columbian fair, dark, historical, landscape architecture, and murder), the ebook format had ten tags (architects, architecture, assassination, Chicago, city building, history, murder, serial killers, true crime, and World's fair), the audiobook CD had seven tags (architecture, Chicago, murder, nonfiction, psychopath, true crime, and World's fair), and the downloadable book format had three tags (historical true fiction, mp3, and true crime). Unless the tag is depicting the item format, one would have expected the same tags to be applicable across different formats of the item—BiblioCommons might want to adopt the synchronization of tags across formats for a particular item just as it synchronizes tags across libraries for an item.

For the items that were tagged, one to two tags per item seem to be very common. This implies that the tagging exercise, as performed by the patrons of these libraries, is not exhaustive, and this might limit the findability of the tagged items via the tag access point. Also, the patrons tend to assign simple tags, with most tags made up of one or two words, but there still tends to be a lot of variation in these words—some are acronyms, abbreviations, slang, one-letter words, very long words, variant spellings of a word, variant word forms, and variant languages (see table 10). For compound words, various styles are used to join words together, including ampersands, slashes, hyphens, no space, conjunctions, prepositions, and so on. By and large, the users' practice in tagging does not conform to subject indexing, which favours nouns over verbs and limits the use of conjunction/preposition in index terms. While most of the tags are subject related, only a few of them are affective or task related.

The cooperative manner of sharing tags among the participating libraries was evident in some of the tags. Even though we used the Seattle Public Library, the Edmonton Public Library and the Christchurch City Libraries for this study, it was not uncommon to find tags that are specific to other libraries—for example, "bostonpl author series" or "nypl book discussion." The sharing of tags among libraries increases the chance of a book being tagged, but it is not certain how useful it is for a tag specific to a library to be assigned to a book in another

Table 10: Tag variations

Tag type	Tag examples
Alphabets, acronyms, abbreviations, and slangs	Audience-ms, bdms, d, d&d, e, exemplar gr, f, glbtq, gvpl 150, sci fi, tech, u.s.a, wcls2012, wwii, ya fiction
Compound words	1 in the time warp trio a gift of barbara o'neil and jean steffano african-american fiction book-within-a-book death and dying guyspreadbookclub mother/son relationships news for business issue #4 parents and kids
Variant spellings, word forms, and languages	1960/1960s/1960s/60s 5th grade/fifth grade architects/architecture chick lit/chicklit first person fiction/first-person fiction homeschooling/home schooling la belle epoque mashup/mash-up non-fiction/nonfiction soins d'espirit wireless/wirelss
Non-subject: affective, task	1001 movies you must . . . 30 books to read before 30 awesome starts here cool gotta read this 2012 lucky day
Non-subject: promotional	April display ideas best chapter books of 2011 best children's books 2012 best of the best 2014 bostonpl author series spring 2013 bpl staff picks bpl suspense listener book review nypl book discussion nypl books to remember opl forthcoming april 2013 opl hot titles may 2013 press book review 11/05/13

library. In addition, some of the tags are promotional in nature—for example, “nypl books to remember,” “opl hot titles January 2013.” Are the users really the ones assigning these promotional tags, or is it being done by the librarians? If it is being done by librarians, then it negates the purpose of tagging, which is meant to be indexing by the users. As a follow-up to this study, it would be interesting to interview some librarians in these libraries to ascertain who is adding these promotional tags and to find out in general the perceptions of the librarians to tagging and its impact on the way they perform their tasks.

When it comes to the effectiveness of the tag as an access point, the precision ratios obtained are good (median values of 0.6667, 0.6015, and 0.4667 for the Edmonton Public Library, the Seattle Public Library, and the Christchurch City Libraries respectively) and somewhat comparable to the effectiveness of a keyword and a subject. This might be due to the fact that most of the tags assigned by users were subject related. We should note here a limitation to our study in the way that relevance assessment was done. We could not use the actual patrons that submitted the queries to make the relevance judgment of the items retrieved because we did not have access to their identities, but the authors based the relevance judgment on an item's details (such as description, excerpts, and reviews), title, notes, and community activities (such as comments and summaries). In terms of the number of items retrieved per query, the tag retrieved far fewer numbers than either the keyword or the subject. In fact, for the seven queries where the users specified a minimum number of relevant items required, while the keyword search or the subject search was able to meet the minimum requirement at least six times for each library, the tag search was able to meet the requirement only three times for each library (see tables 11–13). The fewer number of items retrieved by a tag might be due to the fact that many items in the libraries have still not been tagged, and for those items tagged, the tagging was not exhaustive.

Table 11: Number of relevant items retrieved in response to queries with minimum requirement in the Edmonton Public Library's OPAC

Query number	Minimum number of items required	Number of relevant items retrieved by keyword search	Number of relevant items retrieved by subject search	Number of relevant items retrieved by tag search
3	3–5	5	3	0
6	3	22	27	19
19	5	7	4	1
33	3	21	29	16
35	3	13	11	0
36	3–4	1	1	0
47	5	30	27	12

Table 12: Number of relevant items retrieved in response to queries with minimum requirement in the Seattle Public Library's OPAC

Query number	Minimum number of items required	Number of relevant items retrieved by keyword search	Number of relevant items retrieved by subject search	Number of relevant items retrieved by tag search
3	3–5	7	0	0
6	3	26	29	28
19	5	20	14	1
33	3	29	29	17
35	3	15	10	0
36	3–4	13	8	0
47	5	30	26	15

Table 13: Number of relevant items retrieved in response to queries with minimum requirement in the Christchurch City Libraries' OPAC

Query number	Minimum number of items required	Number of relevant items retrieved by keyword search	Number of relevant items retrieved by subject search	Number of relevant items retrieved by tag search
3	3–5	4	3	0
6	3	30	30	30
19	5	18	18	1
33	3	30	30	24
35	3	10	6	0
36	3–4	3	2	0
47	5	26	25	8

Our findings corroborate the earlier results in the pilot study by Isola Ajiferuke and Jamie Goodfellow (2012) as well as the results of Kun Lu and Margaret Kipp's (2014) study, which found the retrieval performance of tags to be comparable to that of author keywords in searching for journal articles in terms of precision but inferior to the performance of author keywords in terms of recall. Thus, until the level of tagging by users improves considerably, tagging may not be able to serve as a substitute for subject indexing, but tags may serve as complement to subject terms. In that case, BiblioCommons might want to add a combined tag and subject to the list of access points available in the OPAC designed for public libraries. This recommendation is supported by the findings of Lu and Kipp (2014), which suggest that including tags and author keywords in an index for searching journal articles could enhance recall. However, they also noted that doing so might worsen precision.

### Summary and Conclusions

Our study investigated the characteristics and search effectiveness of tagging in public library OPACs. The focus of the study was on public libraries that have adopted the OPAC system designed by BiblioCommons, and Edmonton Public Library was selected in Canada, Seattle Public Library in the United States, and Christchurch City Libraries in Australia/ New Zealand. Queries used for searching the OPACs by two of the authors were adapted mostly from sample reference questions submitted to the London Public Library and the Toronto Public Library.

The results showed that a large percentage of items in the three libraries have not been tagged. For the items that have been tagged, most tags are simple (that is, made up of mostly one or two words) and are subject related. However, we also noticed that some of the tags were for promotional purposes, and we wondered whether such tags were really being assigned by users or by the librarians. In addition, we noticed that in some cases different formats of the same item had different tags. In terms of effectiveness, the precision level of a tag search was found to be acceptable and somewhat comparable to those of keyword and subject searches, but the number of items retrieved by a tag search was far less than what was retrieved by either a keyword or a subject search.



The study recommends to BiblioCommons as well as those in the process of designing a similar OPAC system as BiblioCore, that it should: (1) synchronize tagging across formats so that different formats of the same item could have the same tags and (2) add a combined subject/tag access point to its list of access points for those who might want to use combined human indexing entries for searching. Such a combined access point is likely to result in a higher number of items retrieved than by either a subject search or a tag search but probably fewer than the number retrieved by a keyword search.

## References

- Ajiferuke, Isola, and Jamie L. Goodfellow. 2012. "Evaluation of the Effectiveness of Tag As an Access Point in a Public Library OPAC." Paper presented at the 2012 Annual Conference of the Canadian Association for Information Science, Waterloo, ON, 31 May–2 June. [http://www.caais-acsi.ca/proceedings/2012/cais2012\\_%20ajiferuke\\_goodfellow1.pdf](http://www.caais-acsi.ca/proceedings/2012/cais2012_%20ajiferuke_goodfellow1.pdf) (accessed 1 July 2014).
- BiblioCommons. 2014a. "Canada." <http://www.bibliocommons.com/about/participating-libraries/canada/> (accessed 1 July 2014).
- . 2014b. "United States." <http://www.bibliocommons.com/about/participating-libraries/united-states/> (accessed 1 July 2014).
- . 2014c. "Australia and New Zealand." <http://www.bibliocommons.com/about/participating-libraries/australia-new-zealand/> (accessed 1 July 2014).
- Bischoff, Kerstin, Claudiu S. Firan, Wolfgang Nejdl, and Raluca Paiu. 2008. "Can All Tags Be Used for Search?" Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, 26–30 October. <http://dx.doi.org/10.1145/1458082.1458112>.
- Clough, Paul, and Mark Sanderson. 2013. "Evaluating the Performance of Information Retrieval Systems Using Test Collections." *Information Research* 18 (2): 582. <http://www.informationr.net/ir/18-2/paper582.html> (accessed 1 July 2014).
- DeZelar-Tiedman, Christine. 2011. "Exploring User-Contributed Metadata's Potential to Enhance Access to Literary Works: Social Tagging in Academic Library Catalogs." *Library Review and Technical Services* 55 (4): 221–33. <http://dx.doi.org/10.5860/Lrts.55n4.221>.
- Ding, Ying, Elin K. Jacob, Zhixiong Zhang, Schubert Foo, Erija Yan, Nicolas L. George, and Lijiang Guo. 2009. "Perspectives on Social Tagging." *Journal of the American Society for Information Science and Technology* 60 (12): 2388–401. <http://dx.doi.org/10.1002/asi.21190>.
- Eckert, Kai, Christian Hänger, and Christof Niemann. 2009. "Tagging and Automation: Challenges and Opportunities for Academic Libraries." *Library Hi Tech* 27 (4): 557–69. <http://dx.doi.org/10.1108/07378830911007664>.
- Golder, Scott, and Bernardo A. Huberman. 2006. "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science* 32 (2): 198–208. <http://dx.doi.org/10.1177/0165551506062337>.
- Hall, Catherine E., and Michael A. Zarro. 2011. "What Do You Call It? A Comparison of Library-Created and User-Created Tags." Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Ottawa, Canada, 13–17 June.
- Hedden, Heather. 2008. "How Semantic Tagging Increases Findability." *EContent* 31 (8): 38–43.
- Heymann, Paul, and Hector Garcia-Molina. 2009. "Contrasting Controlled Vocabulary and Tagging: Do Experts Choose the Right Names to Label the Wrong Things?" Proceedings of the Second International Conference on Web Search and Web

- Data Mining, Barcelona, Spain, 9–13 February. [http://www.wsdm2009.org/heymann\\_2009\\_tagging.pdf](http://www.wsdm2009.org/heymann_2009_tagging.pdf) (accessed 1 July 2014).
- Ke, Hao-Ren, and Ya-Ning Chen. 2012. "Structure and Pattern of Social Tags for Keyword Selection Behaviors." *Scientometrics* 92 (1): 43–62. <http://dx.doi.org/10.1007/s11192-012-0718-5>.
- Kipp, Margaret E.I. 2011a. "Tagging of Biomedical Articles on CiteULike: A Comparison of User, Author and Professional Indexing." *Knowledge Organization* 38 (3): 245–61.
- . 2011b. "User, Author and Professional Indexing in Context: An Exploration of Tagging Practices on CiteULike." *Canadian Journal of Information and Library Science* 35 (1): 17–48. <http://dx.doi.org/10.1353/ils.2011.0008>.
- Lawson, Karen G. 2009. "Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers." *Journal of Academic Librarianship* 35 (6): 574–82. <http://dx.doi.org/10.1016/j.acalib.2009.08.020>.
- Lee, Daniel H., and Titus Schleyer. 2012. "Social Tagging Is No Substitute for Controlled Indexing: A Comparison of Medical Subject Headings and CiteULike Tags Assigned to 231,388 papers." *Journal of the American Society for Information Science and Technology* 63 (9): 1747–57. <http://dx.doi.org/10.1002/asi.22653>.
- Lu, Caimei, Jung-ran Park, and Xiaohua Hu. 2010. "User Tags Versus Expert-Assigned Subject Terms: A Comparison of LibraryThing Tags and Library of Congress Subject Headings." *Journal of Information Science* 36 (6): 763–79. <http://dx.doi.org/10.1177/0165551510386173>.
- Lu, Kun, and Margaret E.I. Kipp. 2014. "Understanding the Retrieval Effectiveness of Collaborative Tags and Author Keywords in Different Retrieval Environments: An Experimental Study on Medical Collections." *Journal of the Association for Information Science and Technology* 65 (3): 483–500. <http://dx.doi.org/10.1002/asi.22985>.
- Morrison, Jason P. 2008. "Tagging and Searching: Search Retrieval Effectiveness of folksonomies on the World Wide Web." *Information Processing & Management* 44 (4): 1562–79. <http://dx.doi.org/10.1016/j.ipm.2007.12.010>.
- Rethlefsen, Melissa L. 2007. "Tags Help Make Libraries Del.cio.us." *Library Journal* 132 (15): 26–28.
- Rolla, Peter J. 2009. "User Tags Versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?" *Library Resources and Technical Services* 53 (3): 174–84. <http://dx.doi.org/10.5860/lrts.53n3.174>.
- Spink, Amanda, Dietmar Wolfram, B.J. Jansen, and Tefko Saracevic. 2001. "Searching the Web: The Public and Their Queries." *Journal of the American Society for Information Science and Technology* 52 (3): 226–34. [http://dx.doi.org/10.1002/1097-4571\(2000\)9999:9999::AID-ASI1591>3.0.CO;2-R](http://dx.doi.org/10.1002/1097-4571(2000)9999:9999::AID-ASI1591>3.0.CO;2-R).
- Spiteri, Louise F. 2006. "The Use of Folksonomies in Public Library Catalogues." *Serials Librarian* 51 (2): 75–89. [http://dx.doi.org/10.1300/J123v51n02\\_06](http://dx.doi.org/10.1300/J123v51n02_06).
- . 2007. "The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog." *Information Technology and Libraries* 26 (3): 13–25.
- . 2009. "The Impact of Social Cataloging Sites on the Construction of Bibliographic Records in the Public Library Catalog." *Cataloging and Classification Quarterly* 47 (1): 52–73. <http://dx.doi.org/10.1080/01639370802451991>.
- Thomas, Marliese, Dana M. Caudle, and Cecilia M. Schmitz. 2009. "To Tag or Not to Tag?" *Library Hi Tech* 27 (3): 411–34. <http://dx.doi.org/10.1108/07378830910988540>.
- Voorbij, Henk. 2012. "The Value of LibraryThing Tags for Academic Libraries." *Online Information Review* 36 (2): 196–217. <http://dx.doi.org/10.1108/14684521211229039>.

Appendix 1: List of Queries

Number	Topic	Search syntax	Format	Audience	Others (for example, date, language, and so on)
1	Find youth books on the planet Venus.	Venus	Books	Children and teens	
2	Find magazines on fashion.	Fashion	Magazines		
3	Can you help me find three to five career resources for new university graduates?	University graduates; career			
4	I am looking for biographies about Margaret Thatcher.	Margaret Thatcher; biography			
5	Find books on photograph.	Photography	Books		
6	Can you help me find novels that have dystopian themes? Three good ones should be sufficient.	dystopia	Books		
7	Find information about how to get a divorce.	Divorce; procedure			
8	Find books recommended by Oprah.	Oprah's book club	Books		
9	Find books about meditation.	Meditation	Books		
10	I am looking for some books about Adolph Hitler's life for a project.	Adolph Hitler; biography	Books		
11	I would like information about zombies. The information can be in the following format: books, DVDs, graphic novels, and ebooks, but no audio books and no books aimed at below the young adult age range. The information can include fiction and non-fiction.	zombies	Limit of no audio books	Adults	
12	Find books about hamsters.	Hamsters			
13	Find romance novels featuring hockey players and set in Canada.	Romance; hockey and Canada	Books		
14	Find children books about Italy.	Italy		Children	

## Appendix 1: (continued)

Number	Topic	Search syntax	Format	Audience	Others (for example, date, language, and so on)
15	I am looking for any printed materials on UFOs and conspiracies about them. I only want books or ebooks that are written in English, are for adults, and are non-fiction.	UFOs; conspiracies	Books and ebooks	Adults	English, non-fiction
16	Find books about nursing practices.	Nursing; practice	Books		
17	Find books for adults about drug abuse.	Drug abuse	Books	Adults	
18	I am interested in materials about Algonquin Provincial Park.	Algonquin Provincial Park			
19	I would like to find information on Norse Mythology. I would consider primary and secondary sources and various translations. I am only interested in books and ebooks. I would like at least five items.	Norse mythology	Books and ebooks		
20	I would like information about container vegetable gardens. I do not want information about flowers nor about raised—bed gardens and no information aimed at children.	Container vegetable gardens		Adults	
21	Find books about sharks.	Sharks	Books		
22	I am looking for books on feminism. They don't have to be on a particular type of feminism, just feminism in general.	Feminism	Books		
23	I would like materials on how to preserve herbs.	Herbs; preservation			
24	I am interested in information about extreme sports.	Extreme sports			
25	I want books about Barack Obama. The books must have been written after he was elected president (2009). The books can be ebooks, but no books for children. In addition, the books must be about him or his presidency in general.	Barack Obama	Books and ebooks	Adults	Post-2009

Appendix 1: (continued)

Number	Topic	Search syntax	Format	Audience	Others (for example, date, language, and so on)
26	Find books on time travel.	Time travel	Books		
27	I would like information on modern activism. It could be social activism, public activism, community activism—activism in general really. Books only please, I don't want any audio books.	Modern activism	Books but no audiobooks		
28	Find historical fiction books.	Historical fiction	Books		
29	Find materials on games and gaming.	Games; Gaming			
30	Find materials on health and fitness.	Health; Fitness			
31	I would like to know more about the life of King Henry VIII.	King Henry VIII			
32	I have recently been diagnosed with celiac disease and I'm looking for some resources on gluten-free diets.	Gluten free diet			
33	I want to find some resources that will help me to learn to knit. Three resources should be enough to get started.	Knitting			
34	I am looking for materials on how to train my boxer puppy. The materials should include videos and books but no audio. I want them in English.	Boxer puppy; training	Videos, books but no audiobooks		English
35	My daughter is doing a project on women's hockey and I'm looking for some resources to help her get started. She probably needs about five resources to begin with.	Women; hockey			
36	Can you help to find 3–4 resources on how to teach poetry to elementary students?	Poetry; teaching		Children/youth	

## Appendix 1: (continued)

Number	Topic	Search syntax	Format	Audience	Others (for example, date, language, and so on)
37	I would like to learn about Chinese alternative medicine. I'll prefer materials in DVD format.	Chinese alternative medicine	DVD		
38	I want to know more about skateboarding.	Skateboarding			
39	Find books about child abduction.	Child abduction	Books		
40	I want to learn how to make pastry dishes.	Pastry; cooking			
41	A teacher wants books on respect/manners/treatment of classmates for Grade 4 students who are not well mannered.	Respect; manners	Books	Youth/children	
42	I want to learn how to repair my own plumbing.	Plumbing; repair			
43	I am looking for books that give general information about sustainable living.	Sustainable living	Books		
44	Find picture books for three and four year olds in daycare about chores and taking responsibility.	Chores and responsibility	Picture books	Children	
45	I would like to read about the history of Islam.	History; Islam	No videos, no audiobooks		
46	I just got out of the university and need to know how to write a resume	Resumes; writing			
47	Can you help me find five resources for my project on volcanoes?	Volcanoes			
48	I am interested in materials about parenting.	Parenting			
49	Please find useful materials for me on project management.	Project management			
50	I want materials on how to manage my money.	Personal finance			