Isola Ajiferuke and Jamie Leigh Goodfellow

University of Western Ontario, London, Canada

Paper: Evaluation of the effectiveness of tag as an access point in a public library OPAC

Abstract: The objective of this study is to compare the retrieval effectiveness of using tag as an access point as against subject heading or keyword in a public library OPAC. Thirty queries were searched in Oakville Public Library, and tag retrieved a fewer number of items per query than keyword or subject. However, there was no significant difference in the average precision values.

Résumé: L'objectif de cette étude est de comparer l'efficacité du repérage à l'aide des étiquettes comme point d'accès à celui des vedettes matières ou des mots-clés d'un OPAC de bibliothèque publique. Trente requêtes ont été analysées à la bibliothèque publique d'Oakville et le repérage par étiquette a retourné moins de résultats par requête que les mots-clés ou les vedettes matières. Cependant, aucune différence significative n'a été notée pour les valeurs de précision moyenne.

1. Introduction

Social 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).

The most popular social tagging services include Delicious (http://delicious.com/), where users can tag bookmarks, Flickr (http://www.flickr.com/) , for tagging photos, YouTube (http://www.youtube.com/) , for tagging videos, and Technorati (http://technorati.com/)   for tagging blog posts. In addition, some commercial shopping websites, such as Amazon.com, have incorporated user tagging into their product pages.

Popular academic applications of social tagging include its use in bookmarking academic articles on CiteULike (http://citeulike.org/) , Connotea (http://connotea.org/) or BibSonomy (http://www.bibsonomy.org/) (Eckert, Hänger and Niemann 2009), and personal online cataloging of books on LibraryThing (http://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 catalogs (OPACs) would be beneficial (Spiteri 2006; Rethlefsen 2007; Rolla 2009; Thomas, Caudle and Schmitz 2009). Even though a few libraries,

such as Penn 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 28 other public libraries in Canada (http://www.bibliocommons.com/about/participating-libraries/canada/), 10 public libraries in the United States (http://www.bibliocommons.com/about/participating-libraries/united-states/), and 1 each in Australia and New Zealand (http://www.bibliocommons.com/about/participating-libraries/australia-new-zealand/). BiblioCore, not only allows users in a particular library to add tags to any of the items in that library, but tags added are shared with other libraries in the consortium that have the same item in their collections (http://www.bibliocommons.com/about/about-us/).

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, 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 but 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; Kipp 2011b). So, how does the assignment of varying forms of tags and non-subject tags affect the retrieval performance of tagging systems? Very few studies (Heymann, Koutrika and Garcia-Molina 2008; Kipp and Campbell 2010) have examined this issue, and none on an operational OPAC; Heymann, Koutrika and Garcia-Molina's study was based on the social bookmarking site Delicious while Kipp and Campbell's study was based on an academic publications bookmarking site CiteULike. Hence, the objective of this study is to compare the retrieval effectiveness of using tag as an access point as against subject heading or keyword in a public library OPAC.

2. Methodology

For this study, we selected the Oakville Public Library for evaluation as it was the first to adopt BiblioCore, and hence the one whose users have had the longest period of time to assign tags to the items in the collection. In addition, due to the collaborative nature of tagging in BiblioCore, shared items between Oakville Public Library and other libraries have a greater chance of being tagged.

Twenty two queries were formulated from the information needs expressed by public library users in a qualitative study exploring strategies and behaviors of public library users during interaction with an OPAC, and another eight added from reader's advisory questions that a colleague of one of the authors had received while working behind a public library reference desk. For each of the thirty queries, a search was conducted in the Oakville Public Library OPAC using each of keyword, subject, and tag as access points. The retrieved set of items was then narrowed down by format (i.e. book, magazine or journal, DVD, music CD, audiobook CD, eBook, etc.) and/or audience (i.e. children, teen, or adult) where applicable. For example, for the query "find children books about dogs", "dogs" 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 children. The Oakville Public Library OPAC allows sorting of results by relevance, date acquired, title, author, and published date with sorting by relevance as default. We made use of this default sorting for the final set, and then examined the first 30 items for relevance. The number of items examined was limited to 30 as previous studies have shown that most users hardly view more than the top 30 documents retrieved in response to a query (Spink and Wolfram 2001; Char and Ajiferuke 2011).

Precision was the only practicable effectiveness measure used as the OPAC database was 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. Precision for each query was calculated as (number of relevant items examined)/ (number of items examined).

3. **Results**

One of the queries did not retrieve any document when keyword was used as the access point (it did retrieve a few documents on the topic but none in the required format), three yielded nothing with subject as the access point, and eight queries did not retrieve any document with tag as the access point. The maximum number of documents retrieved for any query was 4482 for keyword, 4089 for subject, and 529 for tag. The distribution of number of items retrieved was skewed for each of the access points, hence the appropriate measure of central tendency is the median. The medians for the number of items retrieved were 90.5, 62.5, and 2.5 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 55.055 with a degree of freedom of 2, and a p-value of .000.   In fact, searching by keyword yielded more results than tag for 29 of the queries and yielded more results than subject for 25 of the queries with the number of results the same for the remaining 4 queries. On the other hand, searching by subject yielded more results than tag for 27 of the queries. It was not surprising to see searching by keyword yielding more results than either subject or tag because while searching by subject or tag in this OPAC

matches the user's terms with the assigned subject headings or tags, searching by keyword matches the user's terms with all indexed fields, including subject and tag.

In the case of precision, we could only obtain values for queries for which at least one item was retrieved. Hence, we obtained precision values for 29 of the queries for keyword, 27 for subject, and 22 for tag. The maximum precision value for each access point was 1.00 while the minimum values were 0.1, 0.0, and 0.0 for keyword, subject, and tag respectively. The mean precision values were .5883, .6096, and .5678 for keyword, subject, and tag respectively but a one-way analysis of variance test showed that these means were not significantly different (F-value = .098, with degrees of freedom of 2 and 75, and a p-value of .906).

4. **Conclusions**

Searching the OPAC using tag as an access point retrieved a far fewer number of items than for keyword or subject, probably due to the fact that too few items in the OPAC have been tagged. However, for the few items retrieved by tag, the mean precision value was comparable to those of keyword and subject, probably due to the small number of non-subject tags assigned by users. The authors are currently examining the types of tags assigned to items by the users of the OPAC, and should, at the end of the exercise, be able to provide a more definitive explanation for the retrieval performance of tag found in this study.

References

Char, D.C.P. & Ajiferuke, I. 2011. The usefulness of related functions in Web of Science and

    Scopus. Evidence Based Library and Information Practice, 6(1), 41-52.

Ding, Y., Jacob, E.K., Zhang, Z., Foo, S., Yan, E., George, N.L. & Guo, L. 2009. Perspectives

    on social tagging. Journal of the American Society for Information Science and

    Technology, 60(12), 2388-2401.

Eckert, K., Hanger, C. & Niemann, C. 2009. Tagging and automation: challenges and

    opportunities for academic libraries. Library Hi Tech, 27(4), 557-569.

Golder, S.A. & Huberman, B.A. 2006. Usage patterns of collaborative tagging systems. Journal

of Information Science, 32(2), 198-208.

Hedden, H. 2008. How semantic tagging increases findability. EContent, 31(8), 38-43.

Heymann, P., Koutrika, G. & Garcia-Molina, H. 2008. Can social bookmarking improve web
search? In: WSDM '08: Proceedings of the International Conference on Web Data Mining,
195-206.

Kipp, M.E.I. & Campbell, D.G. 2010. Searching with tags: do tags help users find things?
Knowledge Organization, 37(4), 239-255.

Kipp, M.E.I. 2011a. 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.

Kipp, M.E.I. 2011b. Tagging of biomedical articles on CiteULike: a comparison of user, author
and professional indexing. Knowledge Organization, 38(3), 245-261.

Lawson, K.G. 2009. Mining social tagging data for enhanced subject access for readers and
researchers. Journal of Academic Librarianship, 35(6), 574-582.

Lu, C.M., Park, J.R. & Hu, X.H. 2010. User tags versus expert-assigned subject terms: a
comparison of LibraryThing tags and Library of Congress Subject Headings. Journal of
Information Science, 36(6), 763-779.

Rethlefsen, M.L. 2007. Tags help make libraries Del.icio.us. Library Journal, 132(15), 26-28.

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.

Slone, D.J. 2000. Encounters with the OPAC: on-line searching in public libraries. Journal of the
American Society for Information Science, 51(8), 757-773.

Spink, A. & Wolfram, D. 2001. Searching the web: the public and their queries. Journal of the
American Society for Information Science and Technology, 59(13), 2070-2086.

Spiteri, L.F. 2006. The use of folksonomies in public library catalogues. The Serials Librarian,
51(2), 75-89.

Spiteri, L.F. 2007. The structure and form of folksonomy tags: the road to the public library catalog. Information Technology and Libraries, 26(3), 13-25.

Thomas, M., Caudle, D.M. & Schmitz, C.M. 2009. To tag or not to tag? Library Hi Tech, 27(3), 411-434.