Characterizing Stages of a Multi-session Complex Search Task through Direct and Indirect Query Modifications

Jiyin He
Centrum Wiskunde en
Informatica
Science Park 123, 1098XG
Amsterdam, the Netherlands
j.he@cwi.nl

Marc Bron
University of Amsterdam
Science Park 107, 1098XG
Amsterdam, the Netherlands
m.m.bron@uva.nl

Arjen de Vries
Centrum Wiskunde en
Informatica
Science Park 123, 1098XG
Amsterdam, the Netherlands
arjen.de.vries@cwi.nl

ABSTRACT

Search systems use context to effectively satisfy a user's information need as expressed by a query. Tasks are important factors in determining user context during search and many studies have been conducted that identify tasks and task stages through users' interaction behavior with search systems. The type of interaction available to users, however, depends on the type of search interface features available. Queries are the most pervasive input from users to express their information need regardless of the input method, e.g., typing keywords or clicking facets. Instead of characterizing interaction behavior in terms of interface specific components, we propose to characterize users' search behavior in terms of two types of query modification: (i) *direct* modification, which refers to reformulations of queries; and (ii) *indirect* modification, which refers to user operations on additional input components provided by various search interfaces. We investigate the utility of characterizing task stages through direct and indirect query reformulations in a case study and find that it is possible to effectively differentiate subsequent stages of the search task. We found that describing user interaction behavior in such a generic form allowed us to relate user actions to search task stages independent from the specific search interface deployed. The next step will then be to validate this idea in a setting with a wider palette of search tasks and tools.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval; H.5.2 User Interfaces (Evaluation/methodology)

Keywords

Query modification, multi-session task

1. INTRODUCTION

In order to effectively satisfy users' information needs as expressed by a query, search systems use context [9]. Many forms of context exist, including previous queries issued and users' interac-

tions with search systems. To correctly interpret contextual signals it is important to determine the underlying factors causing these signals. One salient factor that determines the context of a user is the task type and task stage in which the user is engaged [3]. Tasks can range from simple factual search to complex multi-session search tasks and can consist of single or multiple stages [8]. The interaction of the user with the search system may change depending on task type and task stage. For example, Vakkari et al. [12] found that students' term addition and removal behavior changes during various stages of a research project. For difficult tasks, studies found that users tend to formulate more diverse queries and use more advanced query operators [2, 6]. In exploratory search tasks, users prefer more advanced interface features such as facets and comparisons [7, 13]. Work on aggregated search interfaces found that users have a tendency to click on results from more and more diverse facets in complex tasks [1, 11]. Independently each of these studies contributes to our understanding of user interaction with various types of systems across different tasks and stages. The interaction patterns discovered, however, depend on the type of system used in the study, e.g., a web search engine provides less interface features for interaction than an exploratory search system.

Queries are the most pervasive input from users to express their information need regardless of the input method, e.g., submitting a query by typing keywords in a search box or modifying a query by clicking a facet. Therefore, users' queries and the associated query formulation behavior provide a consistent feature to determine task type and stage. We propose to characterize users' search behavior in terms of two types of query modification: (i) *direct* modification, which refers to reformulations of queries, such as adding, removing or substituting terms in consecutive queries; and (ii) *indirect* modification, which refers to user operations on additional input components provided by various search interfaces such as using filters, and switching between collections.

We report on a case study that investigates how users' search behavior changes throughout different stages of their search task. We focus on two types of search behavior, "exploration" and "focused search", and how these behaviors are reflected in users' interaction with a system in terms of direct as well as indirect query modifications. We follow 25 students during a four week project. We provide students with an interface that provides various interface options, e.g., facets, verticals, and aggregated display. The research project represents a complex search task spanning multiple sessions and stages. We confirm empirically that subjects exhibit different query modification patterns during the various stages of the task.

Our contribution is a demonstration of the effectiveness of using direct and indirect query modifications as a way of characterizing different stages of a complex multi-session search task. The ben-
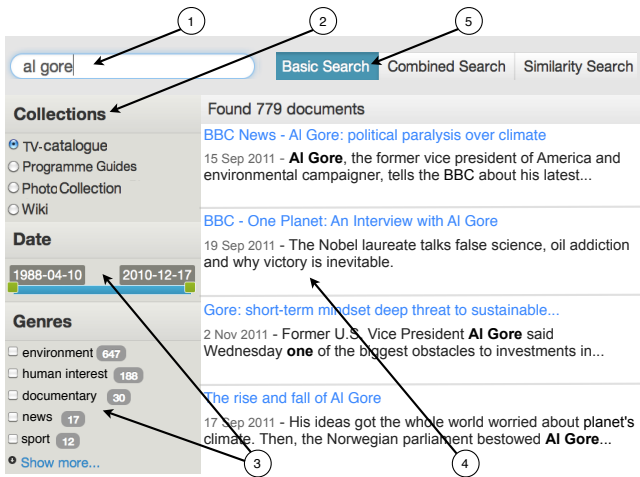
Figure 1: A screen shot of the aggregated search interface. Numbers are used to indicate specific components in the text.

stage	#switch_screen	#switch_collection	#filter
1	46	71	44
2	30	94	37
3	28	81	52

Table 1: Cross-tabulation of the number of times an interface feature is used for each stage of the research project.

efit of transforming interface specific interactions to more general interaction types, i.e., "exploration" vs. "focused search," is that interactions observed with different interfaces become comparable.

2. EXPERIMENTAL SETUP

Search task. The study was conducted as part of a course on "audience reception studies" offered at the Media Studies department of our university. During a four week project, students investigated the historical context of the emancipatory role of female television or film personalities. The students concluded the project by writing a photo essay incorporating primary and secondary sources to place the photos in context. The design of the project assignment can be divided into 3 stages: (i) in the first week students were asked to familiarize themselves with the topic and to pick a number of television or film personalities; (ii) in the second week students collected material, i.e., images and documents that motivated their choice for these images; and (iii) in the third and fourth week students organized the material and prepared a final presentation.

Collections. Six collections, relevant to the assignment about television and film personalities, from various archives were indexed: (1) a television program collection containing 0.5M metadata records; (2) a photo collection with 20K photos of people working at television studio; (3) a wiki dedicated to actors and presenters (20K pages); (4) 25K television guides that are scanned and OCRed; (5) scanned and OCRed newspapers between 1900 and 1995 (6M articles); and (6) digital newspapers between 1995 and 2010 (1M articles). These collection are indexed using Lucene SOLR 4.0 and we use BM25 as the retrieval model.

Search system. We provided subjects with a search system that enables various means of interaction in order to support diverse search behaviors in the various stages of the project. The basic funtionality of the search system is modeled after a standard web search interface, see Figure 1, providing a search box (1) and presenting result snippets in a ranked list (4). Access to the various collections is provided by a navigation menu (2). Additionally, the system supports facets with query preview capabilities (3) that provide users with term suggestions and the ability to quickly filter and zoom in on a specific topic [14]. We also provide two additional interaction capabilities accessible via tabs (5). "Combined Search" supporting exploration of the various collections by providing an aggregated search display. In this display all collections are shown simultaneously. Results for each collection are horizontally orientated and the vertical order of the six collections is fixed. The "Similarity Search" has the same layout as the aggregated display and provides a find-similar feature [10]. By clicking on a document the user submits the current query and the first 100 words of the clicked document as an OR query.

Subjects. In total 25 postgraduate students Media Studies participated. The sample consists of 12 men and 13 women, aged in terms of median (MD) and interquartile range (IQR) around 23 years (MD=23, IQR=22-24). We asked subjects background questions using a 5 point Likert-type scale, where a one indicates no agreement and a five indicates extreme agreement. Subjects reported high levels of experience in general computer use (MD=4, IQR=4-5) and using online search tools (MD=4, IQR=4-5).

Log data. We record all queries submitted to the system and any components clicked by the subjects: switch screen; change collection; check/uncheck filter; paginate; view document; and bookmark document. The switch screen and change collection actions resubmit a query to the system if the user has already provided one during the session, while the filter actions modify a query by adding or removing a term. Due to the limited number of "find similar" operations (4 times in total), they are left out of the analysis.

3. LOG ANALYSIS

3.1 Interface specific interactions

We first determine whether we are able to detect a difference in the interaction behavior of the subjects in different stages of the research project in terms of the features specific to our search system. Table 1 shows a cross-tabulation of the number of times an interface feature is used during each stage of the research project. We find a significant interaction between project stage and the number of times interface features are used: χ^2(df=6, N=483) = 11.4, p < .05). In the first stage the differences between using different interface features are smaller than in later stages in which switching collection is more popular.

Note, that the interactions in Table 1 are highly specific to our interface making comparison to observations of interactions with other search interfaces difficult. Next, we will look at more general forms of characterizing interaction behavior with our system that are not as dependent on a specific search interface.

3.2 Direct query modification

Reformulation types. We consider query reformulations between two *consecutive queries* as direct query modifications. In the literature query reformulation has been defined differently, e.g., term addition, removal, and substitution [5]. We simplify the reformulation types into two types: *new* or *related*. Although reducing the number of reformulation types may give the query reformulation patterns less discriminative power, each of the remaining types has an intuitive interpretation. Let query $q_1 = \{t_i\}_{i=1}^n$ and $q_2 = \{t_j\}_{j=1}^m$
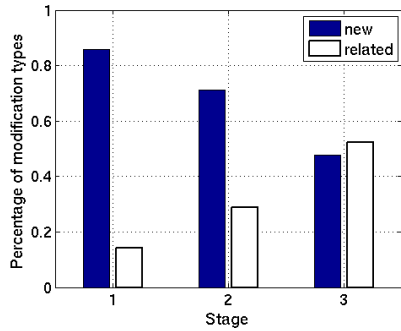
Figure 2: A comparison of the percentage of direct query modification types in 3 stages: "new query" versus "related query".

Stage	Queries
1	"mies bouwman", "liesbeth list", "sylvia kristel", "monique van de ven".
	Note: Each of the query is a Dutch celebrity's name.
2	"sonja barend", "sonja barend op maandag", "sonja barend award", "sonja op maandag", "sonja barend nts".
	Note: Sonja Barend is a Dutch television personality.
3	"wie van de drie", "sonja's goed nieuws show", "goed nieuws show", "sonja goed nieuws show", "wie van de drie", "sonja goed nieuws show".
	Note: "Wie van de drie" and "Goed nieuws show" are two TV programs that Sonja Barend has been involved in.

Table 2: An example of the sequence of queries issued by a user in different stages.

be two consecutive queries. q_2 is a "new" query with respect to q_1 if $q_1 \cap q_2 = \emptyset$, and a "related" query otherwise. That is, two queries are related if the terms in two queries overlap. We assume that users issue "new" queries for different topics, and "related" queries for related topics. From our log data, we obtain 42, 128, and 82 consecutive query pairs in stage 1, 2, and 3, respectively.

Results. Figure 2 shows the composition of direct query modifications in each of the three stages. That is, the percentage of "new" and "related" queries among all direct query modifications. We see that the compositions show clear differences between the three stages. In stage 1, "new" queries are the majority. The percentage of "new" queries decreases as the users move towards further stages of their search task. In stage 3, the percentage of "new" and "related" queries is relatively balanced. This implies that, while reformulating queries, users tend to issue more queries on different topics in the initial stage, and in later stages more queries on related topics. A χ^2 test confirms that there is a significant interaction between project stage and the number of new and related queries issued: $\chi^2(df = 2, N = 252) = 21.14, p < 0.01$.

Figure 3 shows patterns of query reformulation actions. In these graphs, nodes include three states: s (starting query), new (new query), and rel (related query); and edges are weighted by a transition probability from one node to another, calculated as $count(n_i \to n_j)/count(n_i)$. For instance, after issuing a "new" query, the probability of issuing a "related" query in stage 1 is 0.08, and the probability of issuing a "new" query is 0.92. While our data is too sparse to examine the difference between these patterns, we do observe that transition probabilities change across stages.

In summary, patterns of users' query reformulations seem to reflect the underlying stages of the search task. In particular, in initial stages users issue queries associated with more diverse topics than in later stages. That is, initial queries are more likely exploratory, while later queries are more focused. The stages of the assignment, and the necessity to first explore a topic before focussing on a specific aspect both contribute to these variations in search behavior. Note, however that it is possible to discriminate between these behaviors based solely on the queries submitted to the system.

It is of course possible that queries with completely different sets of terms refer to similar topics. Analyzing the semantics in this case of the queries would be useful [4], which we plan for future investigation. Here, without a systematic mapping between queries and their underlying topics, we show in Table 2 an example of a user's queries in different stages, which to some extent illustrates the type of topics being searched. In this example, in stage 1, the user searched for a number of female dutch television personalties, using their names as queries. In stage 2, his/her search concerns

different topics associated with Sonja Barend, a famous Dutch television personality. In stage 3, his/her queries are focused on two specific TV programs that Sonja Barend has been involved in.

3.3 Indirect query modification

Indirect query modification types. Some interfaces provide users with additional input components to a query input box. In such systems direct query modification patterns may not be as salient a feature to differentiate task stages By using these input components, users indirectly revise the queries they have issued. We therefore examine how users use these indirect modifications of the same query in different search stages.

We consider the following operations as indirect query modification: *switch screen*, *change collection*, and *check/uncheck filters*. Switching screen allows users to review top results from different collections. Changing collection allows user to query a different collection, and applying filters allows users to explore results in specific aspects of the query. These operations correspond to different ways to *explore* different aspects of the query.

Intuitively, as shown in the direct query reformulations, users may explore more in the initial stage, and narrow down their search with less exploration in later stages. However, while it could be that users are indeed exploring the content of the collections, an alternative explanation would be that they are simply exploring the functionality of the system. To this end, we investigate an additional set of operations, referred to as *result examination operations*, including *pagination*, *view document*, *bookmark document*. Different from the above exploration operations, with these operations users examine the retrieved documents with respect to a (unmodified) specific query. If users were trying out the system features, intuitively, they would not dig into the details of the retrieved content.

To clarify, the indirect query modifications only include exploration actions, excluding the result examination operations. We investigate the former to see whether users' exploration behavior differs across stages, and the later as signals to indicate whether users explore the content or the functionality of the system.

Results. Unlike direct query modification that occurs between two queries, indirect modification occurs within a same query. Table 3 shows the average number of exploration actions per-query in the three stages. We see that there is a significant difference between the number of exploration actions issued in stage 1 and that of stage 2 and 3. This confirms our expectation that users explore more different aspects/collections with respect to a same query in the initial stage, and less in later stages.

Table 4 shows that, in terms of the per-query number of result examination operations, no significant difference was found between different stages. We find no evidence that users examine less documents/result lists in the first stage compared to the later stages. This suggests that the difference between exploration behaviors in different stages can not be explained by "initial exploration of the
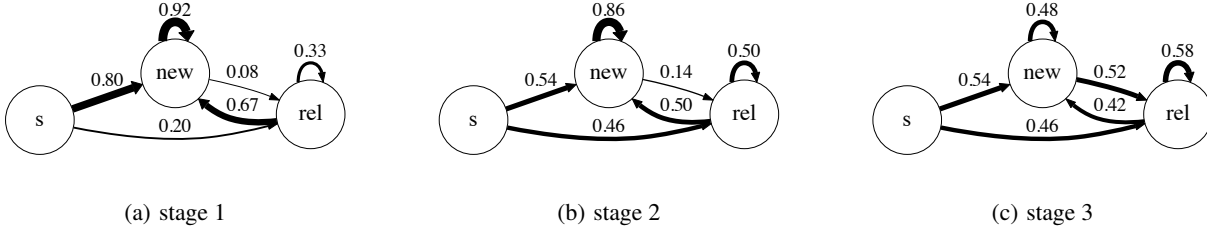
(a) stage 1   (b) stage 2   (c) stage 3

Figure 3: The subfigures (a), (b), and (c) shows the transition probabilities for stage 1, stage 2, and stage 3 of the project respectively.

Stages	Avg.#E	Comparisons	p-values
1	1.81	1 vs. 2	**0.04**
2	0.85	2 vs. 3	0.82
3	0.90	1 vs. 3	**0.03**

Table 3: Comparing the number of "exploring" actions (#E) in the 3 stages, significance tested using Wilcoxon ranksums test.

Stages	Avg.#A	Comparisons	p-values
1	2.63	1 vs. 2	0.15
2	2.13	2 vs. 3	0.21
3	2.58	1 vs. 3	0.67

Table 4: Comparing the number of "examination operations" actions (#A) throughout the 3 stages.

system features", given that users show similar behavior in examine details of retrieved content.

4. DISCUSSION AND CONCLUSION

We have investigated user search behavior in a complex multi-session search task, with a search system that provides various types of input components. We confirm results from previous studies that user search behavior changes during the various stages of an information seeking task and that variations in the use of specific input components are salient features in detecting these changes. We further show that users' query modification behavior, both in terms of direct query reformulations as well as in terms of indirect modifications derived from operations on advanced interface components reflects the underlying task stages users go through. With direct query modification, users explore more diverse topics in the initial stage of their task compared to later stages, which are more focused on related topics. With indirect query modification, users explore more aspects of a single query in the early stage compared to later stages. Being able to transform interface specific interactions to more general interaction types, while maintaining the ability to discriminate between stages, suggests that user interaction behavior can be described in a form comparable across search interfaces with varying input components.

Our study design did not allow for complete control of all conditions. Users were not restricted to our system to accomplish their assignment, reducing the potential number of queries issued and limiting our analysis to just that subset. Also, the size of the search logs we acquired is too small to allow reliable study of higher order query modification sequences. Our planned follow-up research is to acquire search log data from a wider variety of search interfaces

and tasks, to verify the utility of direct and indirect query modifications to analyze user behavior in information seeking tasks.

Acknowledgements

5. REFERENCES

[1] J. Arguello, W.-C. Wu, D. Kelly, and A. Edwards. Task complexity, vertical display and user interaction in aggregated search. In *SIGIR'12*, pages 435–444, 2012.

[2] A. Aula, R. M. Khan, and Z. Guan. How does search behavior change as search becomes more difficult? In *SIGCHI'10*, pages 35–44, 2010.

[3] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: Part i. background and theory. *J. doc.*, 38(2):61–71, 1982.

[4] V. Hollink, T. Tsikrika, and A. P. de Vries. Semantic search log analysis: a method and a study on professional image search. *JASIST*, 62(4):691–713, 2011.

[5] B. J. Jansen, D. L. Booth, and A. Spink. Patterns of query reformulation during web searching. *JASIST*, 60(7): 1358–1371, 2009.

[6] D. Kelly, K. Gyllstrom, and E. W. Bailey. A comparison of query and term suggestion features for interactive searching. In *SIGIR'09*, pages 371–378, 2009.

[7] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *JCDL'09*, pages 313–322, 2009.

[8] Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *IPM*, 44(6):1822–1837, 2008.

[9] M. Melucci. Contextual search: A computational framework. *FnTIR*, 6(4-5):257–405, 2012.

[10] M. Smucker and J. Allan. Find-similar: similarity browsing as a search tool. In *SIGIR'06*, pages 461–468, 2006.

[11] S. Sushmita, H. Joho, M. Lalmas, and R. Villa. Factors affecting click-through behavior in aggregated search interfaces. In *CIKM'10*, pages 519–528, 2010.

[12] P. Vakkari, M. Pennanen, and S. Serola. Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *IPM*, 39(3):445–463, 2003.

[13] R. Villa, I. Cantador, H. Joho, and J. M. Jose. An aspectual interface for supporting complex search tasks. In *SIGIR'09*, pages 379–386, 2009.

[14] R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.