

Exploratory Searching As Conceptual Exploration

Pertti Vakkari
University of Tampere
FIN-33014, Finland
Pertti.Vakkari@uta.fi

ABSTRACT

The aim of this paper is to analyze the characteristics of exploratory searching for inferring ideas on how to evaluate exploratory search systems. Exploratory searching is defined as conceptual exploration. Information search process is divided into major stages. Goals, criteria and measures of attaining goals in explicating information need and formulating search are proposed. They can be applied for evaluating search systems aiming at supporting these two stages in searching.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]

General Terms

Measurement, Performance, Human Factors.

Keywords

Exploratory Search Systems, Search Process, Outcomes, Evaluation

1. INTRODUCTION

Studies on exploratory searching have gained popularity in recent years, although the same phenomenon has been studied earlier by other names like information search process [6] or task-based searching [13; 14]. Exploratory searching is understood here as searching for learning or investigative activities as defined in [8].

If learning is the ultimate goal of the activity generating searching it is evident that the paradigmatic model of evaluating interactive information retrieval [2; 4; 15] is not sufficient for evaluating exploratory searching. It focuses too much on assessing the output of the search system, not sufficiently observing the outcome of the system. Outputs are the products delivered by a system, whereas outcomes are the benefits the system produces to its users [11]. Typical output in IIR evaluation is the number of relevant items retrieved. The outcome of searching like growing understanding of the topic or to which extent the system supports searchers reaching their goals at various stages of search process is typically left without notice [cf. 14]. There is a need to develop ideas for evaluating exploratory search systems based on a deeper understanding of exploratory searching [cf. 15].

The aim of this paper is to analyze exploratory searching for

creating ideas on how to evaluate exploratory search systems. First, the nature of activities leading to exploratory searching is discussed briefly. Based on that, information search process is conceptualized as representations of concepts and their relations in humans and documents [4]. After that exploratory searching is analyzed and ideas for evaluating search systems are introduced consisting of goals, evaluation criteria and their measures for the beginning stages of searching.

2. EXPLORATORY ACTIONS

A common feature of actions leading to exploratory searching is that the actor has insufficient information for solving an ill-structured problem for proceeding in her task [6; 13; 15]. This situation is called e.g. anomalous state of knowledge (ASK) [2] or uncertainty [6]. It is typical that the actor has problems to explicitly express her information need. This kind of ill-structured problems 1) begin with a lack of information necessary to develop a solution or even precisely define the problem, 2) have no single right approach for solution, 3) have problem definitions that change as new information is gathered, and 4) have no identifiable 'correct' solution [3]. Creating a solution is a gradual learning process, which implies integrating new information into existing mental models [3; 5].

Understanding learning as conceptual change [10] allows us to conceptualize mental models as consisting of concepts and their relations. This notion matches with the cognitive view of information retrieval, which conceptualizes information as knowledge structures both in humans and documents [4]. Exploratory searching is generated in situations, when actors' mental models lack concepts and relations between concepts for accurately representing the task. They have insufficient concepts and insufficient relations in their knowledge structure [12]. We claim that exploring concepts and their relations is a major characteristic of exploratory searching. Thus, we may say, that actors search information for obtaining concepts and their relations in order to understand, structure and represent their task more validly for proceeding in its performance.

In an ASK when actors have insufficient conceptual understanding, the goal of information searching is to help to understand and structure their task. By implication, the goal of exploratory search systems is to support the search process so that the searchers' understanding of their task grows and becomes more structured, and therefore produces more useful information items. Consequently, the evaluation of exploratory search systems includes both to which extent they support the search process, and produce useful information items [cf. 2].

3. EVALUATION

Evaluation begins typically by analyzing what is the goal of the system, process or service to be evaluated. It is assessed to what extent the object of evaluation attains the goals defined [11]. This requires that we have an understanding of which factors are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

associated in reaching the goal, i.e. what means are used to attaining it. In evaluation a distinction is made between outputs and outcomes. Outputs are the products delivered by a system, whereas outcomes are the benefits the system produces to its users [11]. Therefore, in evaluating exploratory search systems, one has to focus on the benefits the system produces to users during the search process, and on benefits the information items retrieved produce to the searchers' task.

For analyzing the outcomes of the system to users, we divide the search process into three main stages: explicating information need, search formulation, and evaluating search results. Search formulation can be divided into two elements; term selection and query formulation [1]. The former means expressing information need in search terms, and the latter combining terms for formulating a query. In this paper we focus on explicating information need and search formulation.

4. EXPLICATING INFORMATION NEED

4.1 Measuring conceptual constructs

The goal of explicating information need is to articulate it so that the actor has a growing understanding of the task for proceeding in its performance. Information need refers to insufficient conceptual construct (mental model) representing the task, which lacks necessary concepts and relations [12].

If the actor does not know much about her task, there may be alternative ways of conceptualizing the task and proceeding in its performance. A predefined, given conceptual structure to be used does not exist. The actor has to construct it incrementally while proceeding in the task. Kuhlthau's model indicates how the actor's understanding of a topic changes from vague to clear by formulating a focus [6], i.e. by constructing a necessary conceptual representation. Thus, the conceptual construct representing the task can be reconstructed only afterwards.

The actor's knowledge structure can be measured by observing the number of concepts and their relations that it consists of. Changes in the conceptual construct indicate to which extent the actor has been successful in explicating her information need.

If the goal of an exploratory search system is to help the searcher to express information need by explicating the concepts and relations it consists of, then the degree to which the articulation covers these concepts and their interrelations is the criterion of success.

This criterion is easy to apply to the given search tasks (topics) in retrieval experiments, because they contain the description of the information needs (topics). They are predefined, known in advance and do not change during the search process [4]. Thus, success can be measured by the proportion of articulated concepts of all concepts the topic consists of.

In tasks generated by actors the application of previous criterion is possible only afterwards. In an ASK it is not possible to know in advance the exact conceptual structure of information need [cf. 2]. However, there is some evidence of how conceptual construct changes when actors' understanding grows. In general, it changes from vague to precise [6; 13]. The extension of concepts decreases, the number of sub-concepts increases, and the number of connections between the concepts increases [5; 10]. This hints, that growth of understanding consists of an increasing number of concepts and their relations, and of the specificity of conceptual construct. Thus, increase in the number of concepts and in the

number of interrelations between these concepts, as well as in the specificity of conceptual construct are the criteria for success of explicating information need for searching. Specificity is reflected in the actors' ability to differentiate a concept into sub-concepts. This can roughly be measured by the proportion of sub-concepts of all concepts the conceptual construct consists of [cf. 9]. Measures and criteria of success are presented in table 1.

Table 1. Measures of success in explicating information need

| Measure | Criteria of success |
|--|---|
| Exhaustivity = # of concepts articulated | Increase in the # of concepts articulated <i>in ASK</i> , or Increase in the proportion of concepts articulated <i>in given topics</i> |
| Specificity = the proportion of sub-concepts of all concepts articulated | Increase in the proportion of sub-concepts of all concepts articulated |
| Combined measure | Increase in the # of concepts added by the increase in the number of specified concepts (greater weight) |
| Conceptual integration = # number of links expressed between concepts | Increase in the # links between concepts |

4.2 Measuring conceptual change

Explication of information need for constructing a focus, i.e. when is it fully represented as concepts, may require several iterations within a search session or even several sessions. In the following we focus on changing information needs [cf. 4] for presenting criteria of success in explicating them after the initial search.

If the information need is vague after the initial search, the actor continues explicating it. We may distinguish between two types of conceptual changes. The first one is conceptual continuity, which is based on the concepts explicated in the initial search. Conceptual change refers to the situation when the actor replaces at least an existing concept in the conceptual construct by a new concept with differing extension.

In conceptual continuity new explication includes either new concepts or specifications of the concepts explicated in the initial search or both. The specification of a concept means that its extension is smaller than the original one, but it belongs mainly to original extension. It is a class inclusion. An example of this is the specification of the concept "information seeking" as "information retrieval". Introducing new concepts or specifying old ones is likely to lead to a more specific articulation of information need. As stated earlier the increase in the number of concepts and their interrelations in explicating information need are the criteria of success. Also the proportion of sub-concepts of all concepts explicated can be used as a measure of the specificity of the conceptual construct (table 1). An additional measure of the specificity could be the proportion of concepts in the initial search (previous iteration) specified in the next iteration.

It is possible to form a combined measure for the success of explicating information need after the initial search. For each new concept that is introduced in the conceptual construct the weight of e.g. 1 is given, and for each old concept specified e.g. the

weight 1.5. The greater weight is assigned to specification, because it adds value to the original concept by limiting its extension. The measure is the sum of the values of the new and specified concepts. This combined measure gives a rough estimate in the increase of the articulation of information need.

In conceptual change at least one old concept in the information need is replaced by a new one with differing extension. Other concepts may be unchanged or specified. It is difficult to infer, how replacing concepts increases the actor's understanding of information need. In which cases it is justified to claim that the replacing of concepts has increased the understanding of information need?

If a concept or concepts are replaced without specifying the remaining ones, it is likely that the actor is surveying the conceptual space of the information need not being able to decide how to specify it. She may be looking for alternative explications among which to choose. We call this activity as conceptual mapping. It resembles Bates' vary tactics replacing an existing search term by another [1]. In our case, it is likely that the actor's information need has not become more specific, but it has changed on the same level of specificity.

Introducing new concepts in the conceptual construct typically includes creating new relationships between the new and old ones. These relationships contribute to the meaning of the construct [5]. If the actor specifies at least one of the remaining concepts when replacing one or more concepts, this hints that the relations to the new, replacing concepts have helped her to specify the original concept. The meanings of the replacing concepts have contributed in specifying the meaning of the original concept. Thus, at least in part of the specified concept the information need is more specifically articulated. This specification hints also that the replacing concepts are in some way more specific than the replaced concepts, implying that the explication of the whole information need is more specific.

We may measure the success of explicating conceptually changing information need as follows. In replacing concepts, the number of concepts does not increase, and thus the understanding has not grown in the sense of a more specific information need. However, replacing may lead to selecting a conceptual alternative among the surveyed concepts, and help the searcher to articulate information need. Therefore, only the concept, which is selected after varying concepts, will receive the weight (e.g. 1). If replacement is associated with the specification of at least one original concept, then each specified concept is assigned the weight 1.5.

5. SEARCH FORMULATION

Search formulation is divided into term selection for representing search concepts, and query formulation for combining search terms as a query [1]. We discuss first about term selection, and then query formulation.

5.1 Term selection

For evaluating exploratory search system providing support in term selection, we have to define the goal of term selection in order to be able to infer criteria of success. This goal can be understood from the angle of expressing information need or from the angle of information retrieval.

Term selection based on information need requires that an actor is able to express the concepts it consists of as search terms. The

goal of term selection is thus to express the concepts of information need. The exhaustivity of query refers to the extent to which the concepts of information need are expressed in the query [7]. Thus, the more exhaustive the query, the more successful term selection is from the angle of expressing the information need.

From the angle of information retrieval it has been typical to reduce the success in term selection to the number of relevant items retrieved based on new terms in the query, and inferred measures like precision or recall. Depending on whether the goal of searching is recall or precision, there are known procedures to aim at those goals. Other factors controlled, increasing the exhaustivity and specificity of query increases precision, whereas increasing the extent of the query (# of terms per concept) increases recall [7].

In exploratory searching the actor does not know exactly the concepts her information need consists of, but she tries to articulate them. How would measures like precision or recall reflect to which extent the actor has been successful in articulating the information need and expressing it by search terms? It seems that those measures are not very meaningful in estimating the success of those activities and the help provided by the system.

However, it is suggested that investigative searching is more concerned about recall (maximizing the number of possible relevant objects) than precision (minimizing the number of possibly irrelevant objects) [8]. When actors are exploring possible conceptualizations of their topic in order to formulate a focus, they are typically confused and overwhelmed with information. Information seems inconsistent and incompatible with their prior conceptual constructs [6]. In a situation like this, it is not likely that actors are concerned with recall, but precision. They are not interested in finding most of the documents, which would provide them with ideas in structuring the topic, but a sufficient number for formulating a focus. When the information need is structured containing all the necessary concepts, i.e. when it is stable [4], then it is likely that actors are concerned with recall and aim at comprehensive searches on the topic. Thus, in pre-focus stage, searching is precision oriented, whereas in post-focus stage it is recall oriented.

If searching aims at precision in pre-focus stage, then most productive in term selection is to express all the concepts of information need in search terms. As known, increasing exhaustivity increases precision [7].

5.2 Query formulation

Broadening the view from term selection to query formulation may be a more fruitful approach to evaluate search success. In exploratory searching query formulation is open due to the fact that not all concepts in information need are known. The actor aims at finding documents, which would provide her with ideas of how to structure her topic. She tries to find conceptual alternatives within the scope of her information need, and to compare those alternatives [cf. 8; 9]. It is a question about finding connections between concepts, i.e. finding propositions which connect concepts in a meaningful way. A proposition asserts something about two or more concepts and their relations [10].

As stated earlier, growth of understanding is characterized by the increasing number of concepts and their interrelations (i.e. propositions), and by the specificity of concepts. In order to structure information need, the actor should use search tactics that

would help her in exploring document space accordingly. A typical way of expanding and mapping conceptual structure is by adding a concept in a query and replacing it by another one leading to vary tactics [1]. Empirical findings confirm, that vary tactics are used for mapping conceptually the search topic [9]. The same study also showed, that searchers quite systematically chunked the search topic into conceptually smaller fields by using successive facets for inspecting the items retrieved [9].

A means of specifying concepts is to provide an actor with the sub-concepts of the concepts in her information need [cf. 1]. E.g. if one is interested in evaluating information searching, it would be beneficial, if the system could provide her with sub-concepts like process evaluation and product evaluation, or efficiency and effectiveness. This is likely to help in specifying her information need.

The discussion above hints that structuring searching conceptually both in expressing information need and formulating search would be beneficial in exploratory searching. Consequently, exploratory search systems should provide actors with tools that help them conceptually map and structure the topic of their information need and formulating search tactics accordingly. Distinction between expressing information need and selecting search terms is an analytical one. In searching actors engage in both activities simultaneously. Therefore, it is important that querying facility in exploratory search systems helps searchers to explicitly structure their search formulation conceptually, and that it also provides them with a tool for specifying their search terms e.g. into hierarchically narrower terms.

6. CONCLUSIONS

We have explicated the major stages in exploratory searching, and suggested some criteria for evaluating exploratory search systems especially at the early stages in exploration. Our ideas extend the evaluation paradigm from a focus only on the output of the system onto the whole search process.

In evaluation of systems and services, the point of departure is to assess to which extent the goal of the system is achieved [11]. It seems that one of the major limitations in interactive evaluation has been to focus on only one goal of information retrieval, optimizing the output in terms of the number of relevant items retrieved. Although this is a necessary condition for a successful information retrieval, the search process is in its turn a necessary condition for a good retrieval result. Therefore, in evaluating search systems, it is important to assess to what extent the search process variables reach their objectives, and through those objectives contribute to retrieval effectiveness [cf. 2].

A major implication of our analysis is that in system evaluation it is critical to define the goals of the tools assessed in improving human performance in information searching. Without reflecting and defining the objectives of the system it is difficult to infer appropriate evaluation criteria. This is of special importance in evaluating tools for supporting exploratory searching. In exploration it is as critical to support the search process and structuring of the topic, as it is to retrieve relevant items [13; 15]. Surveying information space with appropriate tools is likely to contribute to structuring the topic and expressing the information need as search terms leading to growing understanding of the topic and as a consequence, more useful search results. Thus, the exploratory search system should help the user to attain several

search goals from explicating the information need to finding documents conceptually matching that need [8; 15]. It is important to assess to which extent the system attains these various goals.

We have excluded from the analysis of search process the evaluation of search results. Our next step is to analyze assessing retrieval results as conceptual exploration. We seek to infer measures of success in this activity understood as conceptual correspondence between searcher's conceptual construct and author's conceptual construct in the document retrieved [12]. It is possible to extend this conceptualization also to cover the benefits information retrieval systems produce to task performance, the ultimate goal of these systems. By representing task as a conceptual construct it is possible to relate task performance process to search process as we have described it in this study. This procedure implies that we are able to model and assess how the search process and the use of information in the items retrieved contribute to task performance.

7. REFERENCES

- [1] Bates, M.J. 1979. Information search tactics. *JASIS* 30, 205-214.
- [2] Belkin, N. J. 1993. Interaction with texts: information retrieval as information-seeking behavior. In *Information Retrieval '93*. Konstanz: Universitätsverlag Konstanz, 55-66.
- [3] Gallagher, S., Sher, B., Stepien, W. and Workman, D. 1995. Implementing problem-based learning in science classrooms. *School Sci. and Math.* 95, 136-140.
- [4] Ingwersen, P. and Järvelin, K. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, Dordrecht.
- [5] Kintsch, W. 1998. *Comprehension. A paradigm for cognition*. Cambridge UP, Cambridge.
- [6] Kuhlthau, C. 1993. *Seeking Meaning*. Ablex, Norwood, N.J.
- [7] Lancaster, W. and Warner, A. 1993. *Information retrieval today*. Information Resources Press, Arlington, VA.
- [8] Marchionini, G. 2006. Exploratory search. *Comm. ACM* 49, 41-46.
- [9] Pennanen, M. and Vakkari, P. 2003. Students' conceptual structure, search process and outcome while preparing a research proposal. *JASIST* 54, 759-770.
- [10] Rebich, S. and Gautier, C. 2005. Concept mapping to reveal prior knowledge and conceptual change in a mock summit course on global climate change. *J. Sci. Educ.* 53, 355-365.
- [11] Rossi, P., Lipsey, M. and Freeman, H. 2004. *Evaluation. A systematic approach*. 7th ed. Sage, Thousand Oaks.
- [12] Vakkari, P. 1999. Task complexity, problem structure and information actions. *IP&M* 35, 819-837.
- [13] Vakkari, P. 2001. A Theory of the task-based information retrieval process. *J. Doc.* 57, 44-60.
- [14] Vakkari, P. 2003. Task-based information searching. In *ARIST 37* (B. Cronin, Ed.) Information Today, Medford, NJ., 413-464.
- [15] White, R.W. and Roth, R. 2009. *Exploratory search*. Morgan & Claypool.