"User Reviews in the Search Index? That'll Never Work!"

Marijn Koolen

Institute for Logic, Language and Computation,
University of Amsterdam, The Netherlands

Abstract. Online book search services allow users to tag and review books but do not include such data in the search index, which only contains titles, author names and professional subject descriptors. Such professional metadata is a limited description of the book, whereas tags and reviews can describe the content in more detail and cover many other aspects such as quality, writing style and engagement. In this paper we investigate the impact of including such user-generated content in the search index of a large collection of book records from Amazon and LibraryThing. We find that professional metadata is often too limited to provide good recall and precision and that both user reviews and tags can substantially improve performance. We perform a detailed analysis of different types of metadata and their impact on a number of topic categories and find that user-generated content is effective for a range of information needs. These findings are of direct relevance to large online book sellers and social cataloguing sites.

Keywords: Book Search, Metadata, Social Media, User-Generated Content.

1   Introduction

Book search behaviour on the web has become more complex with the growing amount of book information generated through social media. Readers can use this user-generated (UGC) content to help them select interesting, engaging, well-written and fun books to read. Yet most book search services, such as on GoodReads, LibraryThing (LT), Amazon and online bookshops, as well as libraries, only allow search via book titles, author names and professional metadata in the form of a small set of descriptive terms from a controlled vocabulary to describe the content of the book. But users often want to read other readers' opinions and summaries before deciding which book they want to read. For the more subjective and non-topical aspects, such retrieval systems are insufficient and force users to browse through many reviews to find the right information or turn to online networks of book readers to ask suggestions. In this paper, we investigate the impact of including UGC to the retrieval index and compare it directly with the professional metadata. Our main research question is:

- What is the impact of professional metadata and user-generated content on book search?

Professional metadata is often based on controlled vocabularies, with trained indexers assigning a small number of descriptive terms to capture the main topics of a book, with all books described in roughly the same amount of detail. With vocabulary control, descriptive terms give access to all and only books relevant to those terms. With UGC this is very different. Popular books are tagged and reviewed more often than obscure books, there is no vocabulary control and reviews can be long or short, useful or useless, honest or misleading or anywhere in between. It is unclear how UGC affects retrieval if added to the index. This leads to the following more specific research questions:

– How do professional metadata and user-generated content compare in terms of vocabulary size and frequency distribution?
– How do they compare in terms of retrieval effectiveness?

We conduct our analysis in the context of the INEX Social Book Search (SBS) Track [11][1]. The track uses a collection of book descriptions from Amazon, with subject headings and user reviews, enriched with user tags from LT. The topics and relevance judgements come from the LT discussion forums. Together, these provide a natural scenario of book search on the web. Most library searches involve keyword searching [7] so we argue our methodology and topic set are appropriate for this evaluation.

This paper is organised as follows: In Section 2 we discuss related work. Next, in Section 3 we detail the SBS Track and quantitatively compare the different types of book information. We explain our experimental setup in Section 4, discuss the evaluation results in Section 5 and provide further analysis in Section 6. Finally, we draw conclusions in Section 7.

2   Related Work

The Cranfield tests for IR evaluation [4] showed that indexing natural language terms from documents was at least as effective as formal indexing schemes with controlled languages. However, controlled vocabularies still hold the potential to improve completeness and accuracy of search results by providing consistent and rigorous index terms and ways to deal with synonymy and homonymy [12, 17]. Lu et al. [13] compared LT tags and Library of Congress Subject Headings (LCSH). They find that social tags can improve accessibility to library collections. A similar finding was reported by Sterken [16]. Peter J. Rolla [15] found that tags increase subject access but introduce noise because of personal and idiosyncratic tags. Yi and Chan [21] explored the possibility of mapping user tags to LCSH. With word matching they can link two-thirds of all tags to LC subject headings. In subsequent work [20], they use semantic similarity between tags and headings to automatically apply headings to tagged resources. This urges us to compare subject headings and user tags in an actual retrieval setting.

Searchers often find it difficult to use controlled vocabularies effectively [7]. On top of that, searchers and indexers might use different terms because they

---

have different perspectives. Buckland [2] describes the differences between vocabularies of authors, cataloguers, searchers, queries as well as the vocabulary of syndetic structure. With all these vocabularies used in a single process, there are many possibilities for mismatches. This is especially pertinent to book search if UGC is indexed, which adds the vocabularies from multiple other users. One of the interesting aspects of UGC in this respect is that it has a smaller gap with the vocabulary of searchers [14]. Golovchinsky et al. [6] constructed queries based on annotations of articles and showed they lead to better perform than feedback based on the user's relevance judgements.

Golder and Huberman [5] finds that tags have different organising functions. Describing what (or who) it is about, what it is and who owns it, refining categories, qualities or characteristics, self referencing and task organising. [1] conducted a user study to find out why LT members tag. For 42%, one of the top 3 reasons to help others find a book, for 73% it is to help their own collection management. Retrieval seems a strong motivation for assigning tags.

Some recent work already looked at social metadata for book search. Kazai and Milic-Frayling [9] incorporated social approval votes for book IR using external resources that refer to books in the corpus—such as lists from libraries and publishers and lists of bestsellers and award winning books. They find that social approval votes can improve a BM25F baseline that indexes both full-text and MARC[2] records. Koolen et al. [10] compared the effectiveness of professional metadata and UGC for book search, but used very minimal professional metadata, which does not reflect the amount in actual library catalogues. They focused on building test collections for book search based on topics and book suggestions from discussion forums. In this paper, we analyse the difference between the two types of metadata in more detail, using professional metadata from two national libraries and identifying specific strengths and weaknesses of each metadata type.

3   Social Book Search

There is a large social component to the way readers discover and select books to read. Chandler [3] analysed which channels are used by readers on GoodReads to select their next books to read and found that suggestions from offline and online friends are important sources for discovering books. Koolen et al. [10] showed that social book search and suggestion represents a different task from traditional search and recommendation tasks used in IR evaluation. On the LT discussion forums, members discuss many aspects of books and regularly start topics to ask for book recommendations. Other members join the topic thread with their suggestions [11]. These topic threads are a form of social book search. Instead of browsing or searching a book catalogue, readers rely on the book knowledge of their peers to discover interesting and fun books to read. The received suggestions are often a subset of all possibly relevant books. Someone asking for good historical fiction set in Tudor England will receive suggestions reflecting what the forum members consider to be good historical fiction, even

---

[2] http://www.loc.gov/marc/

though the LT catalogue contains many more historical fiction books set in Tudor England. The catalogue will not help the reader to identify the best, most fun or most engaging work, which is a reason for the reader to turn to the forums. The information need is more complex than what the catalogue can answer. Such needs contain aspects that are typically covered in reviews but not in professional metadata. However, users cannot search on the content of reviews and tags directly, but can only navigate through individual tags and browse through the reviews of a selected book. It is not clear what information from reviews and tags they use, but it is assumed that searchers use this data to determine whether they want to read a book or not. In terms of library catalogue objectives [18], user reviews can help searchers *choosing* which of the relevant items to read but by excluding reviews from the search index they cannot help *finding or locating* relevant items.

3.1   Comparing Data

How do professional metadata and user-generated content compare in terms of vocabulary size and frequency distribution? The SBS Track uses the INEX Amazon/LibraryThing (A/LT) collection, which contains book records from Amazon, including user reviews, and enriched with UGCfrom LT for a set of 2.8 million books. Each book is identified by ISBN and the record is marked up in XML [11]. Additional information for a large portion of those books is provided by records from the British Library (BL, 1.15 million records) and the Library of Congress (LoC, 1.25 million records). These records contain more subject headings per book than the Amazon records.

Traditional catalogues are based on controlled vocabularies to ensure consistency and rigourousness and to resolve ambiguity. A lot of care has been given to designing systems that give precise and complete results [17]. Social catalogues lack this control, which makes them easier to use for both cataloguers and searchers—they are not restricted in how they express themselves—but often leads to inconsistent use of terminology, ambiguity and inaccurate and incomplete descriptions [19]. Social cataloguing allows a much larger crowd to contribute than traditional methods, with the potential to catalogue a larger corpus of books in more detail. Both methods have advantages and disadvantages, but it is not clear how these affect retrieval performance.

Descriptors from controlled vocabularies are usually not repeated, so there is little information in the term frequency distribution. It is possible in principle to use multiple occurrences of a controlled subject term to reflect the degree to which the described object is about or related to the concept of the subject term. Why is frequency or degree not used in controlled vocabulary access points? It would require a way to measure the degree of relevance of a subject term, but thereby also allow relevance ranking. The indexer would have to analyse each book in much more detail, assigning terms for less prominent topics in the book and to determine the degree of relevance of each subject term. Currently, only subject headings for the main topics of a book are added. In natural language descriptions such as reviews, the term frequency distribution contains more information to measure

Table 1. Total and mean (median) number of types and tokens per metadata field

Field	total		per doc.	
	# types	# tokens	# types	# tokens
Book title	196,977	14,265,785	5 (4)	5 (5)
Author	224,554	7,652,715	3 (2)	3 (2)
Am. subject	87,827	5,225,797	2 (2)	2 (2)
BL/LoC LCSH	96,460	17,307,446	4 (3)	6 (3)
LT tags (set)	216,515	46,865,010	13 (6)	17 (7)
LT tags (bag)	216,515	251,868,997	13 (6)	91 (8)
Am. review	2,601,520	1,184,800,633	170 (0)	426 (0)

term relevance, but with the open nature of the web, there is no guarantee that a review contains a good description of the book. We quantitatively compare the subject headings from Amazon and from BL and LoC (BL/LoC) with the user reviews and tags as well as the book titles and author names. In Table 1 we see the mean (median) number of word types and tokens per metadata field. The numbers shown are based on single words after stopword removal and Krovetz stemming (see the next section on the experimental setup) and averaged over all 2.8 million records. For the LT tags, the bag of tags takes into account the number of users who assigned a tag to a book and the set of tags counts each tag only once per book. The second columns shows the number of term types in the index, i.e. the full vocabulary per metadata type. There are 196,977 distinct terms in the 2.8M book titles. The Amazon and BL/LoC headings have a smaller vocabulary than the other metadata types. The tag vocabulary is similar to the book title and author vocabulary, while the reviews, not surprisingly, have a far larger vocabulary. The third column shows the number of tokens. The BL/LoC subject headings have more than three times the number of tokens of the Amazon headings. Amazon records do not reflect the amount of detail of traditional library catalogues. The tags and especially the reviews have many more tokens, since multiple users can add tags and reviews to a book and reviews are often longer than subject headings. Columns four and five show the mean (median) number of types and tokens per book description. The book titles have a mean (median) of 5 (4) distinct terms and 5 (5) total terms. The author names are shorter. The Amazon subject headings are very short. The BL/LoC headings have a mean of 6 term tokens but 4 term types. In other words, there is some term repetition in subject headings. The tags and reviews show a more skewed distribution. A small number of books have a large number of tags and a lot of review text. This skew has an impact on the ranking produced by standard retrieval models.

4   Experimental Setup

For the professional metadata we use the library catalogue records of the BL and the LoC. For some books we have records from both libraries. In most of these

cases, the subject headings are identical. In the cases where they are not identical, we use both headings, which increases the number of headings and thereby the richness of the subject descriptors. For indexing we use Indri[3], Krovetz stemming and stopword removal. For retrieval we use the standard language model with Dirichlet smoothing ($\mu = 2500$) without any Indri-specific belief operators. We created indexes for four types of descriptions:

Title contains only the book titles.
Am. subject contains the subject headings from the Amazon records.
BL/LoC contains the merged subject headings of the BL and LoC records.
Review contains all Amazon reviews per book.
Tag contains all LT user tags, where the number of users who assigned a tag t to book b is taken as the term frequency $tf(t, d)$.

As Table 1 showed, there is a large difference in term frequency distributions. Therefore, we create two types of indexes:

TF$_1$ based on the term types, that is, for each term the term frequency in the index is 1.
TF$_a$ based on all term tokens, that is, for each term the term frequency in the index is based on the term frequency in the document.

This allows us to investigate the impact of term repetition in descriptions on retrieval effectiveness. The distribution of user reviews and tags is skewed: popular books tend to have many more reviews and tags than obscure books, with a larger set of terms and more term repetition. When matching query terms against document representations based on UGC, popular books have a higher probability of being retrieved and ranked highly than less popular books. To study this impact, we employ a popularity prior probability, based on the number of reviews per book, which we assume is an (imperfect) indicator of popularity. The review prior $P_r(d)$ is computed as the number of reviews $r(d)$ for book d divided by the total number of reviews r_C in the entire collection C. This is combined with the language model score as follows:

$$P(d|q) = P_r(d) \cdot P(q|d) \tag{1}$$

With this popularity prior we can investigate how much of the retrieval effectiveness of reviews and tags comes from differences in popularity.

5   Evaluation

The 2012 SBS task uses a set of 94 topics taken from the LT discussion forums. Members often link the book title they suggest to the catalogue record of the book on LT. The records of those books contain ISBNs associated with these books, which are used to map suggestions on the forum to books in the document

---

[3] Url: http://www.lemurproject.org/indri/

Table 2. Evaluation results for runs on the different indexes using the INEX 2012 SBS topics

Run	nDCG@10				R@1000			
	TF_1	$TF_1\ P_r$	TF_a	$TF_a\ P_r$	TF_1	$TF_1\ P_r$	TF_a	$TF_a\ P_r$
Title	0.0278	0.0722	0.0281	0.0726	0.2175	0.2270	0.2197	0.2277
Author	0.0213	0.0363	0.0120	0.0348	0.1043	0.1033	0.1043	0.1033
Am. subject	0.0027	0.0153	0.0075	0.0214	0.0354	0.0442	0.0376	0.0410
BL/LoC	0.0156	0.0576	0.0203	0.0583	0.1814	0.1936	0.1681	0.1975
Review	0.0184	0.0195	0.0951	0.1242	0.2465	0.1731	0.4579	0.4209
Tag	0.0124	0.0532	0.0835	0.0907	0.2655	0.2622	0.3628	0.3281

collection of the SBS Track. These suggestions are used as relevance judgements. The relevance judgements have a graded scale, with $rv = 1$ for all suggestions with two exceptions. Suggested books that the topic creator already had in her personal catalogue before starting the topic are not relevant ($rv = 0$) and suggestions that the topic creator adds to her personal catalogue after it was suggested on the forum are the most relevant ($rv = 4$) [11].

The evaluation results for the metadata types are shown in Table 2. The official evaluation measure for the task is nDCG@10. We also show recall at rank 1000 (R@1000). Book titles are generally more effective for retrieval than author names and subject headings. Performance on the BL/LoC subject headings is much better than on the Amazon subject headings. Without term frequency differentiation, UGC is not as effective as book titles and author names for precision, but with a larger number of distinct terms it gives better recall. Not surprisingly, using the term frequency information (columns 4, 5, 8 and 9) has little impact on metadata types that have little term repetition, i.e., book titles, author names and subject headings. By far the most effective are the full user reviews and tags, with the reviews resulting in the highest scores for precision and recall. Reviews and tags contain good descriptive terms for retrieving relevant books. The review prior P_{rev} leads to improved precision on all runs except on the Review TF_1 index. This shows that the good performance of reviews is not just based on the inherent popularity information, but it does play a role. However, it hurts R@1000 on the Review and Tag indexes, which is probably due to the most popular books matching any query and getting boosted to the top by their high review prior. In other words, the information needs on the forum have a specific topical focus which is easily lost by the popularity prior.

The effectiveness of UGCis partly derived from the inherent popularity signal and partly from the term distribution. Relevant search terms occur more frequently in reviews and tags than other terms. Even with the presence of misleading, off-topic and badly written reviews, the user reviews still provide a better book representation for retrieval than subject headings, book titles and author names.

What happens if we use multiple types of metadata? Indexes typically contain both book titles, author names and subject headings. Table 3 shows results

Table 3. Evaluation results for combined metadata using the INEX 2012 SBS topics

Fields	nDCG@10	R@1000
Title+Author+BL/LoC	0.0366	0.3457
Title+Author+BL/LoC+Review	0.1344	0.5294
Title+Author+BL/LoC+Tag	0.1020	0.4391
Title+Author+BL/LoC+Review+Tag	0.1640	0.5639

for runs on indexes with combinations of metadata, with full term frequency information and no review prior. The combination of title and author and subject headings leads to improvements over the individual types in both precision and recall, but is well below the performance of the Review and Tag indexes. Adding the reviews and tags leads to improvements over all indexes of the individual types. Both reviews and tags improve the book representation. Combining all metadata gives the best performance, reflecting the importance of poly-representation [8] and suggests the reviews and tags are complementary to each other.

To summarise, professional metadata, including subject headings, provide limited information on the content of books and no information on the quality, popularity and interestingness of books. User reviews and tags can improve retrieval performance when they are included in the search index. Even without any filtering of personal or idiosyncratic tags and unhelpful or misleading reviews, UGC adds more descriptive text to improve recall and better relevance cues for precision through term repetitions.

6   Analysis

The differences in performances between the subject headings on the one hand and the user reviews and tags on the other hand beg for further analysis. A lot of professional effort is spent on creating, maintaining, updating and assigning subject headings and library studies have often stated that UGC is too messy and unreliable to be a suitable alternative for searching, but our findings point in the opposite direction. In this section we break down the results and try to identify in which cases UGC performs better than subject headings and vice versa. We analyse the topic titles and categorise them on the types of aspects they cover, namely formal metadata (name of author, book or series) and content (genre, subject). The formal metadata topics we divide further into sets of topics that ask for the best *edition* of a specific work, best *starting point* for an author or series or for *similar* works to a named work or author. For instance, topic 29129 asks for the best *edition* of 'Canterbury Tales', topic 30984 asks for a good *starting point* for the works of 'Eudora Welty' and topic 10392 asks for books *similar* to 'David Copperfield'. The content-related topics we divide into sets that ask for books on a certain genre (topic 115958 asks for *western horror* recommendations), books about a certain subject, which we further split into named entity subjects (topic 50302 asks for books about *Cedar Creek* which is

Table 4. Evaluation results of the BL/LOC, Review and Tag indexes per topic category

	#tpcs	nDCG@10			Recall$_{set}$		
		BL/LoC	Review	Tag	BL/LoC	Review	Tag
Metadata	25	0.0524	0.1753	0.0860	0.1795	0.7344	0.5860
Edition	3	0.0000	0.0534	0.0000	0.2731	0.5880	0.5185
Similar	8	0.0476	0.0552	0.0847	0.0849	0.6729	0.3638
Episode	14	0.0664	0.2701	0.1051	0.2136	0.8008	0.7274
Metadata & Subject	2	0.0000	0.1034	0.0665	0.4445	1.0000	0.5000
Content	69	0.0108	0.0625	0.0823	0.4174	0.6652	0.6667
Genre	11	0.0000	0.0292	0.1025	0.2972	0.6811	0.7521
Genre & Subject	15	0.0064	0.0795	0.0663	0.3874	0.6961	0.6906
Subject Named	12	0.0080	0.0970	0.0779	0.3201	0.6467	0.6247
Subject Unnamed	3	0.0000	0.0096	0.0200	0.6563	0.8937	0.9542
Subject	43	0.0151	0.0651	0.0962	0.4586	0.6503	0.6364
Named	34	0.0191	0.0720	01216	0.5333	0.6749	0.7018
Unnamed	9	0.0000	0.0391	0.0000	0.1763	0.5574	0.3896
All	94	0.0203	0.0951	0.0835	0.3560	0.6902	0.6422

the name of both a place and an event) and non-entity subjects (topic 26348 asks for books on *portraiture*). Topics can also contain a combination of formal metadata and content (genre and/or subject).

These categories present a broad and challenging but natural set of search tasks. Finding similar books is a form of item-item recommendation. Finding editions of a work is a form of known-item search and the subject-related topics resemble classical topic-relevance search tasks, but with an element of recommendation, as not all books on a subject are suggested.

Performance per category is shown in Table 4. For recall, subject headings score better on content-related topics than on formal metadata-related topics, while for precision they score better on formal metadata-related topics and very badly on subject related-topics. This is surprising, given that their purpose is to support subject access while formal access is supported by title information. The Review scores high on Metadata related topics, suggesting that reviews typically contain metadata terms such book titles and author names. Tags are particularly effective for genre-related topics and subject-topics. When the topic combines both subject and genre, reviews are at least as effective. Both reviews and tags contain relevant genre and subject terms. For precision, tags score well on named subject terms, but fail to score for general subjects. In sum, UGC is not only effective for traditional subject search tasks, but also for known-item search, genre search and complex combinations of these tasks. Adding UGC to the document representations for indexing increases retrieval effectiveness for a broad range of tasks.

On the last row of Table 4, we see set-based recall over all 94 topics. On the BL/LoC index recall is 0.36, on the review index 0.69 and on the tag index 0.64. That is, on average, only 36% of the suggested books can be found through subject

Table 5. Set-based performance with single subject headings optimised for precision, recall and F_1

Optimised for	Precision	Recall	F_1
Precision	0.8878	0.1966	0.2563
Recall	0.1897	0.5307	0.1436
F_1	0.6402	0.2473	0.2818

headings, while the majority of suggestions is found through reviews and tags. This has consequences for potential use of precision devices like low smoothing (missing query terms is heavily punished, resulting in coordination level ranking). The low recall on the BL/LoC index offers little potential for improving precision and might have to be increased first through devices like pseudo-relevance feedback.

Is there a relation between the per-topic recall of the three indexes, i.e. do they perform well on the same topics? For that we look at the per-topic set-based recall distribution of the BL/LoC, Review and Tag indexes and compute the Kendall's τ ranking correlation between runs on the BL/LoC, review and tag indexes. Between the BL/LoC and Review indexes the correlation is low ($\tau = 0.35$), meaning that the topics on which the BL/LoC index scores highest are different from the topics on which the Review index scores highest. Between the BL/LoC and Tag index this correlation is stronger (0.60) indicating that the subject headings are more similar to tags than to reviews. The correlation between reviews and tags is 0.61. For recall, tags are somewhere in between subject headings and reviews.

6.1   Selecting Subject Headings

Subject headings are not meant to be used (only) for keyword searches. In library catalogues, users can select specific headings from a list or from the headings assigned to a specific book. The heading will show the user a list of catalogue items to which the heading is assigned. A good subject heading may be different from a keyword query. This prompts the question whether the user could do better by selecting an appropriate heading instead typing her own query.

In Table 5 we show the set-based precision, recall and F_1 score when the user would choose the optimal subject heading for precision (row 1), recall (row 2) or F_1 (row 3). High precision can be achieved but at the cost of recall. In many cases, the subject heading is very specific and leads to a single book. If the user aims for high recall, the subject headings is more general and contains many books not relevant to her information need. This mismatch between available subject headings and users' information needs is often ignored in analyses of keyword search success in library catalogues.

7   Conclusions

In this paper we investigated the impact of indexing professional metadata and user-generated content on book search.

We first compared professional metadata and UGC in terms of vocabulary size and frequency distribution. Professional book descriptions have a small number of subject headings representing the content of the book. The number of descriptive terms is small and has little term frequency information to signal what the most relevant terms are. The amount of descriptive text is relatively uniform across all books. With UGC this is more skewed, with popular books described with many more terms than obscure books. Descriptive terms are often repeated, thereby providing relevance cues. This popularity effect is a larger for reviews than for tags—assigning tags takes less effort than writing a review—so reviews give access to fewer books than tags, but they do provide a richer set of descriptive terms.

Next, we compared professional metadata and UGC in terms of retrieval effectiveness. Book titles, author names and subject headings lead to low precision, as they provide few cues to what the best, most relevant books are on a topic. Both reviews and tags lead to better performance compared to the professional metadata across many different topic types. Combining them further improves performance. Reviews and tags provide a large number of descriptive terms, leading to higher recall and the relevant terms are repeated more frequently leading to higher precision. On top of that, the skewed distribution of UGC functions like an inherent popularity prior, favouring books that more users want to read.

What is the impact of professional metadata and user-generated content on book search? Professional metadata is too limited to allow effective retrieval of books. The lack of term repetition leads to weak relevance cues and the small number of descriptive terms does not cover the content and non-content aspects of books in sufficient detail. User reviews and tags are more effective for retrieval and to some extent complementary to each other and to professional metadata. Moreover, reviews can contain important information that is lacking from professional metadata. Subject headings also allow users to browse through sets of books on the same topic in a controlled way, but for keyword searches UGC is much more effective.

These findings are of direct relevance to large online book shops and cataloguing sites such as Amazon, GoodReads, LT as well as the numerous online national and international library catalogues. Although including UGC unmodified to the search index opens the search process up to exploitation by spammers and commercial interests, companies with book search services should be aware of the value of tags and reviews for retrieval and can perhaps find safe ways of including (parts of) the user data in the index. Although we have only looked at books, we expect these findings generalise to others product categories where metadata, reviewing and searching are similar in nature, like music and films.

References

[1] Bartley, P.: Book tagging on librarything: How, why, and what are in the tags? Proceedings of the American Society for Information Science and Technology 46(1), 1–22 (2009)

[2] Buckland, M.: Vocabulary as a Central Concept in Library and Information Science. In: Proceedings of CoLIS3 (1999)

[3] Chandler, O.: How Consumers Discover Books Online (2012), http://cdn.oreillystatic.com/en/assets/1/event/73/How

[4] Cleverdon, C.W.: The Cranfield tests on index language devices. Aslib 19, 173–192 (1967)

[5] Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)

[6] Golovchinsky, G., Price, M.N., Schilit, B.N.: From reading to retrieval: free-form ink annotations as queries. In: SIGIR 1999, pp. 19–25 (1999)

[7] Gross, T., Taylor, A.G.: What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results. College & Research Libraries 66(3) (2005)

[8] Ingwersen, P.: Polyrepresentation of information needs and semantic entities: Elements of a cognitive theory for information retrieval interaction. In: Croft, W.B., van Rijsbergen, C.J. (eds.) SIGIR, pp. 101–110. ACM/Springer (1994)

[9] Kazai, G., Milic-Frayling, N.: Effects of social approval votes on search performance. In: Third International Conference on Information Technology: New Generations, pp. 1554–1559 (2009)

[10] Koolen, M., Kamps, J., Kazai, G.: Social Book Search: The Impact of Professional and User-Generated Content on Book Suggestions. In: CIKM 2012. ACM (2012)

[11] Koolen, M., Kazai, G., Kamps, J., Doucet, A., Landoni, M.: Overview of the INEX 2012 Social Book Search Track. In: Geva, S., Kamps, J., Schenkel, R. (eds.) INEX 2011. LNCS, vol. 7424, pp. 1–29. Springer, Heidelberg (2012)

[12] Lancaster, F.W.: Vocabulary control for information retrieval, 2nd edn. Information Resources Press, Arlington (1986)

[13] Lu, C., Jung-ran, P., Hu, X.: 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 (2010)

[14] Mathes, A.: Folksonomies - cooperative classification and communication through shared metadata (December 2004)

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

[16] Sterken, V.: Folksonomy as a thing for a library: An analysis of user generated metadata in Library Thing. Cahiers de la Documentation–Bladen Voor Documentatie 1, 9 (2009)

[17] Svenonius, E.: Unanswered questions in the design of controlled vocabularies. Journal of the American Society for Information Science 37(5), 331–340 (1986)

[18] Svenonius, E.: The Intellectual Foundation of Information Organization. MIT Press (2000) ISBN 0-262-19433-3

[19] Voss, J.: Tagging, folksonomy & co - renaissance of manual indexing? CoRR, abs/cs/0701072 (2007)

[20] Yi, K.: A semantic similarity approach to predicting library of congress subject headings for social tags. JASIST 61(8), 1658–1672 (2010)

[21] Yi, K., Chan, L.M.: Linking folksonomy to Library of Congress subject headings: an exploratory study. Journal of Documentation 65(6), 872–900 (2009)