

# Overview of the INEX 2014 Interactive Social Book Search Track

Mark Hall<sup>1</sup>, Hugo Huurdeman<sup>2</sup>, Marijn Koolen<sup>2</sup>, Mette Skov<sup>3</sup>, and David Walsh<sup>1</sup>

<sup>1</sup> Department of Computing, Edge Hill University  
Ormskirk L39 4QP, United Kingdom

{mark.hall,david.walsh}@edgehill.ac.uk

<sup>2</sup> University of Amsterdam, Netherlands

{h.c.huurdeman,marijn.koolen}@uva.nl

<sup>3</sup> Department of Communication and Psychology  
Aalborg University, Denmark

skov@hum.aau.dk

**Abstract.** Users looking for books online are confronted with both professional meta-data and user-generated content. The goal of the Interactive Social Book Search Track was to investigate how users used these two sources of information, when looking for books in a leisure context. To this end participants recruited by four teams performed two different tasks using one of two book-search interfaces. Additionally one of the two interfaces also investigated whether user performance can be improved by providing a user-interface that supports multiple search stages.

## 1 Introduction

The goal of the Interactive Social Book Search (SBS) task is to investigate how book searchers use professional metadata and user-generated content at different stages of the search process. The purpose of this task is to gauge user interaction and user experience in social book search by observing user activity with a large collection of rich book descriptions under controlled and simulated conditions, aiming for as much "real-life" experiences intruding into the experimentation. The output will be a rich data set that includes both user profiles, selected individual differences (such as a motivation to explore), a log of user interactivity, and a structured set of questions about the experience.

The Interactive Social Book Search (ISBS) Task is a merge of the INEX Social Book Search (SBS, (??)) track and the Interactive task of CHiC (??). The SBS Track started in 2011 and has focused on system-oriented evaluation of book search systems that use both professional metadata and user-generated content. Out of three years of SBS evaluation arose a need to understand how users interact with these different types of book descriptions and how systems could support user to express and adapt their information needs during the search process.

The CHiC Interactive task focused on interaction of users browsing and searching in the Europeana collection. One of the questions is what types of metadata searchers use to determine relevance and interest. The collection, use case and task were deemed not interesting and useful enough to users.

The SBS contributes a new document collection, use case and search tasks. The first year of the ISBS will therefore focus on switching to the SBS collection and use case, with as few other changes as possible.

The goal of the interactive book search task is to investigate how searchers interact with book search systems that offer different types of book metadata.

The addition of opinionated descriptions and user-supplied tags allows users to search and select books with new criteria. User reviews may reveal information about plot, themes, characters, writing style, text density, comprehensiveness and other aspects that are not described by professional metadata.

In particular, the focus is on complex goal-oriented tasks as well as non-goal oriented tasks. For traditional tasks such as known-item search, there are effective search systems based on access points via formal metadata (i.e. book title, author name, publisher, year, etc). But even here user reviews and tags may prove to have an important role.

The long-term goal of the task is investigate user behaviour through a range of user tasks and interfaces and to identify the role of different types of metadata for different stages in the book search process. Research Questions

For the Interactive task, the main research question is:

**RQ** : How do searchers use professional metadata and user-generated content in book search?

This can be broken down into a few more specific questions:

**RQ1** How should the UI combine professional and user-generated information?

**RQ2** How should the UI adapt itself as the user progresses through their search task?

In this paper, we report on the setup and the results of the ISBS track. First, in Section 2 lists the participating teams. The experimental setup of the task is discussed in detail in Section 3 and the results in Section 4. We close in Section 5 with a summary and plans for 2015.

## 2 Participating Teams

In this section we provide information on the participating teams. In Table 1 we show which institutes participated in this Track and the number of users that took part in their experiments.

## 3 Experimental Setup

### 3.1 Social Book Search

The goal of the interactive Social Book Search (iSBS) track is to investigate how searchers make use of and appreciate professional metadata and user-generated

**Table 1.** Overview of the participating teams and number of users per team

| Institute | # users |
|-----------|---------|
| Aalborg   | 7       |
| Amsterdam | 7       |
| Edge Hill | 10      |
| Humboldt  | 17      |
| Total     | 41      |

content for book search on the Web and to develop interfaces that support searchers through the various stages of their search task. The user has a specific information need against a background of personal tastes, interests and previously seen books. Through social media, book descriptions are extended far beyond what is traditionally stored in professional catalogues. Not only are books described in the users' own vocabulary, but they 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 (?), 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 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 pure search (?) or pure recommendation.

The iSBS track investigates book requests and suggestions from the LibraryThing (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 (?), which contains 1.5 million book descriptions from Amazon, enriched with content from LT. This collection contains both professional metadata and user-generated content.<sup>4</sup>

The records contain title information as well as a Dewey Decimal Classification (DDC) code (for 61% of the books) and category and subject information supplied by Amazon. We note that for a sample of Amazon records the subject descriptors are noisy, with a number of inappropriately assigned descriptors that seem unrelated to the books. Each book is identified by an ISBN. Since different editions of the same work have different ISBNs, there can be multiple records for

<sup>4</sup> This collection is a subset of a larger collection of 2.8 million description. The subset contains all book description that have a cover image.

a single intellectual work. Each book record is an XML file with 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.

### 3.2 User Tasks

Two tasks were created to investigate the impact of different task types on the participants interactions with the interfaces and also the professional and user-generated book meta-data.

**The *goal-oriented* task** was developed as a “simulated leisure task” (?), with the topic derived from the LibraryThing collection. The LibraryThing collection contains discussion fora in which users asked other users for advice on which books to read for a given topic, question, or area of interest. From this list of discussion topics, a discussion on “layman books for physics and mathematics” was selected as the book collection contained a significant number of books on the topic, it is a neutral topic, it provides guidance, but it is also sufficiently flexible that participants can interpret it as needed. The following instruction text was provided to participants:

Imagine you are looking for some interesting physics and mathematics books for a layperson. You have heard about the Feynman books but you have never really read anything in this area. You would also like to find an “interesting facts” sort of book on mathematics

**The *non-goal* task** was developed based on the open-ended task used in the iChIC task at CLEF 2013 (?). The aim of this task is to investigate how users interact with the system when they have no pre-defined goal in a more exploratory search context. It also allows the participants to bring their own goals or sub-tasks to the experiment in line with the “simulated work task” ideas (?). The following instruction text was provided to participants:

Imagine you are waiting to meet a friend in a coffee shop or pub or the airport or your office. While waiting, you come across this website and explore it looking for any book that you find interesting, or engaging or relevant...

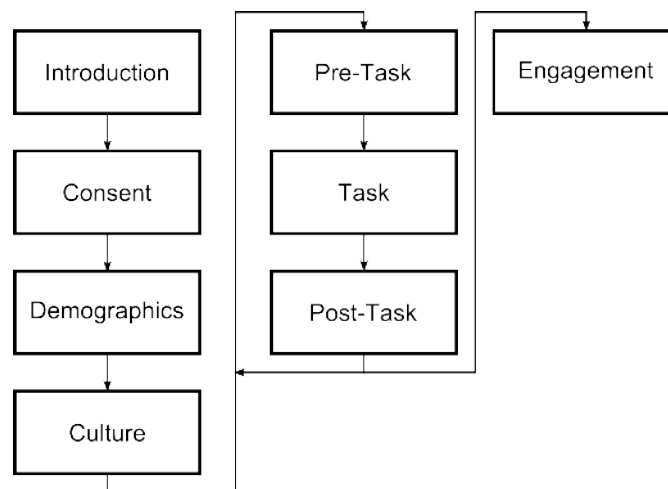
### 3.3 Questionnaires

The experiment was conducted using the SPIRE system<sup>5</sup> (?), which uses a set of question pages to acquire the following information:

<sup>5</sup> Based on the Experiment Support System – <https://bitbucket.org/mhall/experiment-support-system>

- *Consent* – all participants had to confirm that they understood the tasks they would be asked to undertake and the types of data collected in the experiment. Participants also specified who had recruited them;
- *Demographics* – the following factors were acquired in order to characterise the participants: gender, age, achieved education level, current education level, and employment status;
- *Culture* – to quantify language and cultural influences, the following factors were collected: country of birth, country of residence, mother tongue, primary language spoken at home, languages used to search the web;
- *Post-Task* – in the post task questions, participants were asked to judge how useful each of the interface components and meta-data parts that they had used in the task were, using 5-point Likert-like scales;
- *Engagement* – after participants had completed both tasks, they were asked to complete O’Brien et al.’s (?) engagement scale.

Figure 1 shows the path participants took through the 8 steps of the experiment. The SPIRE system automatically rotated the task order to avoid ordering biases.



**Fig. 1.** The path participants took through the experiment. Each participant completed the *Pre-Task*, *Task*, *Post-Task* twice (once for each of the tasks). The SPIRE system automatically balanced the task order. No data was acquired in the *Introduction* and *Pre-Task* steps.

### 3.4 System and Interfaces

The two tested interfaces (*baseline* and *multi-stage*) were both built using the PyIRE<sup>6</sup> workbench, which provides the required functionality for creating interactive IR interfaces and logging all interactions between the participants and the system. This includes any queries they enter, the books shown for the queries, pagination, facets selected, books viewed in detail, meta-data facets viewed, books added to the book-bag, and books-removed from the book-bag. All log-data is automatically timestamped and linked to the participant and task.

Both interfaces used a shared IR backend implemented using ElasticSearch<sup>7</sup>, which provided free-text search, faceted search, and access to the individual books complete meta-data.

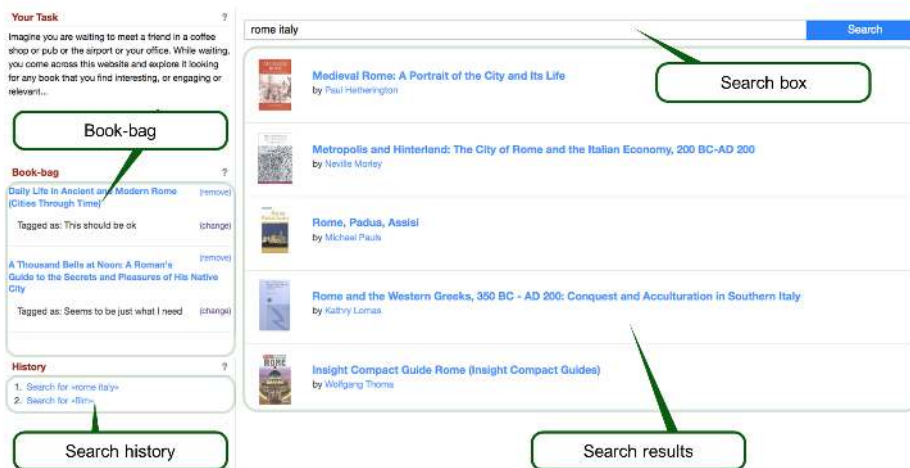


Fig. 2. *Baseline* interface – results view.

The *baseline* interface shown in figure 2 represents a standard web-search interface, with the left column containing the task instructions, book-bag, and search history and the main area showing the results. The book-bag resembles a shopping-cart system and contained those books participants collected to complete the task. Items added to the book bag could be viewed again and also be removed from the book bag if the participants felt that they did not need them after all. The history panel at the bottom recorded all searches undertaken, to

<sup>6</sup> Python interactive Information Retrieval Evaluation workbench – <https://bitbucket.org/mhall/pyire>

<sup>7</sup> ElasticSearch – <http://www.elasticsearch.org/>

enable easy re-running of queries. The main panel on the right contained the search box and a paginated list of search results. The result listings contained a thumbnail image, title, author(s), and aggregate review ratings, if there were any reviews for the book.

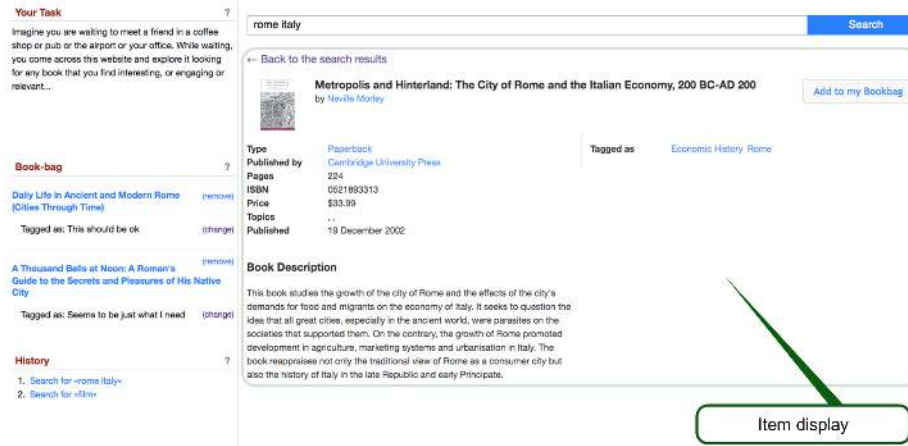


Fig. 3. *Baseline* interface – item view.

The item view shown in figure 3 was displayed once a book was selected from the search results. The item view contained two distinct sets of information. The left hand side was dedicated to professional meta data, including publisher, price, isbn, published date, number of pages and a detailed description, followed by a grid of similar books if any had been linked to the main book. The right-hand side contained user generated content such as tags and customer reviews, where each review consisted of a description and a 1-5 star rating.

**The *multi-stage* interface** aims to support users by taking the different stages of the search process into account. The idea behind the *multi-stage* interface design is supported by two theoretical components.

Firstly, several information search process models look at stages in the search process. A well-known example is ?, who discovered “common patterns in users’ experience” during task performance. She developed a model consisting of six stages, which describe users’ evolving thoughts, feelings and actions in the context of complex tasks. ? later summarized Kuhlthau’s stages into three categories (pre-focus, focus formulation, and post-focus), and points to the types of information searched for in the different stages. Building on ?, the proposed multi-stage search interface for ISBS includes three stages: *Explore*, *focus*, and

*refine*. Secondly, when designing a new search interface for social book search it has also been relevant to look more specifically at the process of choosing a book to read.

A model of decision stages in book selection (?) identifies the following decision stages: browse category, selecting, judging, sampling, and sustained reading. This work supports the need for a user interface that takes the different search and decision stages into account. However, the different stages in (?) closely relate to a specific full text digital library, and therefore the model was not applicable to the present collection.

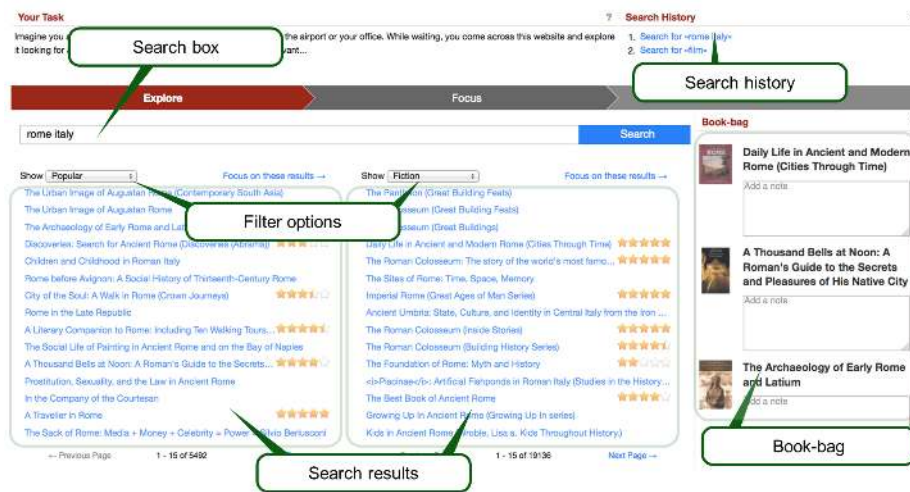


Fig. 4. Multistage interface – explore view.

The initial *explore* stage shown in Figure 4 aimed to support the initial exploration of the data-set and contains a very similar feature set to the baseline, including task instructions, search box, search results, book bag, and search history. The two main differences to the *baseline* interface were the navigation bar that allows the participants to switch between the stages and the dense, multi-column search results. The search results were shown in two columns with each book showing only a title and review rating. Each column showed a different faceted filter of the same query to allow for a wider overview of results to be explored. The "filter options" allowed the participant to select what filter to apply to each of the columns. Additionally the book-bag was extended to include a note field for each book, but this feature was not used heavily.

Figure 5 shows the popup interface which is displayed when a book is selected in the *explore* or *refine* stage's search results. All of the book meta-data is available via a tabbed structure, including the professional meta-data in the form



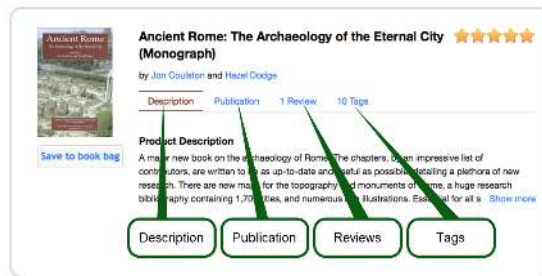


Fig. 5. *Multistage* interface – item popup view.

of the description and publication data as well as the user generated content of reviews and tags.

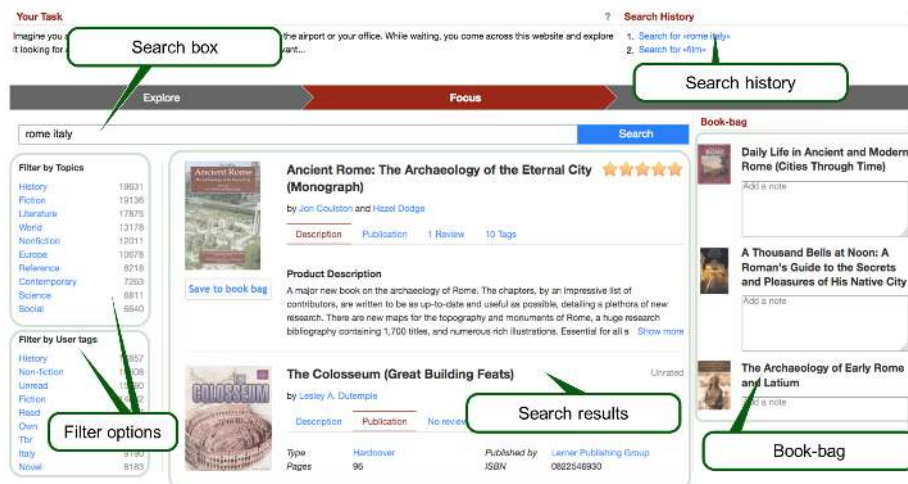


Fig. 6. *Multistage* interface – focus view.

The *focus* stage shown in Figure 6 supports in-depth searching and provides detailed search results that directly include the full meta-data that in the other stages is shown via a popup (Figure 5). A category filter was also provided in the left column which provided a means to reduce and refine the search results. In the *explore* stage, the participant could choose to focus on one of the two result columns, which would show the *refine* view with that column's filter pre-selected.

The *refine* stage shown in Figure 7 supports the refining of the final list of books the participants want to choose. It thus focuses on the books the user has

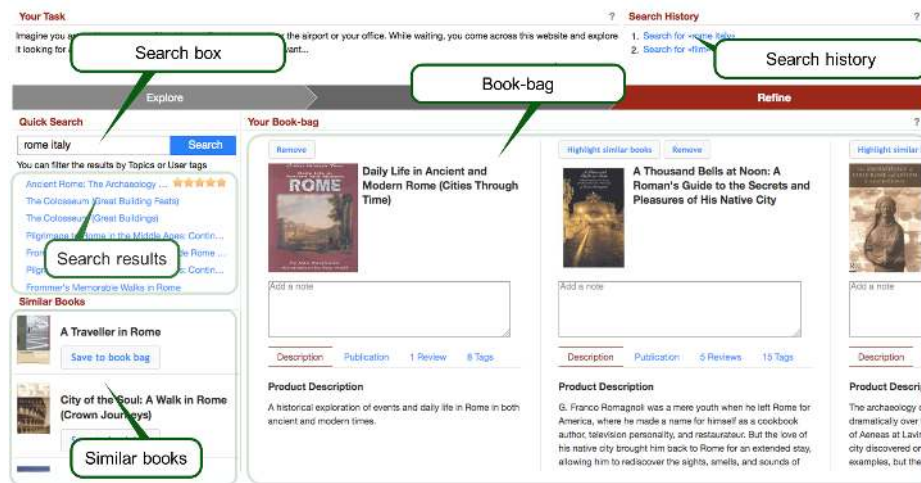


Fig. 7. *Multistage* interface – refine view.

already added to their book-bag and this stage cannot be entered until at least one book has been added to the book-bag.

The search feature is a lot less prevalent in this interface and is confined to the left column with minimal result details shown. The query, filters, and results are the same as on the focus stage. Directly underneath the results is the similar books panel, which shows all books that are defined as similar to any of the books in the book-bag, in order to support the participant in augmenting their list of collected books. When a book in the book-bag has similar books a button is shown above the book title which highlights that book's similar books in yellow and raises them to the top of the similar books list. The book-bag fills the rest of the interface and works as it does on the other pages but provides more detail per book.

### 3.5 Participants

A total of 41 participants were recruited (see Table 1), 27 female and 14 male. 16 were between 18 and 25, 21 between 26 and 35, 3 between 36 and 45, and 1 between 46 and 55. 9 were in employment and 32 were students. Participants came from 8 different countries including Germany, Denmark, UK, the Netherlands, Colombia, Brazil, Romania, and Suriname. Participants' mother tongues included German, Dutch, English, Danish, Romanian, Farsi or Portuguese. The majority of participants executed the tasks at a lab (29), only 12 users conducted the experiment remotely. 22 participants used the novel *multi-stage* interface, while 19 used the *baseline* interface.

### 3.6 Procedure

Participants were invited by the individual teams, either using e-mail (Aalborg, Amsterdam) or by recruiting students from a lecture or lab (Edge Hill, Humboldt). Where participants were invited by e-mail, the e-mail contained a link to the online experiment, which would open in the participant’s browser. Where participants were recruited in a lecture or lab, the experiment URL was distributed using e-learning platforms. All browsers and operating systems had been tested and worked, with the exception of Safari and Chrome on OS X, where there was a conflict caused by the security certificate, which was outside of our control.

After participants had completed the experiment as outlined above (3.3), they were provided with additional information on the tasks they had completed and with contact information, should they wish to learn more about the experiment. Where participants that completed the experiment in a lab, teams were able to conduct their own post-experiment process, which mostly focused on gathering additional feedback on the system from the participants.

## 4 Results

Based on the participant responses and log data we have aggregated summary statistics for a number of basic performance metrics.

**Session length** was measured automatically using JavaScript and stored with the participants’ responses. Table 2 shows median and inter-quartile ranges for all interface and task combinations. While the results seem to indicate that participants spent longer in the *baseline* interface and also longer on the *goal-oriented* task, the differences are not statistically significant (Wilcoxon signed-rank test). Interestingly, for the *non-goal* task, the median times are roughly similar to the session lengths in the iCHiC experiment that the task was taken from ?. This might indicate that that is the approximate time that participants can be expected to spend on any kind of open-ended leisure-task.

**Table 2.** Session lengths for the two interfaces and tasks. Times are in minutes:seconds and are reported median (inter-quartile range).

| Interface          | Goal-oriented     | Non-goal          |
|--------------------|-------------------|-------------------|
| <i>Baseline</i>    | 6:25min (3:42min) | 3:42min (3:45min) |
| <i>Multi-Stage</i> | 3:35min (4:24min) | 2:40min (6:21min) |

**Number of queries** was extracted from the log-data. In both interfaces it was possible to issue queries by typing keywords into the search box or by clicking on a meta-data field to search for other books with that meta-data field value. Both types of query have been aggregated and Table 3 shows the number of queries for each interface and task. The results are in line with the session length results, with participants executing slightly more queries in the *goal-oriented* task (Wilcoxon rank-sum test  $p < 0.05$ ). However, the interface did not have a significant impact on the number of queries executed.

**Table 3.** Number of queries executed. Numbers are reported median (inter-quartile range).

| <b>Interface</b>   | <b>Goal-oriented</b> | <b>Non-goal</b> |
|--------------------|----------------------|-----------------|
| <i>Baseline</i>    | 4 (5.5)              | 2 (4.5)         |
| <i>Multi-Stage</i> | 3 (2.75)             | 2 (3)           |

**Number of books viewed** was extracted from the log-data. Table 4 shows the results. Participants viewed fewer books in the *non-goal* task (Wilcoxon rank-sum test  $p < 0.05$ ), which was to be expected considering that they also executed less queries and spent less time on the task. As with the number of queries the number of books viewed is not significantly influenced by the interface participants used.

**Table 4.** Number of books viewed. Numbers are reported median (inter-quartile range).

| <b>Interface</b>   | <b>Goal-oriented</b> | <b>Non-goal</b> |
|--------------------|----------------------|-----------------|
| <i>Baseline</i>    | 4 (5.5)              | 2 (4.5)         |
| <i>Multi-Stage</i> | 3 (2.75)             | 2 (3)           |

**Number of books collected** was extracted from the log-data. Participants collected those books that they felt were of use to them. The numbers reported in Table 5 are based on the number of books participants had in their book-bag when they completed the session, not the total number of books collected over the course of their session, as participants could always remove books from their book-bag in the course of the session, as participants could always remove books from their book-bag in the course of the session.

Unlike the other metrics, where the interface had no significant influence on the metric, in the *non-goal* task, participants collected significantly more books

**Table 5.** Number of books collected. Numbers are reported median (inter-quartile range).

| Interface          | Goal-oriented | Non-goal |
|--------------------|---------------|----------|
| <i>Baseline</i>    | 3 (3)         | 1 (2)    |
| <i>Multi-Stage</i> | 3.5 (3)       | 2 (3)    |

using the *multi-stage* interface than with the *baseline* interface. Considering that there are no significant interface effects for the *non-goal* task in any of the other metrics and that there is no significant difference in the *goal-oriented* task, this strongly suggests that the *multi-stage* interface provides a benefit to open-ended leisure tasks, while at the same time working just as well as the *baseline* interface for more focused tasks.

## 5 Conclusions and Plans

This was the first year of the Interactive Social Book Search Track. Because of time constraints, the data-gathering period was short and only a small number of users participated in the study. However, their data provides valuable lessons for the future.

Plans for the next year are to make improvements to the interfaces based on user feedback and to run a full data-gathering phase to get user data from a much larger group of participants.

## 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.
- P. Borlund and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of documentation*, 53(3): 225–250, 1997.
- M. M. Hall and E. Toms. Building a common framework for iir evaluation. In *CLEF 2013 - Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pages 17–28, 2013. doi: 10.1007/978-3-642-40802-1\_3.
- M. M. Hall, R. Villa, S. A. Rutter, D. Bell, P. Clough, and E. G. Toms. Sheffield submission to the chic interactive task: Exploring digital cultural heritage. In *CLEF 2013 Evaluation Labs and Workshop*.
- 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, 2012a.

- M. Koolen, G. Kazai, J. Kamps, A. Doucet, and M. Landoni. Overview of the INEX 2011 books and social search track. In S. Geva, J. Kamps, and R. Schenkel, editors, *Focused Retrieval of Content and Structure: 10th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2011)*, volume 7424 of *LNCS*. Springer, 2012b.
- M. Koolen, G. Kazai, J. Kamps, M. Preminger, A. Doucet, and M. Landoni. Overview of the INEX 2012 social book search track. In S. Geva, J. Kamps, and R. Schenkel, editors, *Focused Access to Content, Structure and Context: 11th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX'12)*, LNCS. Springer, 2013a.
- M. Koolen, G. Kazai, M. Preminger, and A. Doucet. Overview of the INEX 2013 social book search track. In *CLEF 2013 Evaluation Labs and Workshop, Online Working Notes*, 2013b.
- C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5): 361–371, 1991. ISSN 1097-4571. doi: 10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#. URL [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<361::AID-ASI6>3.0.CO;2-#](http://dx.doi.org/10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#).
- H. L. O'Brien and E. G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, 2009.
- V. Petras, T. Bogers, E. Toms, M. Hall, J. Savoy, P. Malak, A. Pawowski, N. Ferro, and I. Masiero. Cultural heritage in clef (chic) 2013. In P. Forner, H. Mller, R. Paredes, P. Rosso, and B. Stein, editors, *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, volume 8138 of *Lecture Notes in Computer Science*, pages 192–211. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-40801-4. doi: 10.1007/978-3-642-40802-1\_23. URL [http://dx.doi.org/10.1007/978-3-642-40802-1\\_23](http://dx.doi.org/10.1007/978-3-642-40802-1_23).
- K. Reuter. Assessing aesthetic relevance: Children's book selection in a digital library. *JASIST*, 58(12):1745–1763, 2007.
- M. Skov and P. Ingwersen. Exploring information seeking behaviour in a digital museum context. In *Proceedings of the second international symposium on Information interaction in context*, pages 110–115. ACM, 2008.
- E. Toms and M. M. Hall. The chic interactive task (chici) at clef2013. <http://www.clef-initiative.eu/documents/71612/1713e643-27c3-4d76-9a6f-926cdb1db0f4>, 2013a.
- E. G. Toms and M. M. Hall. The CHiC Interactive Task (CHiCi) at CLEF2013. In *CLEF 2013 Evaluation Labs and Workshop, Online Working Notes*, 2013b.
- P. Vakkari. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation*, 57(1): 44–60, 2001.