Overview of the SBS 2016 Suggestion Track

Koolen, M.; Bogers, T.; Kamps, J.

[Link to publication](#)

Overview of the SBS 2016 Suggestion Track

Marijn Koolen[1,2], Toine Bogers[3], and Jaap Kamps[1]

[1] University of Amsterdam, Netherlands
{marijn.koolen,kamps}@uva.nl
[2] Netherlands Institute of Sound and Vision
mkoolen@beeldengeluid.nl
[3] Aalborg University Copenhagen
toine@hum.aau.dk

Abstract. The goal of the SBS 2016 Suggestion Track is to evaluate approaches for supporting users in searching collections of books who express their information needs both in a query and through example books. The track investigates the complex nature of relevance in book search and the role of traditional and user-generated book metadata in retrieval. We consolidated last year's investigation into the nature of book suggestions from the LibraryThing forums and how they compare to book relevance judgements. Participants were encouraged to incorporate rich user profiles of both topic creators and other LibraryThing users to explore the relative value of recommendation and retrieval paradigms for book search. In terms of systems evaluation, the most effective systems include ...

1 Introduction

The goal of the Social Book Search 2016 Suggestion Track[4] is to investigate techniques to support users in searching for books in catalogues of professional metadata and complementary social media. Towards this goal the track is building appropriate evaluation benchmarks, complete with test collections for social, semantic and focused search tasks. The track provides opportunities to explore research questions around two key areas:

- Evaluation methodologies for book search tasks that combine aspects of retrieval and recommendation,
- Information retrieval techniques for dealing with professional and user-generated metadata,

The *Social Book Search* (SBS) 2016 Suggestion Track, framed within the scenario of a user searching a large online book catalogue for a given topic of interest, aims at exploring techniques to deal with complex information needs—that go beyond topical relevance and can include aspects such as genre, recency, engagement, interestingness, and quality of writing—and complex information

---

[4] See http://social-book-search.humanities.uva.nl/#/suggestion

Table 1. Active participants of the INEX 2014 Social Book Search Track and number of contributed runs

Institute	Acronym	Runs
Aix-Marseille Université CNRS (LSIS-OpenEdition)	LSIS	4
Chaoyang University of Technology	CYUT-CSIE	6
Indian School of Mines Dhanbad	ISMD	6
Know-Center	Know	2
Laboratoire d'Informatique de Grenoble	MRIM	6
Oslo & Akershus University College of Applied Sciences	OAUC	3
Research Center on Scientific and Technical Information	CERIST	6
University of Amsterdam	UvA	1
University of Neuchtel, Zurich University of Applied Sciences	UniNe-ZHAW	6
University of Science and Technology Beijing	USTB-PRIR	6
Total		46

sources that include user profiles, personal catalogues, and book descriptions containing both professional metadata and user-generated content.

The Suggestion Track has been part of the SBS Lab since 2015 and is a continuation of the INEX SBS Track that ran from 2011 up to 2014. The focus is on search requests that combine a natural language description of the information need as well as example books, combining traditional ad hoc retrieval with query-by-document. The information needs are derived from the LibraryThing (LT) discussion forums. LibraryThing forum requests for book suggestions, combined with annotation of these requests resulted in a topic set of 120 topics with graded relevance judgments. A test collection is constructed around these information needs and the Amazon/LibraryThing collection, consisting of 2.8 million documents. The Suggestion Track runs in close collaboration with the SBS Interactive Track,[5] which is a user-centered track where interfaces are developed and evaluated and user interaction is analysed to investigate how book searchers make use of professional metadata and user-generated content.

2 Participating Organisations

A total of 29 organisations registered for the track, with 10 teams submitting runs (see Table 1). In 2015 25 teams registered and 10 submitted runs, so participation is stable.

---

[5] See http://social-book-search.humanities.uva.nl/#/interactive

3 Suggestion Track Setup

3.1 Track Goals and Background

The goal of the Suggestion Track is to evaluate the value of professional metadata and user-generated content for book search on the Web and to develop and evaluate systems that can deal with both retrieval and recommendation aspects, where the user has a specific information need against a background of personal tastes, interests and previously seen books.

Through social media, book descriptions have extended far beyond what is traditionally stored in professional catalogues. Not only are books described in the users' own vocabulary, but are also reviewed and discussed online, and added to online personal catalogues of individual readers. This additional information is subjective and personal, and opens up opportunities to aid users in searching for books in different ways that go beyond the traditional editorial metadata based search scenarios, such as known-item and subject search. For example, readers use many more aspects of books to help them decide which book to read next (Reuter, 2007), such as how engaging, fun, educational or well-written a book is. In addition, readers leave a trail of rich information about themselves in the form of online profiles, which contain personal catalogues of the books they have read or want to read, personally assigned tags and ratings for those books and social network connections to other readers. This results in a search task that may require a different model than traditional ad hoc search (Koolen et al., 2012) or recommendation.

The SBS track investigates book requests and suggestions from the Library-Thing (LT) discussion forums as a way to model book search in a social environment. The discussions in these forums show that readers frequently turn to others to get recommendations and tap into the collective knowledge of a group of readers interested in the same topic.

The track builds on the INEX Amazon/LibraryThing (A/LT) collection (Beckers et al., 2010), which contains 2.8 million book descriptions from Amazon, enriched with content from LT. This collection contains both professional metadata and user-generated content.

The SBS Suggestion Track aims to address the following research questions:

– Can we build reliable and reusable test collections for social book search based on book requests and suggestions from the LT discussion forums?
– Can user profiles provide a good source of information to capture personal, affective aspects of book search information needs?
– How can systems incorporate both specific information needs and general user profiles to combine the retrieval and recommendation aspects of social book search?
– What is the relative value of social and controlled book metadata for book search?

3.2 Scenario

The scenario is that of a user turning to Amazon Books and LT to find books to read, to buy or to add to their personal catalogue. Both services host large collaborative book catalogues that may be used to locate books of interest.

On LT, users can catalogue the books they read, manually index them by assigning tags, and write reviews for others to read. Users can also post messages on discussion forums asking for help in finding new, fun, interesting, or relevant books to read. The forums allow users to tap into the collective bibliographic knowledge of hundreds of thousands of book enthusiasts. On Amazon, users can read and write book reviews and browse to similar books based on links such as "customers who bought this book also bought... ".

Users can search online book collections with different intentions. They can search for specific known books with the intention of obtaining them (buy, download, print). Such needs are addressed by standard book search services as offered by Amazon, LT and other online bookshops as well as traditional libraries. In other cases, users search for a specific, but unknown, book with the intention of identifying it. Another possibility is that users are not looking for a specific book, but hope to discover one or more books meeting some criteria. These criteria can be related to subject, author, genre, edition, work, series or some other aspect, but also more serendipitously, such as books that merely look interesting or fun to read or that are similar to a previously read book.

3.3 Task description

The task is to reply to a user request posted on a LT forum (see Section 4.1) by returning a list of recommended books matching the user's information need. More specifically, the task assumes a user who issues a query to a retrieval system, which then returns a (ranked) list of relevant book records. The user is assumed to inspect the results list starting from the top, working down the list until the information need has been satisfied or until the user gives up. The retrieval system is expected to order the search results by relevance to the user's information need.

The user's query can be a number of keywords, but also one or more book records as positive or negative examples. In addition, the user has a personal profile that may contain information on the user's interests, list of read books and connections with other readers. User requests may vary from asking for books on a particular genre, looking for books on a particular topic or period or books written in a certain style. The level of detail also varies, from a brief statement to detailed descriptions of what the user is looking for. Some requests include examples of the kinds of books that are sought by the user, asking for similar books. Other requests list examples of known books that are related to the topic, but are specifically of no interest. The challenge is to develop a retrieval method that can cope with such diverse requests.

The books must be selected from a corpus that consists of a collection of curated and social book metadata, extracted from Amazon Books and LT, extended with associated records from library catalogues of the Library of Congress

and the British Library (see the next section). Participants of the Suggestion track are provided with a set of book search requests and user profiles and are asked to submit the results returned by their systems as ranked lists.

The track thus combines aspects from retrieval and recommendation. On the one hand the task is akin to directed search familiar from information retrieval, with the requirement that returned books should be topically relevant to the user's information need described in the forum thread. On the other hand, users may have particular preferences for writing style, reading level, knowledge level, novelty, unusualness, presence of humorous elements and possibly many other aspects. These preferences are to some extent reflected by the user's reading profile, represented by the user's personal catalogue. This catalogue contains the books already read or earmarked for future reading, and may contain personally assigned tags and ratings. Such preferences and profiles are typical in recommendation tasks, where the user has no specific information need, but is looking for suggestions of new items based on previous preferences and history.

3.4 Submission Format

Participants are asked to return a ranked list of books for each user query, ranked by order of relevance, where the query is described in the LT forum thread. We adopt the submission format of TREC, with a separate line for each retrieval result (i.e., book), consisting of six columns:

1. topic_id: the topic number, which is based on the LT forum thread number.
2. Q0: the query number. Unused, so should always be Q0.
3. isbn: the ISBN of the book, which corresponds to the file name of the book description.
4. rank: the rank at which the document is retrieved.
5. rsv: retrieval status value, in the form of a score. For evaluation, results are ordered by descending score.
6. run_id: a code to identify the participating group and the run.

Participants are allowed to submit up to six runs, of which at least one should use only the *title* field of the topic statements (the topic format is described in Section 4.1). For the other five runs, participants could use any field in the topic statement.

4 Test Collection

We use and extend the Amazon/LibraryThing (A/LT) corpus crawled by the University of Duisburg-Essen for the INEX Interactive Track (Beckers et al., 2010). The corpus contains a large collection of book records with controlled subject headings and classification codes as well as social descriptions, such as tags and reviews. See https://inex.mmci.uni-saarland.de/data/nd-agreements.jsp for information on how to gain access to the corpus.

Table 2. A list of all element names in the book descriptions

tag name			
book	similarproducts	title	imagecategory
dimensions	tags	edition	name
reviews	isbn	dewey	role
editorialreviews	ean	creator	blurber
images	binding	review	dedication
creators	label	rating	epigraph
blurbers	listprice	authorid	firstwordsitem
dedications	manufacturer	totalvotes	lastwordsitem
epigraphs	numberofpages	helpfulvotes	quotation
firstwords	publisher	date	seriesitem
lastwords	height	summary	award
quotations	width	editorialreview	browseNode
series	length	content	character
awards	weight	source	place
browseNodes	readinglevel	image	subject
characters	releasedate	imageCategories	similarproduct
places	publicationdate	url	tag
subjects	studio	data	

The collection consists of 2.8 million book records from Amazon, extended with social metadata from LT. This set represents the books available through Amazon. Each book is identified by an ISBN. Note that since different editions of the same work have different ISBNs, there can be multiple records for a single intellectual work. Each book record is an XML file with formal metadata fields like *isbn*, *title*, *author*, *publisher*, *dimensions*, *numberofpages* and *publicationdate*. Curated metadata comes in the form of a Dewey Decimal Classification in the *dewey* field, Amazon subject headings in the *subject* field, and Amazon category labels in the *browseNode* fields. The social metadata from Amazon and LT is stored in the *tag*, *rating*, and *review* fields. The full list of fields is shown in Table 2.

To ensure that there is enough high-quality metadata from traditional library catalogues, we extended the A/LT data set with library catalogue records from the Library of Congress (LoC) and the British Library (BL). We only use library records of ISBNs that are already in the A/LT collection. These records contain formal metadata such as title information (book title, author, publisher, etc.), classification codes (mainly DDC and LCC) and rich subject headings based on the Library of Congress Subject Headings (LCSH).[6] Both the LoC records and the BL records are in MARCXML[7] format.

---

[6] For more information see: http://www.loc.gov/aba/cataloging/subject/

[7] MARCXML is an XML version of the well-known MARC format. See: http://www.loc.gov/standards/marcxml/

Fig. 1. A topic thread in LibraryThing, with suggested books listed on the right hand side.

4.1 Information needs

LT users discuss their books on the discussion forums. Many of the topic threads are started with a request from a member for interesting, fun new books to read. Users typically describe what they are looking for, give examples of what they like and do not like, indicate which books they already know and ask other members for recommendations. Members often reply with links to works catalogued on LT, which, in turn, have direct links to the corresponding records on Amazon. These requests for recommendations are natural expressions of information needs for a large collection of online book records. We use a sample of these forum topics to evaluate systems participating in the Suggestion Track.

Each topic has a title and is associated with a group on the discussion forums. For instance, topic 99309 in Figure 1 has the title *Politics of Multiculturalism Recommendations?* and was posted in the group *Political Philosophy.* The books suggested by members in the thread are collected in a list on the side of the topic thread (see Figure 1). A feature called *touchstone* can be used by members to easily identify books they mention in the topic thread, giving other readers of the thread direct access to a book record in LT, with associated ISBNs and links to Amazon. We use these suggested books as initial relevance judgements for evaluation. In the rest of this paper, we use the term *suggestion* to refer to a book that has been identified in a touchstone list for a given forum topic. Since all suggestions are made by forum members, we assume they are valuable judgements on the relevance of books. Additional relevance information can be gleaned from

the discussions on the threads. Consider, for example, topic 129939[8]. The topic starter first explains what sort of books he is looking for, and which relevant books he has already read or is reading. Other members post responses with book suggestions. The topic starter posts a reply describing which suggestions he likes and which books he has ordered and plans to read. Later on, the topic starter provides feedback on the suggested books that he has now read. Such feedback can be used to estimate the relevance of a suggestion to the user.

In the following, we first describe the topic selection and annotation procedure, then how we used the annotations to assign relevance values to the suggestions, and finally the user profiles, which were then provided with each topic.

Topic selection The topic set of 2016 is a newly selected set of topics from the LibraryThing discussion forums. A total of 2000 topic threads were assessed on whether they contain a book search request by four judges, with 272 threads labelled as book search requests. To establish inter-annotator agreement, 453 threads were double-assessed, resulting a Cohen's Kappa of 0.83. Judges strongly agree on which posts are book search requests and which are not. Of these 272 book search requests, 124 (46%) are known-item searches from the *Name that Book* discussion group. Here, LT members start a thread to describe a book they know but cannot remember the title and author of and ask others for help. In earlier work we found that known-item topics behave very differently from the other topic types ((Koolen et al., 2015)). We remove these topics from the topic set so that they do not dominate the performance comparison. Furthermore, we removed topics that have no book suggestions by other LT members and topics for which we have no user profile of the topic starter, resulting in a topic set of 120 topics for evaluation of the 2016 Suggestion Track.

Topics are distributed to participants in XML format with as richly annotated requests. As an example, the topic in Figure 1 (topic 99309) would look like this:

```
<topic>
  <topicid>99309</topicid>
  <query>Politics of Multiculturalism</query>
  <title>Politics of Multiculturalism Recommendations?</title>
  <group>Political Philosophy</group>
  <request> I'm new, and would appreciate any recommended reading on
    the politics of multiculturalism. <a href="/author/parekh">Parekh
    </a>'s <a href="/work/164382">Rethinking Multiculturalism: Cultural
    Diversity and Political Theory</a> (which I just finished) in the end
    left me unconvinced, though I did find much of value I thought he
    depended way too much on being able to talk out the details later. It
    may be that I found his writing style really irritating so adopted a
    defiant skepticism, but still... Anyway, I've read
    <a href="/author/sen">Sen</a>, <a href="/author/rawles">Rawls</a>,
    <a href="/author/habermas">Habermas</a>, and
```

```
  <a href="/author/nussbaum">Nussbaum</a>, still don't feel like I've
  wrapped my little brain around the issue very well and would
  appreciate any suggestions for further anyone might offer.
</narrative>
<examples>
  <example>
    <workid>164382</workid>
    <booktitle>Rethinking Multiculturalism: Cultural Diversity and Political Theory</booktitle>
    <author>Bhikhu Parekh</author>
  </example>
</examples>
<catalog>
  <work>
    <workid>9036</workid>
    <booktitle>The Confessions of St. Augustine</booktitle>
<author>Saint Augustine, Bishop of Hippo</author>
    <publication-year>397</publication-year>
    <cataloguing-date>2007-09</cataloguing-date>
    <rating>0.0</rating>
    <tags></tags>
  </work>
  <work>
    ...
```

The hyperlink markup, represented by the `<a>` tags, is added by the *Touchstone* technology of LT. The rest of the markup is generated specifically for the Suggestion Track. Above, the book with *work ID* 164382 is annotated as an example of what the requester is looking for.

Suggestion provided by other LT members are often marked up in touchstones so they are easy to extract and use as relevance judgements. As for the 2015 Suggestion Track, we manually annotated all book suggestions that were not marked up as touchstones by LT members to provide a more complete recall base.

Operationalisation of forum judgement labels In previous years the Suggestion Track used a complicated decision tree to derive a relevance value from a suggestion. To reduce the number of assumptions, we simplified the mapping of book suggestions to relevance values. By default a suggested book has a relevance value of 1. Books that the requester already has in her personal catalogue before starting the thread (pre-catalogued suggestions) have little additional value are assumed to have a relevance value of 0. On the other hand, suggestions that the requester subsequently adds to her catalogue (post-catalogued suggestions) are assumed to be the most relevant suggestions and receive a relevance value of 8, to keep that relevance level the same is in 2014 and 2015. Note that some of the books mentioned in the forums are not part of the 2.8 million books in our collection. These suggestions removed from the suggestions any books that are not in the INEX A/LT collection. The numbers reported in the previous section were calculated after this filtering step.

User profiles and personal catalogues From LT we can not only extract the information needs of social book search topics, but also the rich user profiles of the topic creators and other LT users, which contain information on which books they have in their personal catalogue on LT, which ratings and tags they assigned to them and a social network of friendship relations, interesting library relations and group memberships. These profiles may provide important signals on the user's topical and genre interests, reading level, which books they already know and which ones they like and don't like. These profiles were scraped from the LT site, anonymised and made available to participants. This allows Track participants to experiment with combinations of retrieval and recommender systems. One of the research questions of the SBS task is whether this profile information can help systems in identifying good suggestions.

Although the user expresses her information need in some detail in the discussion forum, she may not describe all aspects she takes into consideration when selecting books. This may partly be because she wants to explore different options along different dimensions and therefore leaves some room for different interpretations of her need. Another reason might be that some aspects are not related directly to the topic at hand but may be latent factors that she takes into account with selecting books in general.

To anonymise all user profiles, we first removed all friendship and group membership connections and replaced the user name with a randomly generated string. The cataloguing date of each book was reduced to the year and month. What is left is an anonymised user name, book ID, month of cataloguing, rating and tags. We distributed a set of 94,656 user profiles containing over 33 million transactions.

ISBNs and Intellectual Works Each record in the collection corresponds to an ISBN, and each ISBN corresponds to a particular intellectual work. An intellectual work can have different editions, each with their own ISBN. The ISBN-to-work relation is a many-to-one relation. In many cases, we assume the user is not interested in all the different editions, but in different intellectual works. For evaluation we collapse multiple ISBN to a single work. The highest ranked ISBN is evaluated and all lower ranked ISBNs of the same work ignored. Although some of the topics on LibraryThing are requests to recommend a particular edition of a work—in which case the distinction between different ISBNs for the same work are important—we ignore these distinctions to make evaluation easier. This turns edition-related topics into known-item topics.

However, one problem remains. Mapping ISBNs of different editions to a single work is not trivial. Different editions may have different titles and even have different authors (some editions have a foreword by another author, or a translator, while others have not), so detecting which ISBNs actually represent the same work is a challenge. We solve this problem by using mappings made by the collective work of LibraryThing members. LT members can indicate that two books with different ISBNs are actually different manifestations of the same

intellectual work. Each intellectual work on LibraryThing has a unique work ID, and the mappings from ISBNs to work IDs is made available by LibraryThing.[9]

The mappings are not complete and might contain errors. Furthermore, the mappings form a many-to-many relationship, as two people with the same edition of a book might independently create a new book page, each with a unique work ID. It takes time for members to discover such cases and merge the two work IDs, which means that at any time, some ISBNs map to multiple work IDs even though they represent the same intellectual work. LibraryThing can detect such cases but, to avoid making mistakes, leaves it to members to merge them. The fraction of works with multiple ISBNs is small so we expect this problem to have a negligible impact on evaluation.

5 Evaluation

This year, 10 teams submitted a total of 46 runs (see Table 1). The evaluation results shown in Table 3. The official evaluation measure for this task is nDCG@10. It takes graded relevance values into account and is designed for evaluation based on the top retrieved results. In addition, P@10, MAP and MRR scores are also reported.

The best runs of the top 5 groups are described below:

1. *USTB-PRIR - run1.keyQuery_active_combineRerank* (rank 1): This run was made by a searching-re-ranking process where the ini- tial retrieval result was based on the selection of query keywords and a small index of active books, the re-ranking results based on a combination of several strategies (number of people who read the book from profile, similar-product from amazon.com, popularity from LT forum, etc.). At indexing time, the collection is enriched with book metadata from other sources, particularly books that have little metadata from Amazon and LibraryThing. The full topic starter post in the `<request>` field is filtered based key word lists from topics of the SBS Tracks of 2011–2015. The top 1000 retrieval results of a Language Model with default parameters is re-ranked using a number of query-independent features.

2. *CERIST - all_features* (rank 7): The topic statement in the *request* field is treated as a verbose query and is reduced using several features based on term statistics, Part-Of-Speech tagging, and whether terms from the *request* field occur in the user profile and example books.

3. *CYUT-CSIE - 0.95Averageword2vecType2TGR* (rank 11): This run uses query expansion based on word embeddings using word2vec, on top of a standard Lucene index and retrieval model. For this run, queries are represented by a combination of the *title*, *group* and *request* fields. Results are re-ranked using a linear combination of the original retrieval score and the average Amazon user ratings of the retrieved books.

---

[9] See: http://www.librarything.com/feeds/thingISBN.xml.gz

Table 3. Evaluation results for the official submissions. Best scores are in bold. Runs marked with * are manual runs.

Group	Run	nDCG@10	P@10	MRR	MAP
USTB-PRIR	run1.keyQuery_active_combineRerank	0.2157	0.5247	0.1253	0.3474
USTB-PRIR	run2.keyQuery_active_userNumRerank	0.2047	0.4700	0.1177	0.3474
USTB-PRIR	run6.keyQuery_AllRerank	0.2030	0.4868	0.1144	0.3146
USTB-PRIR	run5.keyQuery_readByOne	0.2009	0.4767	0.1128	0.3146
USTB-PRIR	run3.keyQuery_active_readByOneReRank	0.1989	0.4923	0.1157	0.3474
USTB-PRIR	run4.keyQuery_active_similarRerank	0.1935	0.4685	0.1106	0.3474
CERIST	all_features	0.1567	0.3513	0.0838	0.4330
CERIST	all_no_field_feature	0.1438	0.3275	0.0754	0.3993
CERIST	all_with_filter	0.1418	0.3335	0.0780	0.3961
CERIST	stat_ling_features	0.1290	0.2970	0.0816	0.4560
CYUT-CSIE	0.95Averageword2vecType2TGR	0.1158	0.2563	0.0563	0.1603
CYUT-CSIE	0.95AverageType2TGR	0.1137	0.2718	0.0572	0.1626
CYUT-CSIE	word2vecType2TGR	0.1107	0.2479	0.0542	0.1614
CERIST	stat_features	0.1082	0.2279	0.0749	0.4326
CERIST	topic_profil_features	0.1077	0.2635	0.0627	0.4368
CYUT-CSIE	Type2TGR	0.1060	0.2545	0.0550	0.1635
UvA-ILLC	base_es	0.0944	0.2272	0.0548	0.3122
MRIM	RUN2	0.0889	0.1889	0.0518	0.3491
MRIM	RUN6	0.0872	0.1914	0.0538	0.3652
MRIM	RUN3	0.0872	0.1914	0.0538	0.3652
MRIM	RUN1	0.0864	0.1858	0.0529	0.3654
MRIM	RUN5	0.0861	0.1866	0.0525	0.3652
MRIM	RUN4	0.0861	0.1866	0.0524	0.3652
ISMD	ISMD16allfieds	0.0765	0.1722	0.0342	0.2157
UniNe-ZHAW	Pages_INEXSBS2016_SUM_SCORE	0.0674	0.1512	0.0472	0.2556
UniNe-ZHAW	RatingsPagesPrice_INEXSBS2016_SUM_SCORE	0.0667	0.1499	0.0462	0.2556
UniNe-ZHAW	PagesPrice_INEXSBS2016_SUM_SCORE	0.0665	0.1442	0.0461	0.2556
ISMD	ISMD16titlefield	0.0639	0.1197	0.0333	0.1933
ISMD	ISMD16requestfield	0.0613	0.1454	0.0287	0.1870
UniNe-ZHAW	Ratings_INEXSBS2016_SUM_SCORE	0.0584	0.1332	0.0419	0.2556
UniNe-ZHAW	INEXSBS2016	0.0561	0.1251	0.0396	0.2556
UniNe-ZHAW	Price_INEXSBS2016_SUM_SCORE	0.0542	0.1114	0.0386	0.2556
ISMD	ISMD16titlewithoutreranking	0.0531	0.1329	0.0355	0.1933
LSIS	Run1_ExeOrNarrativeNSW_Collection	0.0450	0.1166	0.0251	0.2050
ISMD	similaritytitlefieldreranked	0.0445	0.0966	0.0307	0.1933
CYUT	0.95RatingType2TGR	0.0392	0.1363	0.0145	0.1089
CYUT	0.95Ratingword2vecType2TGR	0.0373	0.1265	0.0136	0.1055
LSIS	Run2_ExeOrNarrativeNSW_UserProfile	0.0239	0.1018	0.0144	0.1742
OAU	oauc_reranked_ownQueryModel	0.0228	0.0766	0.0127	0.1265
OAU	oauc_basic	0.0217	0.0778	0.0118	0.1265
LSIS	Run3_ExeOrNarrativeNSW_Collection_AddData	0.0177	0.0533	0.0101	0.2050
LSIS	Run4_ExeOrNarrativeNSW_UserProfile_AddData	0.0152	0.0566	0.0079	0.1742
ISMD	ISMD16groupfield	0.0104	0.0527	0.0069	0.0564
know	sbs16suggestiontopicsresult2	0.0058	0.0227	0.0010	0.0013
OAU	oauc_reranked_attachedWorksModel	0.0044	0.0081	0.0021	0.1265
know	sbs16suggestiontopicsresult1	0.0018	0.0084	0.0004	0.0004

4. *UvA-ILLC - base_es* (rank 17): This run is based on a full-text elasticsearch (Elasticsearch) index of the A/LT collection, where the Dewey Decimal Codes are replaced by their textual representation. Default retrieval parameters are used, the query is a combination of the topic *title*, *group* and *request* fields. This is the same index that is used for the experimental system of the Interactive Track and serves as a baseline for the Suggestion Track.
5. *MRIM - RUN2* (rank 18): This run is a weighted linear fusion of a BM25F run on all fields, an Language Model (LM) run on all fields, and two query expansion runs, based on the BM25 and LM run respectively, using as expansion terms an intersection of terms in the user profiles and word embeddings from the query terms.

Most of the top performing systems, including the top performing run pre-process the rich topic statement with the aim of reducing the *request* to a set of most relevant terms. Two of the top five teams use the user profiles to modify the topic statement. This is the first year that word embeddings are used for the Suggestion Track and not without success. Both CYUT-CSIE and MRIM found that word embeddings improved performance over configurations without them. From these results it seems clear that topic representation is an important aspect in social book search. The longer narrative of the *request* field as well as the metadata in the user profiles and example books contain important information regarding the information need, but many terms are noisy, so a filtering step is essential to focus on the user's specific needs.

6   Conclusions and Plans

This was the second year of the SBS Suggestion Track as part of the SBS Lab. The overall goal remains to investigate the relative value of professional metadata, user-generated content and user profiles, but the specific focus for this year is to construct a test collection to evaluate systems dealing with complex book search requests that combine an information need expressed in a natural language statement and through example books.

We kept the evaluation procedure the same. Like last year, we added annotated example books with each topic statement, so that participants can investigate the value of query-by-example techniques in combination with more traditional text-based queries. Whereas in 2015 we only provided the book ID of example books in the topic statement, this year we also provided book title and author name.

The evaluation has shown that the most effective systems either adopt a learning-to-rank approach or incorporate keywords from the example books in the textual query. The effectiveness of learning-to-rank approaches suggests the complexity of dealing with multiple sources of evidence—book descriptions by multiple authors, differing in nature from controlled vocabulary descriptors, free-text tags and full-text reviews and information needs and interests represented by both natural language statements and user profiles—requires optimizing parameters through observing users' interactions.

Next year, we continue this focus on complex topics with example books and consider including an recommender systems type evaluation. We are also thinking of a pilot task in which the system not only has to retrieve relevant and recommendable books, but also to select which part of the book description—e.g. a certain set of reviews or tags—is most useful to show to the user, given her information need.

Bibliography

T. Beckers, N. Fuhr, N. Pharo, R. Nordlie, and K. N. Fachry. Overview and results of the inex 2009 interactive track. In M. Lalmas, J. M. Jose, A. Rauber, F. Sebastiani, and I. Frommholz, editors, *ECDL*, volume 6273 of *Lecture Notes in Computer Science*, pages 409–412. Springer, 2010. ISBN 978-3-642-15463-8.

Elasticsearch. Elasticsearch, version 2.1 (2015).

M. Koolen, J. Kamps, and G. Kazai. Social Book Search: The Impact of Professional and User-Generated Content on Book Suggestions. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM 2012)*. ACM, 2012.

M. Koolen, T. Bogers, A. van den Bosch, and J. Kamps. Looking for books in social media: An analysis of complex search requests. In A. Hanbury, G. Kazai, A. Rauber, and N. Fuhr, editors, *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29 - April 2, 2015. Proceedings*, volume 9022 of *Lecture Notes in Computer Science*, pages 184–196, 2015. doi: 10.1007/978-3-319-16354-3_19. URL http://dx.doi.org/10.1007/978-3-319-16354-3_19.

K. Reuter. Assessing aesthetic relevance: Children's book selection in a digital library. *JASIST*, 58(12):1745–1763, 2007.