Overview of the INEX 2007 Book Search Track (BookSearch'07)

Gabriella Kazai[1] and Antoine Doucet[2]
[1]Microsoft Research Cambridge, United Kingdom
gabkaz@microsoft.com
[2]University of Caen, France
doucet@info.unicaen.fr

Abstract

This paper provides an overview of the newly launched Book Search Track at INEX 2007 (Book-Search'07), its participants, tasks, book corpus, test topics and relevance assessments, as well as some results.

1 Introduction

Libraries around the world and commercial companies like Amazon, Google and Microsoft are digitizing hundreds of thousands of books in an effort to enable online access to these collections. The Open Content Alliance (OCA)[1], a library initiative formed after Google announced its library book digitization project, has brought library digitization efforts into the public eye, even though libraries have been digitizing books for decades before that. However, unlike most library digitization projects of the past, which centered around preservation and involved the careful and individual selection of materials to be digitized, the recent mass-digitization efforts aim at the conversion of materials on an industrial scale with minimum human intervention [2].

The increasing availability of the full-text of digitized books on the Web and in digital libraries, both enables and prompts research into techniques that facilitate storage, access, presentation and use of the digitized content. Indeed, the unprecedented scale of the digitization efforts, the unique characteristics of the digitized material as well as the unexplored possibilities of user interactions make full-text book search an exciting area of research today.

Motivated by this need, the book search track was launched in 2007 as part of the INEX initiative. INEX was chosen as a suitable forum due to its roots in the evaluation of structured document retrieval (SDR) approaches and since searching for information in a collection of books can be seen as one of the natural application areas of SDR. For example, in focused retrieval a clear benefit to users is to gain direct access to parts of books (of potentially hundreds of pages) relevant to the information need.

The ultimate goal of the INEX book search track is to investigate book-specific relevance ranking strategies, UI issues and user behaviour, exploiting special features, such as back of book indexes provided by authors, and linking to associated metadata like catalogue information from libraries. However, searching over large collections of digitzed books comes with many new challenges that need to be addressed first. For example, proper infrastructure has to be developed to allow for the scalable storage, indexing and retrieval of the digitized content. In addition, the setting up of a new track requires identifying suitable usage scenarios and tasks, establishing an evaluation framework complete with relevance

[1]http://www.opencontentalliance.org/

criteria, judgement procedures and evaluation metrics, as well as the development of a support system infrastructure. In its first year, the track set to explore these issues with the aim to investigate the requirements for such an infrastructure.

This paper reports on the outcome of the BookSearch'07 track. It provides an overview of its participants, tasks, book corpus, test topics and relevance assessments, as well as some results and findings. Since, at the time of writing, the relevance assessments for one of the tasks (Page in Context) were still outstanding, the results for this task are not reported here.

This paper is organised as follows. Section 2 gives a brief summary of the participating organisations. In Section 3, we briefly describe the retrieval tasks defined at BookSearch'07. Section 4 details the book corpus, test topics, and relevance assessments. Section 5 presents the results of the evaluation. Finally, we close with a summary and plans for BookSearch'08.

2 Participating organisations

In response to the call for participation, issued in April 2007, 27 organisations registered for BookSearch'07. Throughout the year, however, a number of groups droped out and only about a third remained active by the end of the year. Most groups reported difficulties due insufficient resources, including lack of space to store the dataset or scalable approach to process it, as well as lack of time or human resources required to tackle the various tasks.

The 27 groups along with details of their participation are summarized in Table 1. As it can be seen, only 10 groups remained active throughout. 16 groups downloaded the book corpus, 7 groups contributed search topics, and only 2 groups managed to submit runs.

3 Retrieval Tasks

The track defined four tasks: 1) Book Retrieval, 2) Page in Context retrieval, 3) Classification and 4) User intent taxonomy building. A summary of these are given in the following sections. Further details and the various submission DTDs are available in the track's Tasks and Submission Guidelines [7].

3.1 Book Retrieval Task

The goal of this task was to investigate the impact of book specific features on the effectiveness of book search systems, where the unit of retrieval is the (complete) book. Users are thus assumed to be searching for (whole) books relevant to their information need that they may want to, e.g., purchase or borrow from a library.

Participants of this task were invited to submit pairs of runs with the following condition: one of the runs would be generated using generic IR techniques, while the other run would extend this technique by exploiting book-specific features (e.g. back-of-book index, citation statistics, library catalogue information, etc.) or specifically tuned algorithms. In both cases, the ranked list of books could contain a maximum of 1000 books estimated relevant to the given topic, ranked in order of estimated relevance to the query.

Participants were permitted to submit up to 3 pairs of runs.

3.2 Page in Context Task

This task was set up similarly to the ad hoc track's Relevant in Context task [3], but here the task is applied to a collection of digitized books with shallow structural markup. Accordingly, based on the

ID	Organisation	Status A/C/P	Corpus download	Topics created	Runs submitted
1	University of Kaiserslautern	C	Y	-	-
2	University of California, Berkeley	A	Y	-	4 BR
4	University of Granada	C	Y	-	-
5	Lexiclone Inc	P	-	-	-
9	Queensland University of Technology	A	Y	-	-
10	University of Otago	C	-	-	-
12	University of Strathclyde	C	-	-	-
14	Wuhan University, China	P	-	-	-
19	Indian Statistical Institute	C	Y	-	-
20	LAMSADE	P	-	-	-
22	Doshisha University	A	Y	1	-
23	Kyungpook National University	A	Y	1	-
25	Max-Planck-Institut für Informatik	P	Y	-	-
26	Dalian University of Technology	A	Y	5	-
28	University of Helsinki	A	Y	2	-
32	RMIT University	P	-	-	-
33	Information Engineering Lab, CSIRO	P	-	-	-
36	University of Amsterdam	A	Y	3	-
37	University of Waterloo	C	Y	-	-
40	Carnegie Mellon University	P	Y	-	-
42	LIP6	P	-	-	-
53	Ecoles des Mines de Saint-Etienne	P	-	-	-
54	Microsoft Research, Cambridge	A	Y	13	-
55	University of Tampere	A	Y	5	-
61	Hong Kong Uni. of Sci. and Tech.	P	-	-	-
68	University of Salford, UK	P	-	-	-
92	Cairo Microsoft Innovation Center	A	Y	-	6 BR, 7 PiC
	Total (27 organisations)	10/6/11	16	30	10 BR, 7 PiC

Table 1: Participating groups at BookSearch'07 (In the Status column, A stands for Active, C for Cancelled, and P for Passive; In the Runs column, BR stands for Book Retrieval task, and PiC for Page in Context task)

assumption of a focused informational request, the task of a book search system was to return the user a ranked list of books estimated relevant to the request, and then present within each book, a ranking of relevant non-overlapping XML elements, passages, or book pages. The difference from the Relevant in Context task is that book search systems were required to rank the relevant elements/passages inside the books (instead of returning sets of elements/passages). The challenge for existing INEX participants was to test the scalability of their XML IR approaches and the adaptability of their search engines to the new domain. This task, however, is, and has proven to be, rather ambitious for most of the participants. For example, the Wikipedia corpus used in the ad hoc track experiments totals only about 1GB, whereas the size of the BookSearch'07 corpus is around 210GB (see Section 4).

Participants were allowed to submit up to 10 runs. One automatic (title-only) and one manual run were compulsory. Additional manual runs were encouraged in order to help the construction of a reliable test collection. Each run could contain for each test topic a maximum of 1000 books estimated relevant to the topic, ordered by decreasing value of relevance. For each book, participants were asked to provide a ranked list of non-overlapping XML elements, passages, or book page results that were estimated relevant to the query, ordered by decreasing value of relevance. A minimum of 1 element/passage/page result per book was required. A submission could only contain one type of result, i.e., only book pages or only passages; alas, result types could not be mixed.

3.3 Classification Task

In this task, systems were tested on their ability to assign the correct classification labels from the Library of Congress (LoC) classification scheme to the books of the test corpus based only on information available from the full text of the books. The distributed corpus of about 42,000 books (see Section 4) served as the training corpus, where classification labels were available for 20,692 books out of the 39,176 that had an associated MARC record. The test corpus contained 2 sets of 1,000 books.

Participants were allowed to submit up to three runs per test set. Each run was required to contain all 1,000 books of the given test set. For each book, systems needed to return a ranked list of classification labels, with a minimum of one label.

The list of Library of Congress classification headings extracted from the MARC records of the 20,692 books was made available by organisers on the INEX web site.

3.4 User Intent Taxonomy Task

User intent is a critical component in the understanding of users' search behaviour. It defines what kinds of search tasks users engage in. In traditional information retrieval, a user's intent is assumed to be informational in nature: It is driven by the user's need for information in order to complete a task at hand. Observations of Web use resulted in further two categories: navigational and transactional [1]. It is clear that these can also be applied to the book domain. However, it is possible that there are additional classes of user intent which are specific to books. It may also be the case that user tasks and user behaviour in the book domain will have specific traits and characteristics that may, for example, depend on genre. What are the possible classes of user intent and user tasks and what properties they have is a research question that this task was set to explore.

The goal of this task was to derive a taxonomy of user intent with its associated properties and search tasks. The use of examples of (actual or hypothetic) information needs demonstrating each class of intent and task was encouraged. Such an investigation could extend to include both research and design questions and possible answers regarding how a given user behaviour might be supported by a search system and its user interface. For example, a user hoping to buy a book is likely to be more interested

in a price comparison feature, while an informational query will more likely benefit from a "find related books" feature.

Examples of questions that could be explored included: How is user intent dependent on book genre? What book specific features best support the different types of intent and tasks? How could intent be extracted from query logs? How should one design experiments to allow for the identification of user intent from system logs? What data would enable the prediction of intent in order to aid users? What user behaviour follows from them?

Participation in this task involved the submission of a research or opinion paper detailing the proposed taxonomy. Participants could choose to report findings from the analysis of collected user log data or provide recommendations for the design of user studies to help elicit such data.

4 Test Collection

4.1 Book corpus

The corpus was provided by Microsoft Live Search Books and the Internet Archive (for non-commercial purposes only). It consists of 42,049 digitized out-of-copyright books, and totals around 210Gb in size. The collection contains books from a wide range of genre and includes reference works as well as poetry books. Most of the corpus is made up of history books (mostly American history), biographies, literary studies, religious texts and teachings. There are also encyclopedias, essays, proceedings and novels.

The OCR content of the books is stored in djvu.xml format. 39,176 of the 42,049 books also have associated metadata files (*.mrc), which contain publication (author, title, etc.) and classification information in MAchine-Readable Cataloging (MARC) record format.

The basic XML structure of a book (djvu.xml) is as follows:

```
<DjVuXML>
 <BODY>
  <OBJECT data="file..." [...]>
   <PARAM name="PAGE" value="[...]">
   [...]
   <REGION>
    <PARAGRAPH>
     <LINE>
      <WORD coords="[...]"> Moby </WORD>
      <WORD coords="[...]"> Dick </WORD>
      <WORD coords="[...]"> Herman </WORD>
      <WORD coords="[...]"> Melville </WORD>
      [...]
     </LINE>
     [...]
    </PARAGRAPH>
   </REGION>
   [...]
  </OBJECT>
  [...]
 </BODY>
</DjVuXML>
```

An <OBJECT> element corresponds to a page in a digitized book. A page counter is embedded in the @value attribute of the <PARAM> element which has the @name="PAGE" attribute. The actual page numbers (as printed inside the book) can be found (not always) in the header or the footer part

of a page. Note, however, that headers/footers are not explicitly recognised in the OCR, i.e., the first paragraph on a page could be a header and the last one or more paragraphs could be part of a footer. Depending on the book, headers may include chapter titles and page numbers (although due to OCR error, the page number is not always present).

Inside a page, each paragraph is marked up. It should be noted that an actual paragraph that starts on one page and ends on the next would be marked up as two separate paragraphs within two page elements.

Each paragraph element consists of line elements, within which each word is marked up. Coordinates that correspond to the four points of a rectangle surrounding a word are given as attributes of word elements, and could be used to enable text highlighting.

No further structural markup is currently available, although some books have rich logical structure, including chapters, sections, table of contents, bibliography, back-of-book index, and so on.

4.2 Topics

The test topics in BookSearch'07 are representations of users' informational needs, i.e, where the user is assumed to search for information on a given subject. For this year, all topics were limited to deal with content only aspects (i.e., no structural query conditions). The structure of books, however, could still be used by search engines to improve their ranking of books or book parts estimated relevant to a query.

Two sets of topics were used: 1) a set of 250 queries extracted from the query log of Live Search Books was used for the Book Retrieval task; 2) a set of 30 topics was created by the participating organisations for the Page in Context task. The next sections detail the topic format, the topic creation process for the Page in Context task, and provide a summary of the collected topics.

4.2.1 Topic Format.

Topics are made up of three parts, each of which describe the same information need, but for different purposes and at different level of detail:

<title> represents the search query that is to be used by systems. It serves as a short summary of the content of the user's information need.

<description> is a natural language definition of the information need.

<narrative> is a detailed explanation of the information need and a description of what makes an element/passage relevant or irrelevant. The narrative must be a clear and precise description of the information need in order to unambiguously determine whether or not a given text fragment in a book fulfills the need. The narrative is taken as the only true and accurate interpretation of the user's needs. Relevance assessments are made on compliance to the narrative alone.

Precise recording of the narrative is also important for scientific repeatability. To aid this, the narrative should explain not only what information is being sought, but also the context and motivation of the information need, i.e., why the information is being sought and what work-task it might help to solve. The narrative thus has the following two parts:

<task> is a description of the task for which information is sought, specifying the context, background and motivation for the information need.

<infneed> is a detailed explanation of what information is sought and what is considered relevant or irrelevant.

An example topic is given in Figure 1.

```
<title> Octavius Antony Cleopatra conflict </title>
<description> I am looking for information on the conflict between
    Octavius, Antony and Cleopatra. </description>
<narrative>
 <task> I am writing an essay on the relationship of Antony and Cleopatra
    and currently working on a chapter that explores the conflict between
    Octavius (the brother of Antony's wife, Octavia) and the lovers.
 </task>
 <infneed> Of interest is any information that details what motivated the
    conflict, how it developed and evolved through events such as the
    ceremony known as the Donations of Alexandria, Octavious' propaganda
    campaign in Rome against Antony, Antony's divorce from Octavia, and
    the battle of Actium in 31BC. Any information on the actions and
    emotions of the lovers during this period is relevant. Any
    non-documentary or non-biographical information, such as theatre plays
    (e.g., Shakespeare's play) or their critics are not relevant.
 </infneed>
</narrative>
```

Figure 1: Example topic (not part of the BookSearch'07 test set).

4.3 Topic Creation and Collected Topics

4.3.1 Topics for the Book Retrieval Task.

250 queries were extracted from the query logs of Live Search Books for which the test corpus contained at least one relevant book. No additional background or context information was available for these queries. Therefore these topics only have topic titles; and both the description and the narrative fields are left empty.

On average, a query contained 2.188 words, the longest query being 6 words long. The distribution of queries by length (in number of words) is shown in Figure 2.

4.3.2 Topics for the Page in Context Task.

Participants were asked to submit candidate topics for which at least 3 but no more than 20 relevant books were found during the collection exploration stage [8]. Participants were provided with a collection exploration tool to assist them in their topic creation. A screenshot of the tool is given in Figure 3.

This tool gave participants the means to search and explore the book corpus. This was achieved by building an interface to the search service provided by Live Search Books[2]. The tool took advantage of the fact that all books in the BookSearch'07 collection are indexed by Live Search Books. It worked by first sending the query entered by the user to the the Live Search Books search engine, and filtering the result list so that only books of the BookSearch'07 corpus were shown. Clicking on any of the books in the ranked list, took the user directly to the book viewer of Live Search Books (see Figure 3). Using this tool, participants could familiarize themselves with the collection and determine whether a candidate topic met the necessary criteria to be considered as a test topic: creation of topics with too few or too many relevant answers had to be aborted as these were deemed unsuitable for testing system performance [8]).

A total of 30 topics were created, complete with topic title, description and narrative (as described in Section 4.2.1). Table 1 shows the number of topics each participating group contributed.

[2] http://books.live.com, or http://search.live.com/books#q=\&mkt=en-US

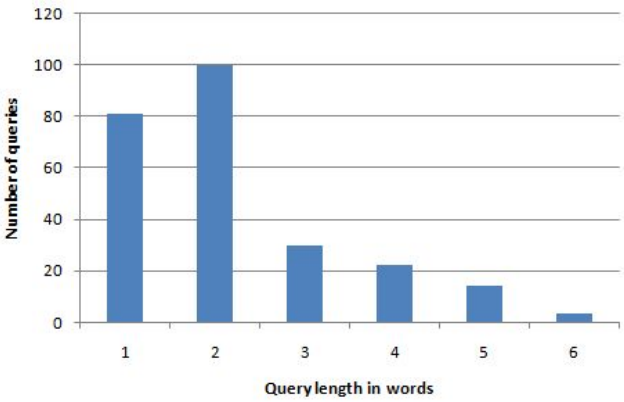

Figure 2: Distribution of queries by length (in number of words).

Participants were allowed to create topics based around the queries used for the Book Retrieval task. A query thus served as a starting point for participants to build up an information need during the collection exploration phase. Based on the similarity between the topic title of the created topic and the original query, we can distinguish the following categories: full, partial and no match. 10 topics belong in the full match category, meaning that the created topic title is exactly the same as the original query. 9 topics have partial matches, where participants refined the focus of the query, usually narrowing its scope. The remaining 11 topics were created from scratch.

4.4 Relevance Assessments

4.4.1 Relevance Assessments for the Book Retrieval Task.

The relevance assessments for the 250 queries used for this task were collected by Live Search Books from human judges. Judges were presented with a query and a set of books to judge. Assessment were made along a four point scale: Excellent, Good, Fair, and Non-relevant.

In total, 3,918 relevant books are contained in the assessment set. These include 1061 Excellent, 1,655 Good, and 1,202 Fair judgments. The average number of relevant books per query is 15.672, and the maximum is 41. The distribution of number of relevant books per topic is shown in Figure 4.

4.4.2 Relevance Assessments for the Page in Context Task.

Relevance judgments for the Page in Context task are to be provided by the participants of the track. At the time of writing, the relevance assessment system was still being finalised and thus the collection of judgments has not yet started.

The assessment system was implemented by adapting XRai [9] used at the ad hoc track and integrating it with the book viewer of Live Search Books. Figure 5 shows a screenshot of the assessment system. For

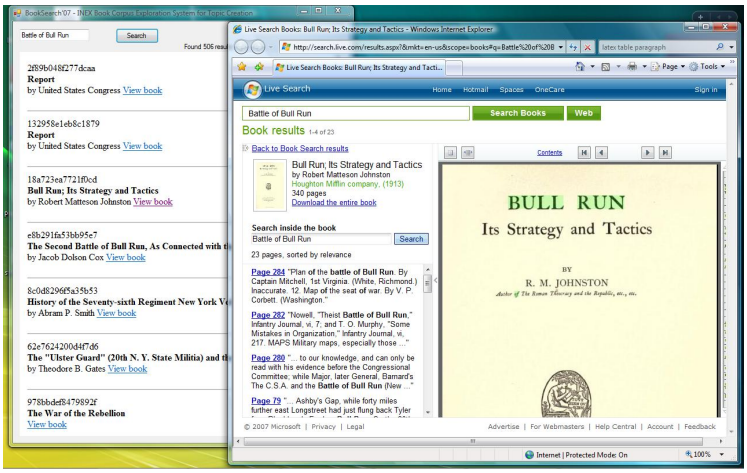

Figure 3: Screenshot of the system used to assist topic creation at BookSearch'07. The window on the left shows the main search window of the system. It allowed participants to enter a query and view the ranked list of books. The screen on the right is the book viewer window of the Live Search Books service, which allowed participants to explore the contents of a book for relevant information.

a given topic, the list of books to be assessed forms the assessment pool. On accessing a book, the book is opened in the Live Search Books viewer while the list of pages inside the book that are to be judged are displayed as hyperlinks on the left hand side of the browser window. Clicking on a page link displays the corresponding page in the book viewer. To ease the assessment process, judges can assess groups of pages using the "select all", "select none", and "select range" options. The latter offers a simple syntax for selecting a number of pages with little user effort (e.g., "18-36; 49" selects all pages between 18 and 36 (inclusive), and page 49). For each book, assessors will be required to judge all pages in the assessment pool and will be encouraged to explore additional pages in the book. The location of additional, possibly relevant pages is supported through the relevance bar feature of the Live Search Books scrollbar, which highlights pages where the query terms occur.

5 Submissions and Evaluation Results

As shown in Table 1, only two groups submitted retrieval runs to the Book Retrieval task and only one group to the Page in Context task. The University of California, Berkeley has also participated in the Classification task, but only evaluated its results unofficially. Furthermore, experiments on Book Retrieval were also conducted by participants at Microsoft Research Cambridge, the results of which were published in [12]. No submissions were received for the User Intent Taxonomy task. This was a bit of a surprise as our intent with this task was to attempt to open up the track allowing participation

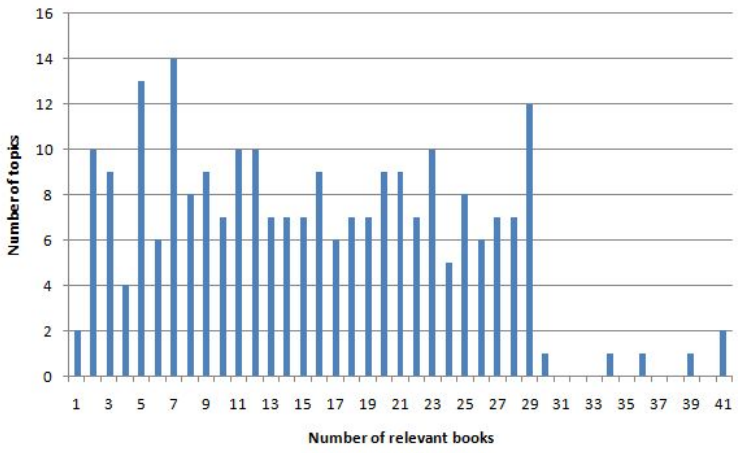

Figure 4: Distribution of relevant books per query.

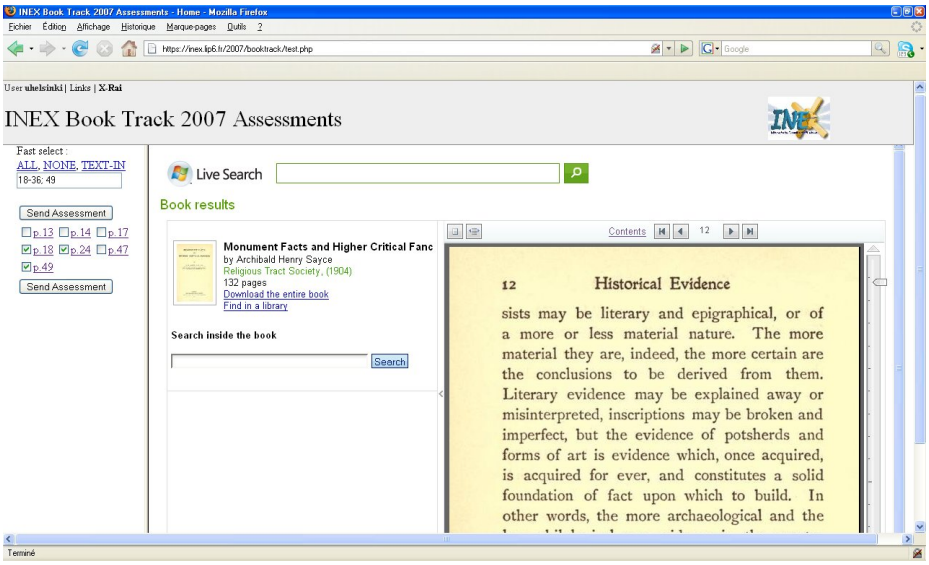

Figure 5: Screenshot of the Page in Context assessment system.

ID	RunID	Query A/M	Method N/B	NDCG @1	NDCG @5	NDCG @10	NDCG @25	NDCG @100	NDCG @1000
2	BERK_T2_OBJ **	A	N	0.351	0.316	0.316	0.349	0.427	0.478
2	BERK_T2_OBJ2 *	A	N	0.351	0.316	0.316	0.349	0.427	0.478
2	BERK_MARC_T2FB **	A	B	0.446	0.349	0.334	0.343	0.394	0.439
2	BERK_T2_CC_MARC *	A	B	0.453	0.375	0.359	0.371	0.422	0.462
92	Indri-F-C-A	A	N	0.521	0.477	0.479	0.503	0.562	0.604
92	Indri-NF-C-A	A	N	0.527	0.490	0.490	0.514	0.573	0.613
92	Indri-NF-PC-A	A	N	0.319	0.319	0.331	0.359	0.425	0.488
92	Indri-F-TOC-A	A	B	0.241	0.218	0.206	0.219	0.262	0.305
92	Indri-NF-TOC-A	A	B	0.257	0.233	0.225	0.235	0.275	0.316
92	Indri-NF-H-A	A	B	0.511	0.421	0.398	0.399	0.440	0.478

Table 2: Performance scores for the Book Retrieval task (In the Query column, A stands for Automatic, and M is for Manual; in the Method column, N stands for Non-book-specific retrieval approach, and B for Book-specific approach). Paired runs are indicated by * (where the information was available).

without a fully working book search engine.

To evaluate runs submitted to the Book Retrieval task, we adopted the Normalized Discounted Cumulated Gain measure of [6], using the discount function of $1/\log_b(i)$, where $b = 2$, and where the discounted cumulated gain for a given ranking is calculated as:

$$DCG[i] = \begin{cases} \sum_i G[i] & \text{if } i \leq b, \\ DCG[i-1] + \frac{\sum_i G[i]}{\log_b(i)} & \text{if } i > b. \end{cases} \quad (1)$$

The normalized DCG scores were obtained by dividing the DCG vector of the system's ranking by the DCG vector of the ideal ranking. The gain associated with a relevant book was 3 for a book rated Excellent, 2 for a book judged Good, and 1 for a Fairly relevant book. Irrelevant and unjudged books gave 0 gain.

Table 2 shows the NDCG scores, reported at various rank cutoffs. From the University of California, Berkeley results, it appears that performance at top ranks (up to rank 25) is improved using book-specific features or ranking methods. However, the results of Cairo Microsoft Innovation Center show that superior performance is achieved by simply applying traditional document retrieval techniques. Overall, we can observe a large variation in the performance scores, from 0.206 to 0.613. Interestingly, the runs `Indri-F-TOC-A` and `Indri-NF-TOC-A`, which only relied on pages that contained the table of contents or the back-of-book index performed the worst of all runs. On the other hand, the best performance was produced by the simplest strategy in `Indri-NF-C-A`, using a standard document retrieval framework. These findings highlight the need for further study in order to better understand the utility of book-specific retrieval features and suggest that there is plenty of room for further development. For details on the approaches of the two groups, please refer to [10] and [11].

6 Summary of BookSearch'07 and Plans for BookSearch'08

The Book Search track in 2007 focused on investigating infrastructure issues that come with the setting up of a new track. A range of tasks were defined: some of them extending established focused retrieval tasks studied at INEX into the book domain and some novel, book-specific tasks. The tasks were designed with the aim to provide new challenges for participants with existing search engines, as well as to attract new

groups with an interest in digitized books. Although most of these tasks proved to be rather ambitious, they represent a significant step in the shaping of a research agenda for the future of book search. The level of interest (27 registered groups) suggests that book search is an area that is set to grow considerably in the coming years, especially as more and more groups will be able to muster the necessary resources to tackle the range of challenges.

The Book Search track in 2008 (BookSearch'08)[3] will aim to look beyond the topic of search and extend to issues that touch a wider research community. BookSearch'08 will aim to bring together researchers in Information Retrieval, Digital Libraries, Human Computer Interaction, and eBooks with the goal to work on a common research agenda around digitized book collections. Towards this goal, the track will investigate the following topics:

- Users' interactions with e-books and collections of digitized books
- IR techniques for searching full texts of digitized books
- Digital library services to increase accessibility of digitized books

We plan to propose five tasks for 2008 and invite participants in the setting up process. The tasks are 1) Structure extraction from digitized books, 2) Creation of virtual bookshelves, 3) Supporting active reading, and the 4) Book retrieval, and 5) Page in Context tasks from BookSearch'07. The different tasks will make use of different sets of digitized books, ranging from a collection of 100 books to 50,000.

The track is set to start in mid April 2008.

Acknowledgements

We thank Steven Robertson, Nick Craswell and Natasa Milic-Frayling for their valuable comments on aspects of the organisation of this track.

References

[1] Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

[2] K. Coyle. Mass digitization of books. *Journal of Academic Librarianship*, 32(6):641–645, 2006.

[3] N. Fuhr, J. Kamps, M. Lalmas, S. Malik, and A. Trotman. Overview of the INEX 2007 ad hoc track. In Fuhr et al. [4].

[4] Norbert Fuhr, Jaap Kamps, Mounia Lalmas, and Andrew Trotman, editors. *Focused access to XML documents, 6th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX), Revised Selected Papers*, Lecture Notes in Computer Science. Springer, 2008.

[5] Norbert Fuhr, Mounia Lalmas, and Andrew Trotman, editors. *Pre-proceedings of the 6th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2007)*, http://inex.is.informatik.uni-duisburg.de/2007/inex07/pdf/2007-preproceedings.pdf, 2007.

[6] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (ACM TOIS)*, 20(4):422–446, 2002.

[3] http://www.inex.otago.ac.nz/tracks/books/books.asp

[7] Gabriella Kazai. Tasks and submission guidelines for the INEX 2007 book search track. In Fuhr et al. [5], pages 473–480.

[8] Gabriella Kazai. Topic development guidelines for the INEX 2007 book search track. In Fuhr et al. [5], pages 464–472.

[9] Mounia Lalmas and Benjamin Piwowarski. INEX 2007 relevance assessment guide. In Fuhr et al. [4].

[10] R. R. Larson. Logistic regression and EVIs for XML books and the heterogeneous track. In Fuhr et al. [4].

[11] W. Magdy and K. Darwish. CMIC at INEX 2007: Book search track. In Fuhr et al. [4].

[12] Hengzhi Wu, Gabriella Kazai, and Michael Taylor. Book search experiments: Investigating IR methods for the indexing and retrieval of books. In *Advances in Information Retrieval. Proceedings of the 30th European Conference on Information Retrieval*, volume 4956 of *Lecture Notes in Computer Science*, pages 234–245. Springer, 2008.