Edge Hill Computing @ Interactive Social Book Search 2015

Daniel Campbell[1], Mark Hall[1], and David Walsh[1]

Edge Hill University, St Helens Road, Ormskirk, L39 4QP, United Kingdom,
{Daniel.Campbell, Mark.Hall, David.Walsh}@edgehill.ac.uk

Abstract. In our contribution to the CLEF2015 Interactive Social Book Search task we use log-analysis to investigate how users interacted with the system over their session. We investigated what participants' first interactions with the collection are, how they interact with the multi-stage interface, and how users interactions with the multi-stage interface change over the course of their session.

Keywords: human computer information retrieval, user study, log analysis

1 Introduction

The CLEF[1] Social Book Search lab's *Interactive* task gathered data from users using one of two interfaces to complete two tasks. The *baseline* interface implemented a standard Information Retrieval (IR) interface [1] consisting of a search box, a search result list, and an individual item display. The second interface (*multi-stage*) attempted an implementation of Kuhlthau's multi-stage search process[4], filtered through Vakkari's simplification of the model [6]. Two tasks were tested, the first an *non-goal* task where participants were asked to look for any book they might find interesting, and a *goal-oriented* task where participants were asked to find five books for a given topic ("survival on a desert island"). Each participant completed both tasks in one of the two interfaces. Task order was automatically balanced to avoid ordering bias.

In this paper we investigated the following three research questions:

2 Research Questions

RQ1 : What are users' first interactions with an unknown collection?
RQ2 : How do users interact with the multi-stage interface?
RQ3 : How do users interactions change over the course of the session?

[1] Conference and Labs of the Evaluation Forum

3 First Interactions

To answer RQ1 an investigation of how the participant interacts with a new system and data set via the search box was conducted focusing initially on the first query by each participant and then the first 3 queries by each participant.

To do this we process the log data, removing any session that lasted less than 20 seconds. The remaining data was then analysed for each interface / task combination.

3.1 First Query Word Counts

Investigating initial search queries, potentially provides an insight to how a user learns about an unknown data set/system. To analyse this the four data sets were filtered to only included rows of data where the action was 'start' and the first query action that followed.

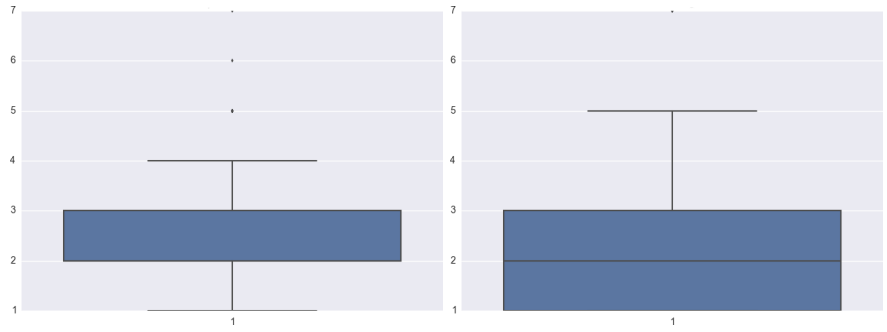
By simply processing the first queries and counting the number of words; there is an interesting pattern that emerges around the high volume of one and two word search terms in all tasks and interface combinations, and also that all but one of the long tail search terms (over 4 words) are in the *goal-oriented* tasks. Table 1 summarises the numbers of words per initial query per session. Figure 1 shows that for the *goal-oriented* tasks there is a difference in the distribution, with more one and two word search terms being used in the multistage interface.

Table 1. Number of words in the first query for all task / interface combinations

	Non-Goal Task				**Goal-Oriented Task**			
# Words	**Baseline**		**Multi-Stage**		**Baseline**		**Multi-Stage**	
	# queries	**%**	**# queries**	**%**	**# queries**	**%**	**# queries**	**%**
1	40	55%	29	55%	16	18%	33	36%
2	26	36%	16	30%	39	43%	29	32%
3	4	5%	7	13%	19	21%	14	15%
4	1	1%	1	2%	5	5%	6	7%
5	2	3%	0	0%	10	11%	6	7%
6	0	0%	0	0%	1	1%	0	0%
7	0	0%	0	0%	1	1%	3	3%
8	0	0%	0	0%	0	0%	0	0%

The interface does not seem to effect the interactions in the *non-goal* task (rank sum p value = 0.84), which is interesting as the participants provided with the Multi-stage interface are not presented straight away with the search box and so have the opportunity to explore using the faceted hierarchy before actively changing to the search view.

Table 2 shows that either or both of the interface or task are having an effect on the focus task. When comparing the two *goal-oriented* tasks there is a p-value of **0.037**. There are also clear differences when comparing the different



Baseline interface *Multi-stage* interface

Fig. 1. Distribution of number of words per query in the *Goal-oriented* task

types of task in the same interface. The *non-goal* vs *goal-oriented* tasks in the Baseline interface are the most significant with a p-value of **0.0000000118** which shows that the number of words being used per query in the *goal-oriented* task are significantly higher than in the *non-goal* task where the participant is not provided with any cues to search for.

Table 2. Average query length for all interface / task combinations (mean query length and standard deviation)

	Non-Goal Task		**Goal-oriented Task**		
	Words per Query		**Words per Query**		*p-value*
Baseline	1.61	(.87)	2.59	(1.32)	0.0000000118
Multi-Stage	1.62	(.78)	2.28	(1.46)	0.011
p-value		0.84		0.037	

The type of task however does appear to have an impact on the length of the initial query as search terms with 4 words or more account for less then 4% in either of the *non-goal* tasks and are 16% in the *goal-oriented* Multi-stage and 19% in the *goal-oriented* Baseline.

In the *non-goal* tasks the largest percentage of queries that were conducted as an initial search were single word queries, accounting for 55% (Table 1). In the *goal-oriented* task it is interesting to observe that there were more single word queries conducted in the multi-stage interface than the Baseline, and also that there were more two word queries conducted in the Baseline than the multi-stage.

This would suggest that presenting users directly into the facet hierarchy (explore view) has little to no effect on the user learning about the data set they are searching. Or it could show that for a *known item search* (which it appears the participants took the *goal-oriented* task for) participants prefer to search rather than browse a hierarchy.

3.2 Type of Things Participants Searched For

Using the same filtered log data, each initial query term was manually coded. The coding's were generalizations extrapolated from the search keywords themselves. Table 3 summarises the manually coded categorization of all first queries.

Table 3. Query classification for all task / interface combinations

	Non-Goal Task		Goal-oriented Task	
# Words	**Baseline**	**Multi-Stage**	**Baseline**	**Multi-Stage**
Place	**11**	**8**	**2**	**5**
Media Technology	1	0	0	0
Book title	3	2	**2**	2
Genre	6	1	0	0
Person	**13**	**7**	**2**	2
Hobby	4	1	1	1
Study Skill	1	0	**86**	**77**
Technical	2	0	0	0
History	5	1	0	0
Technology	2	3	0	1
Cartoon	1	1	0	0
Game	1	3	0	0
Sexuality	1	0	0	0
Tv Show	1	1	0	0
Nothing	3	1	0	0
Non-Sense	1	0	0	0
Other	1	**9**	0	2

The results show that there is no significant difference in the *non-goal* task across either interface. Table 3 shows that the Place and Person categories were consistently high ranking in all of the tasks and interfaces. This also indicates that the browse facets feature on the multistage interface has no/very little impact on initial query choice. For the *goal-oriented* task results show that 94.5% (172 of 182) of queries on the *goal-oriented* task were relevant to the first part of the task "Select one book about surviving on a desert island", nearly all were variants of the actual task e.g. "survive desert island".

This indicates a possibility that participants were using the task keywords as known item searches.

3.3 First Three Queries

To explore the findings from the initial query further, the data set was widened to include the first 3 queries to see if there were any beginnings of search patterns emerging. The coding strategy was expanded to all queries then the data was processed in the same way starting with the word counts.

The pattern of the amount of keywords in a query continues after widening the scope. The percentages of amounts of keywords remain almost the same as the first queries (Table 1) with one exception. The data shows that the *non-goal* task participants started to create more specific (long tail (over 4 words)) queries (2%). Table 4 summarises the numbers of words per query over the first 3 queries per session.

Table 4. Number of words per query calculated over the first three queries for all task / interface combinations

	Non-Goal Task				**Goal-oriented Task**			
# Words	**Baseline**		**Multi-Stage**		**Baseline**		**Multi-Stage**	
1	79	50%	59	46%	56	21%	85	34%
2	54	34%	40	31%	99	37%	84	34%
3	16	10%	14	11%	66	25%	45	18%
4	4	3%	5	4%	18	7%	15	6%
5	3	2%	1	1%	23	9%	10	4%
6	1	1%	1	1%	3	1%	2	1%
7	0	0%	0	0%	4	1%	7	3%
8	0	0%	0	0%	0	0%	1	0%

3.4 Keyword duplication

A qualitative analysis looked at the frequency of duplication of the search queries. For this study we classify that a duplicate query is when 2 participants have used the same term, not when a single participant searches multiple times for the same term.

In the *non-goal* tasks there were only 2 terms that appeared more than once (Table 5). It is important to note that there is no obvious linkage between the two search terms. But Berlin was part of the training task and it appears a couple of participants re-searched this term in the main tasks.

Table 5. Most frequent first queries in the *non-goal* task in the *baseline* interface

Term	**Occurrences**
Art	2
Berlin	2

The *non-goal* Multi-Stage had no duplicates at all on the initial query. However in the *goal-oriented* tasks there were 40 queries with duplicates in the baseline (Table 6) and 25 duplications in the multi-stage (Table 7).

Table 6. Most frequent first queries in the *goal-oriented* task in the *baseline* interface

Term	Occurrences
surviving on a desert island	9
desert island	8
survival	7
surviving desert island	6
desert island survival	4
surviving	3
survive island	3

Table 7. Most frequent first queries in the *goal-oriented* task in the *multi-stage* interface

Term	Occurrences
survival	11
surviving on a desert island	6
desert island	5
surviving	3

In the initial 3 queries for the *non-goal* tasks there were only 4 terms that appeared more than once (Tables 8 and 9).

Table 8. Most frequent first queries in the *non-goal* task in the *baseline* interface

Term	Occurrences
Art	2
Game of Thrones	2
Harry Potter	2
Berlin	2

The *goal-oriented* task shows a significantly different set of results with many terms being used by multiple participants. There is also clear relation with all of the terms to the task (Tables 10 and 11).

From the findings above there is a stronger indication that the participants were using parts of the instructions as not the intended Que but instead as actual queries (Known item searches). To investigate this apparent pattern further a brief manual study of the complete log data (all queries, actions in the sessions) was undertaken and whilst further study is needed there is a strong suggestion of a pattern emerging throughout the sessions where the participants use the task as a known item search list.

Whilst an amount of duplication is to be expected with the *goal-oriented* type of task, the numbers here show that most users certainly attempted to

Table 9. Most frequent first queries in the *non-goal* task in the *multi-stage* interface

Term	Occurrences
Harry Potter	2

Table 10. Most frequent first queries in the *goal-oriented* task in the *baseline* interface

Term	Occurrences
Survival	15
Surviving on a desert island	11
Surviving desert island	10
desert island	10
desert island survival	8
Surviving island	7
Surviving	6
Survival guide	5
Survival island	5

Table 11. Most frequent first queries in the *goal-oriented* task in the *multi-stage* interface

Term	Occurrences
Survival	19
desert island	16
Surviving on a desert island	8
Surviving	6
Desert	6
Survival guide	4
Surviving desert	3
Survive nature	3
How to Survive on a desert island	3
survival desert island	3

deal with the task starting with the the first given item "Select one book about surviving on a desert island". It also shows that a number started with an exact copy of the given task.

3.5 Query Re-Formulation

All of the three queries in each participants sessions for each interface were coded to identify any query reformulation patterns. The study followed the typical approach of manually analysing the transitions between query pairs in the session [2, 3, 5]. An example of two Query pairs in one session can be seen in Table 12.

Table 12. Examples of query reformulation

Term	Code
q=desert+island	first
q=desert+island+survive	specialisation
q=desert+island+survive	specialisation
q=survive	generalisation

Table 13. Number of reformulations for all task / interface combinations

	Non-goal Task		Goal-oriented Task		
# Words	**Baseline**	**Multi-Stage**	**Baseline**	**Multi-Stage**	**Total**
Specialisation	27	13	51	13	**104**
Generalisation	3	0	21	7	**31**
Parallel	18	3	43	21	**85**
New	28	24	47	57	**156**
First	72	51	88	86	**297**
Backtrack	3	1	3	7	**14**
Repeat	4	19	14	21	**58**

To code the reformulations, the three commonly used groupings/categories were adopted: **Specialisation**, **Generalisation** and **Parallel**. As a method of dealing with the log data a **First** category was added to quickly identify the start of a session. Next a **New** category was required to handle the queries that were not related to the previous. Two other categories were also added to handle the remaining interactions in the log data: **Repeat** for use where a previous query was used straight away, and **Backtrack** where a query was returned to. Table 13 shows the total occurrences of each category by task and interface. There are a couple of interesting figures in this table:

: There are significantly higher numbers of New classifications in the the *goal-oriented* tasks than the *non-goal* tasks.
: There are more Specialisations made in the Baseline interface tasks.
: The number of Generalisations is lower than one would expect.
: There are more Parallel moves in the Baseline interfaces than there are in the Multi-Stage.

Every instance of users starting the *goal-oriented* task as a known item search using the search term "surviving on a desert island" is followed by a **Generalisation** of this initial search, most to just "desert island". Opposed to this is where participants started with single word query of "survival" most followed this with a **Specialisation** to "desert island survival" or "island survival"

To fully answer RQ1 there needs to be more work undertaken. But from the findings above there is a an indication that just from the query terms and the order they were queried in, a prediction on the type of task the participant had undertaken could be accurately made.

4 Using the Multi-Stage Interface

The next question investigated how participants interacted with the multi-stage interface, in particular whether participants used the three stages differently for the two tasks, whether distinct user groups can be identified in the log data, and whether there are any patterns in how participants used the subject hierarchy exploration interface.

4.1 *Non-goal* vs *Goal-oriented* Sessions

The first question investigated is whether there is a difference in the uses of the "Explore", "Search", and "Book-bag" sections. The analysis is based on the log data, identifying when the participants switch between the three stages in the multi-stage interface. The time spent in each of the stages is then aggregated by participant. Table 14 summarises the time spent in each of the three sections.

Table 14. Time spent in each of the three stages in the *multi-stage* interface

	Non-goal Task		**Goal-oriented Task**	
	Mean (sec)	**%**	**Mean (sec)**	**%**
Browse	427	(.56)	296	(.32)
Query	240	(.22)	474	(.51)
Book-bag	171	(.20)	165	(.15)

One aspect that stands out clearly from last year's experiment is the much higher amount of time spent in the "Book-bag". In the first iteration of the interactive Social Book Search task, participants barely spent any time interacting

with the "Book-bag" (median time spent 0), whereas here we see that participants spend on average about two and three-quarter minutes interacting with the "Book-bag". It seems that the changes made to the interface and task based on last year's results have had an impact on how participants interact with the "Book-bag". The other interesting aspect is that the amount of time spent in the "Book-bag" is the same for both tasks. It seems that the "Book-bag" stage is used in the same way in both tasks.

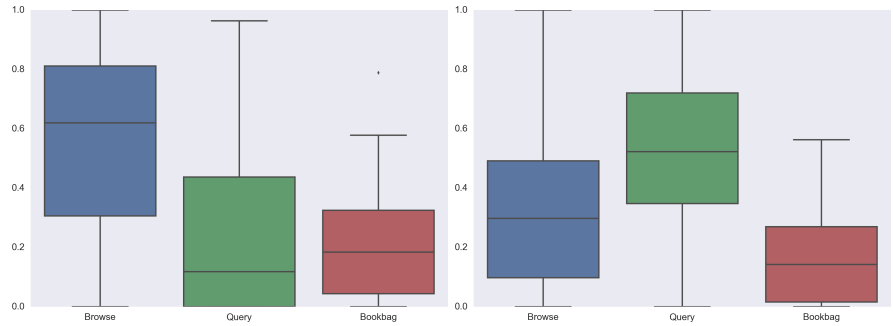


Fig. 2. Boxplots showing the normalised time spent in each of the three stages (*Browse*, *Query*, *Bookbag*)

While the actual times spent in the three sections are not statistically significantly different between the two tasks, when normalised to the scale 0 to 1 (Figure 2), there is a significant difference between the two tasks (Welch' t-test, Browse & Query $p < .01$; Book-bag $p < .05$). Looking at the boxplots in Figure 2 it seems clear that the browsing functionality is of more help in the *non-goal* task, while for the *goal-oriented* task participants focus on using the standard faceted query functionality. This is as expected, as the open-ended nature of the non-goal task is perfectly suited to using the subject hierarchy tree to explore the collection.

4.2 Cluster Analysis

The boxplots in Figure 2 indicate that there is a lot of variation in how the participants use the different sections of the system. To determine whether there are groups of users with similar behaviours the log data was processed and the actions classified into the categories shown in Table 15. For each participant the number of times they used each of the actions was counted and the resulting vectors then used to cluster the participants.

Average, single-linkage hierarchical clustering was used to cluster the participants in both tasks (Figure 3), using cosine similarity as the distance metric. A cosine similarity of 0.2 was used as the cut-off value to define the clusters. In

Table 15. Action Classification Scheme

Browse	The participant clicked on a topic in the hierarchy browser.
Query	The participant entered a query into the search box or clicked on a meta-data element to search by that element.
Facet	The user added or removed a facet.
Paginate	The user used the pagination functionality after either browse or query
Item	The user viewed an item or switched between item meta-data views.
Bookbag	The user added or removed a book to/from the book-bag, added a note to a bookbag item, or re-ordered the books in the book-bag.
Similar	The user used the similar-books feature in the book-bag

the *non-goal* task eight clusters are identified, while the *goal-oriented* task has seven clusters.

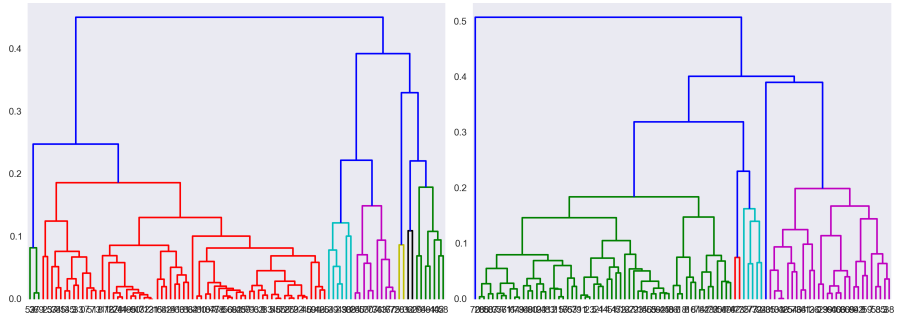


Fig. 3. Average, single-linkage, hierarchical clustering results

In the *non-goal* task there are three clusters with participants that primarily used the "browse" functionality (total 77 participants), one that used both the "browse" and "query" functions (10 participants), and one cluster with participants who primarily used the "query" function (3 participants).

Of the three "browse" clusters the main cluster (61 participants) participants use the "browse" functionality and interact quite heavily with the individual items. The second cluster (10 participants) shows the same basic behaviour, but they also use the pagination functionality to view more than just the first page of books for a topic. The third cluster (6 participants) again mostly uses the "browse" functionality, but they do not view any of the items' meta-data in detail. Overall it is clear that focusing on the "browse" functions is the preferred interaction pattern when exploring in an open-ended exploration task.

In the *goal-oriented* task the situation is more diverse. The biggest cluster (51 participants) still primarily uses the "browse" functionality, but unlike in the *non-goal* task, participants in this cluster augment this with a few queries. There are three query clusters (36 participants in total), which distinguish themselves based on their use of other features. As in the *non-goal* task, one cluster (10 participants) uses only the query and the result list, not interacting with the items at all. The second cluster (16 participants) uses the item meta-data more heavily. The third cluster (10 participants) makes heavy use of the faceting functionality, and has item meta-data use roughly between the other two groups. There is also a mixed browse & query group (6 participants). Interesting is that even though the *goal-oriented* task would seem to lend itself towards using queries, we still see a major focus on using the "browse" function.

4.3 Browse Patterns

To investigate how participants interacted with the topic hierarchy in the "Explore" section, the patterns of moves between topics in the hierarchy were analysed. When the user selects a topic in the hierarchy, based on the previously selected topic the action is classified into one of the following five "moves":

- **Start**: the participant has not previously selected a topic and selects a top-level topic;
- **Depth**: the participant selects a child topic of the currently selected topic;
- **Breadth**: the participant selects a sibling of the currently selected topic or a sibling of one of the current topic's ancestors;
- **Backtrack**: the participant selects one of the ancestor topics of the current topic;
- **Restart**: the participant selects a top-level topic that is not related to the current topic.

The next pre-processing step merges consecutive runs of the same move into a single move for the **Depth**, **Breadth**, and **Backtrack** moves. This was done as there was a large amount of variation in how many of the different moves the participants used, thus merging consecutive moves makes the analysis possible. Next each participant's action sequences were split at the **Start** and **Restart** moves to create the final list of sequences, based on which the sequence frequencies were calculated (Table 16). In the *non-goal* task 27 sequences are identified of which 13 occur two or more times. In the *goal-oriented* task 27 sequences are identified of which 19 occur two or more times.

Overall the first thing that stands out is that in the majority of cases participants investigate one sub-tree in the hierarchy at a time. The "Start → Breadth → Depth" sequence is the only one where participants viewed a number of top-level topics before digging down into one of them. It seems that breadth-first searching is generally not a strategy preferred by users, regardless of task.

The stand-alone "Start" sequences are caused by the participant switching to the "Search" page and then back, which causes the hierarchy to reset and

Table 16. Top five most frequent browse sequences

Non-goal		Goal-oriented	
Sequence	**#**	**Sequence**	**#**
Start → Depth → Breadth	(26)	Start → Depth	(42)
Restart → Depth	(25)	Start	(24)
Start → Depth	(25)	Restart → Depth	(20)
Start → Breadth → Depth	(17)	Start → Depth → Breadth	(19)
Restart	(13)	Start → Breadth → Depth	(17)

thus there is no previously selected topic. The stand-alone "Restart" is caused by participants using one of the other sequences, then clicking on a top-level topic and then moving to the "Search" stage.

Interesting is also that while in the *non-goal* task, the depth and then breadth is significantly more frequent than breadth and then depth; in the *goal-oriented* task, depth then breadth is roughly as frequent as breadth then depth. Clearly the two tasks induce different uses of the hierarchy. One possible explanation is that in the *non-goal* task participants pick the first top-level topic that seems interesting and then dig down to see what is available and if there is anything they might be interested in. The use of breadth after digging down indicates that they are then browsing around. Potentially the use of Breadth then Depth browsing pattern indicates a less clear idea of what the participant might be interested in, something that could potentially be taken advantage of by the interface to suggest areas of interest to the user.

5 Temporal Activity Analysis

Time provides us with an interesting perspective into user interaction with the systems. By studying how the interactions are distributed across a relative session time line, provides an insight into the behavioural patterns of the users. Our investigation began by processing and filtering the actions in relation to the task and interface they occurred in. The next phase was to adjust the timestamps of each action via normalisation so each users session and their actions could be measured on a common relative scale. This enabled us to utilise kernel density estimation resulting in a representation of the relative distribution of the users actions in relation to session length. During the analysis of these results we were able to compare the *goal-oriented* task against the open task for each interface which revealed some interesting preliminary results.

Within both interfaces it appears the task influences the users behaviour when interacting with the same system. After analysing and comparing the distribution of user actions, the following comparisons where found to be of most interest.

Baseline Interface: Book Bag

For each of the tasks, the way in which the users interact with the book bag within the baseline interface is different. In Figure 4 we found a lot of the user interaction when adding to the book bag occurs in first quarter of the session and follows a very similar distribution to that of the query. However in the open task the *remove facet* distribution appears to indicate the participants may already have an idea of the books or types of books they are searching for and only adding books which met the participants selection criteria.

Multi-stage Interface: Book Bag Interaction

Figure 5 *add-to-bookbag* distributions for both the *goal-oriented* and open tasks are very similar, However there is a clear difference in the distribution of the *remove-from-booking* action. In the *goal-oriented* task, users are more likely to remove books from their book bag throughout the duration of the session with a slight increase in probability the final quarter. This is to be expected in a goal orientated experiment since the participants where asked to select 5 books. Whereas in the open task the data suggests that users are more likely to remove a book from their book bag in the final half of the session, with the probability increasing as they approach the end. This "clean-up" phase may indicate that participants are collecting more books which meet their criteria and filtering those who don't make the cut at the end of their session. Suggesting the removal of books in the open task is an indicator that the user is about to finish using the system.

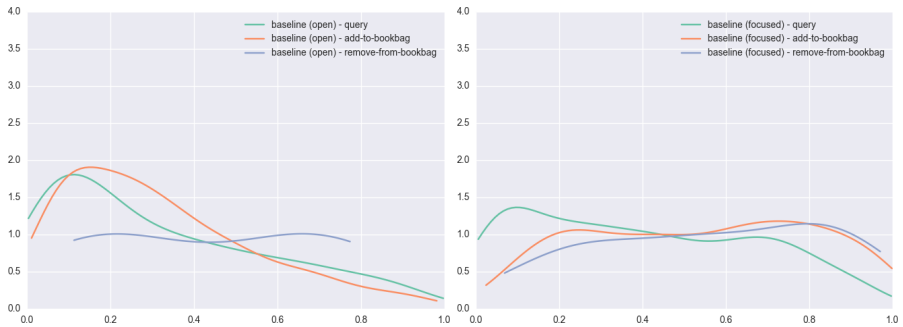


Fig. 4. Action-time KDEs for the query (green), add-to-bookbag (orange), and remove-from-bookbag (blue) actions using the *Baseline* interface. Times have been normalised to $[0, 1]$.

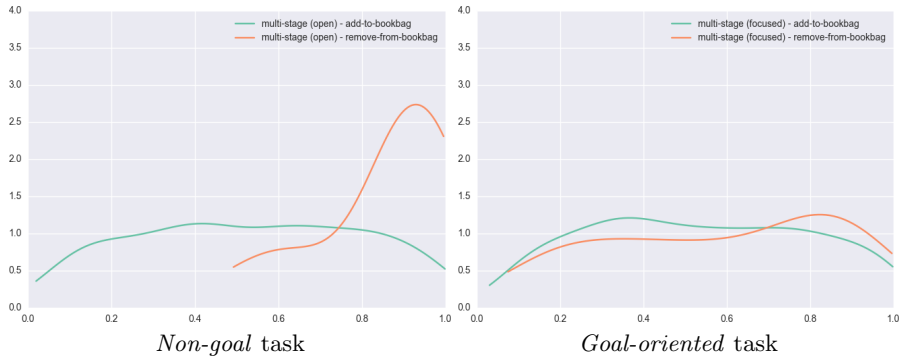


Fig. 5. Action-time KDEs for the add-to-bookbag (green) and remove-from-bookbag (orange) actions using the *Multi-stage* interface. Times have been normalised to $[0, 1]$.

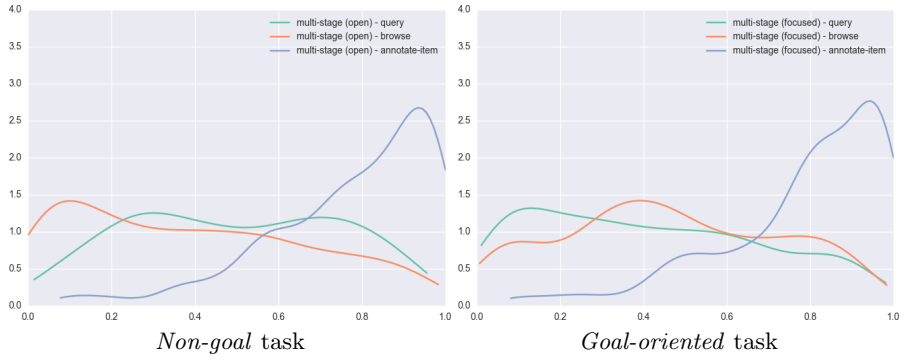


Fig. 6. Action-time KDEs for the query, browse, and annotate-item actions using the *Multi-stage* interface. Times have been normalised to $[0, 1]$.

Multi-stage Interface: Query, Browse and Annotate

Both tasks share a similar distribution with regards to users annotating books they have selected gradually increasing in probability towards the end of the session. However it is interesting to see how users are utilising queries and browse actions in both tasks (Figure 6). In the very early stages of the *goal-oriented* task we can see that the majority of users begin by performing a query and as time progresses, users began to use the browse functionality. In the open task it appears to be quite the opposite. Users initially begin by browsing the system however as time progresses they are less likely to browse but instead resort to using queries to explore the collection.

Multi-stage Interface: Facet Search Interaction

In Figure 7 our immediate attention was drawn towards the differences between the two *remove facet* action distributions.In the *goal-oriented* task the *remove facet* interactions follows a very similar distribution pattern to that of the *add facet*. However in the open task, the *remove facet* distribution suggests that a high volume of users at a similar relative point in their session performed this action. This maybe the result of users fine tuning their search or exploring a different topic completely or users switching between the stages which clears all currently selected facets. However at this stage we can only speculate what factors are causing this peak, as this requires further investigation

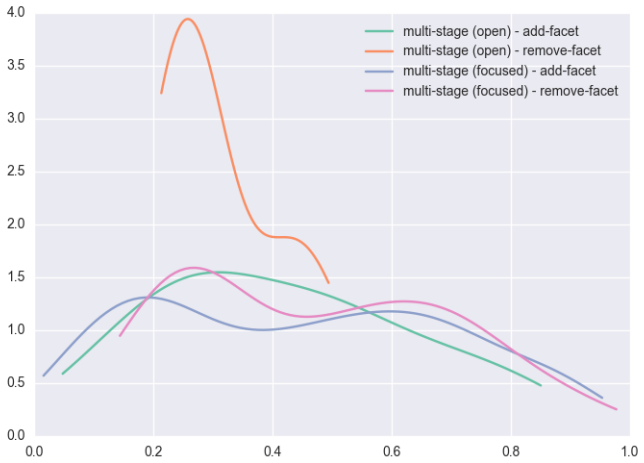


Fig. 7. Multi-Stage Interface: Kernel density estimation for user interaction with the facet search functionality for the *goal-oriented* and open tasks

6 Discussion

In this paper we have analysed the 2015 Interactive Social Book Search dataset looking at the initial queries, use of the multi-stage interface, and also how particiants use of the interfaces changed over the duration of their sessions.

The first queries participants issued showed a clear differentiation between the two tasks. In the *goal-oriented* task, participants made heavy use of the task's instruction text to define their initial query, while in the *non-goal* task there is almost no overlap between the query words used. Participants also generated longer queries in the *goal-oriented* task. While we have undertaken some initial classification of the queries, we are planning to conduct a more in-depth analysis of the query types and topics, particularly in the *non-goal* task.

An analysis of the *multi-stage* interface indicates that the re-design of the *multi-stage* interface, based on last year's results, has achieved its aim and participants consistently used all three of the stages. Encouraging is that participants made full use of the third stage, particularly in the *non-goal* task where participants used the third stage to filter down the set of books they had collected. We also saw differences between the two tasks. Participants made more use of the *Browse* stage in the *non-goal* task, while in the *goal-oriented* task the focus was on using the *Query* stage. This is also supported by a temporal analysis, where participants focused on querying first, before then using the *Browse* functionality in the *goal-oriented* task, while in the *non-goal* task, the order is reversed.

Finally, we have looked at using clustering to distinguish different interaction styles in the multi-stage interface. While significant further work is needed here, there is a clear distinction between users who prefer the *Browse* functionality versus those who prefer *Query*. Also it looks like it might be possible to distinguish between different groups of users within those two major groups, which could be used to adapt the interface to the users' preferences.

Bibliography

[1] M. A. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.

[2] B. J. Jansen, D. L. Booth, and A. Spink. Patterns of query reformulation during web searching. *Journal of the american society for information science and technology*, 60(7):1358–1371, 2009.

[3] S. Jesper, P. Clough, and M. Hall. Regional effects on query reformulation patterns. In *Research and Advanced Technology for Digital Libraries*, pages 382–385. Springer, 2013.

[4] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *JASIS*, 42(5):361–371, 1991.

[5] S. Y. Rieh et al. Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management*, 42(3):751–768, 2006.

[6] 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.