Information Search Patterns in Complex Tasks

By Pertti Vakkari

Abstract

The aim of my talk is to analyze information search process in complex tasks[1]. By task I mean larger tasks, which lead people to engage in search tasks for finding information to advance those tasks. Search process consists of activities from query formulation to working with sources selected for task outcome. I approach task performance from the cognitive point of view conceptualizing it as changes in knowledge structures. These structures consist of concepts and their relations representing some phenomenon. I analyze how changes in knowledge structures are associated to query formulation and search tactics, selecting contributing sources and working with sources for creating task outcome. As a result, I suggest hypotheses concerning associations between changes in knowledge structures and search behaviors. I also present some ideas for success indicators at various stages of search processes.

Introduction

Our understanding of search processes triggered by complex tasks is limited (Belkin et al 2017). It is not well known how the information search process evolves during task performance and how search behaviors vary by task process. How do changes in information needs reflect in search formulations and tactics, in selecting contributing sources and interacting with sources for task outcome? A better understanding of these questions helps in identifying success criteria for various parts of search process. The results contribute also to designing support tools for complex search tasks.

The aim of my talk is to analyze information search processes in complex tasks. By task I mean larger tasks, which lead people to engage in search systems for finding information to advance these larger tasks (Vakkari 2003). Larger tasks may be e.g. work tasks or activities related to hobbies. Task complexity refers to the *apriori* determinability of task outcomes, process and information requirements (Vakkari & Kuokkanen 1997). Thus, actors have a vague notion in advance of what kind of information is needed for the task, what kind of procedure is required for creating task outcome, and what is the content of this outcome. Typically, complex tasks are vague and ill-structured and generate several successive search sessions. In my talk I consider only complex, information intensive tasks. In this context, I am interested in tasks with small extent of *apriori* determinability. Naturally, when a task proceeds, actors' understanding concerning it grows reducing *apriori* determinability.

[1] This text is an updated and modified version of my keynote presentation at CHIIR'18 Conference at Rutgers University, 12.3.2018.

Search process consists of activities from query formulation to working with sources selected for task outcome (Järvelin et al. 2015). I approach task performance from the cognitive point of view conceptualizing it as changes in knowledge structures (Brookes 1980, Ingwersen 1992). These structures consist of concepts and their relations representing some phenomenon. I analyze how changes in knowledge structures are associated to query formulation and search tactics, selecting contributing sources and working with sources for creating task outcome. As a result, I suggest hypotheses concerning associations between changes in knowledge structures and search behaviors. I present also some ideas for success indicators at various stages of search processes.

Although I am speaking about tasks, I mean task topic, what the task is about, i.e. it's subject matter. I focus on the declarative knowledge what is needed to perform the task, not so much on procedural knowledge for realizing the task.

Knowledge structures

People engage in information systems when they do not have sufficient knowledge for performing a task. This lack of knowledge can be called e.g. ASK (Belkin 1980) or uncertainty (Kuhlthau 1993). It can be called as "information need" as well. Human knowledge can be described as knowledge structures, which consist of concepts and their relations representing some phenomenon like "tide in Gran Canarias" or "evaluation of information retrieval". These structures can be called e.g. as mental models or schemas (Robertson 2001). Knowledge structures change when people encounter information in various contexts. One needs information inputs for changing or reinforcing current knowledge structures. Typically, this implies working with information acquired for adjusting it to existing or adapted knowledge structures. This means that the development of new knowledge is based on prior knowledge.

Bertram Brookes' (1980) famous fundamental equation for information science represents the growth of human knowledge as changes in knowledge structures.
$K(S)+\Delta I \rightarrow K(S+\Delta S)$.
Knowledge structures K(S) are changed by information input ΔI into a new modified state of knowledge K(S+ΔS).

In learning theories cognitive changes have been typically categorized as assimilation and accommodation (Zhang & Soergel 2014). Assimilation means addition of information into existing knowledge structures, whereas accommodation means modifying or changing existing knowledge structures. The latter refers to adding or removing concepts and their relations in the knowledge structure. The former means that the conceptual construct the knowledge structure consists of, does not change, but merely new instances are added in to concepts.

Accommodation is refined according to the degree of structural change (Zhang & Soergel 2014). Tuning or weak revision does not include replacing concepts or relations between concepts in the structure, but merely tuning of the scope and meaning of concepts and their relations. This may include e.g. generalizing or

specifying a concept. Restructuring means changing and replacing concepts and their relations in the knowledge structure.

Information searching in complex tasks is generated in situations, when actors' mental models lack concepts and relations between concepts for accurately representing task topic. They have insufficient concepts and insufficient relations in their knowledge structure (Vakkari 1999). I claim that exploring concepts and their relations is a major characteristic of information search for complex tasks, in the beginning of the process, in particular. Thus, we may say, that actors search information for obtaining concepts and their relations in order to understand, structure and represent their task topic more validly for proceeding in task performance. Based on the results of searches and their own reflections actors create little by little a more coherent understanding of task leading to task outcome.

Search process as a part of task process

For the analysis, we divide the search process into four main stages: search formulation, selecting sources, interacting with sources and synthesizing and presenting. Search formulation can be divided into expressing information need, term selection and query formulation

Expressing information need for search formulation and executing the query is the starting point of a search process (Marchionini 1995). Search formulation includes the articulation of search terms and the use of search operators if there are any. It includes also the search tactics used for reaching the search goal. The choice of search terms is a reflection of the searcher's understanding of her information need, what is insufficient in her knowledge structure. In practice, a searcher expresses her information need by selecting search terms and search tactics, i.e. by formulating and reformulating queries. The searcher likely reflects and reforms her understanding of the topic, and consequently search terms for querying.

In selecting sources searchers explore document surrogates in the result list or full documents to assess the value (relevance, usefulness) of the documents found to satisfy their information needs, i.e. to reduce the insufficiency of their knowledge. Especially, various document information elements are used for inferring the value of a source (Wang & Soergel 1998). The aim of this interaction is to select sources, which likely contribute to the activity triggering the search so that they are available for immediate or later use (Vakkari 2003). Thus, I expect that users' criterion for selecting sources is the contribution to the sub-task at hand, how the information obtained supports it in proceeding in task performance. This criterion can be called usefulness (Belkin et al. 2009, Vakkari 2003). Usefulness is more than topical relevance, because information obtained should be utilized for some purpose.

Interacting with sources is the phase when the searcher familiarizes with the sources to realize the task for which they were selected. It may include reading, making notes, extracting pieces of information, combining and organizing

information and outlining (Kuhlthau 1993). The aim of this activity is to analyze, categorize and structure information in documents for preparing the final outcome of the task. The borderline to synthetizing and presenting information for task outcome is blurred, because it is impossible to determine exactly when e.g. combining and outlining information by writing for the task outcome becomes synthetizing or presenting. It is likely that in a complex task when the actor is not familiar with it, the characteristics of interacting with sources can be equated to Kuhltau's ISP model's stages prior to information collection (Järvelin et al. 2015) Anyway, synthetizing and presenting include preparing the final outcome of the task. The form of the outcome depends naturally on the nature of the task. It can be a financing decision, a newspaper article, or a greater understanding about some societal issue.

Search formulation

In the beginning of search a person seeks to articulate her information need into search terms. Her understanding of the topic is vague. Her knowledge structure is insufficient for expressing exactly what she wants to know. It lacks sufficient concepts and their relations for representing the topic in order to proceed in the task. Therefore, she is searching for how to structure the topic, actually conceptual structures (Vakkari 1999, Qu & Furnas 2008) for constructing a focus. There are seldom ready-made conceptual structures available, but the person has to create them incrementally. There is typically not a single right solution, but many options available for how to proceed and how to specify the structure.

It is likely that in the search formulation stage the person seeks to explore the conceptual space in documents related to her own conceptual structure of the topic. This usually starts by articulating first her conceptual understanding as a query, and then extending it by additional concepts or specifying it by reformulating queries. The aim of this activity is to find information that helps in structuring a necessary conceptual representation of task topic. The representation changes from vague to precise by narrowing its scope (Kuhlthau 1993, Vakkari 2016). This typically includes introducing new concepts, relating them to old ones, and specifying old concepts (Robertson 2001). If this is the case, then success in explicating the conceptual structure can be measured by observing the changes in the number and specificity of concepts in the representation of task topic. Specificity refers to the number of sub-concepts and the specificity of concepts in the knowledge structure (Vakkari 2010; cf. Lancaster & Warner 1993). Measures and criteria for success are presented in table 1.

Table 1. Measures of success in articulating information need (conceptual structure).

	Measure	Criteria of success
Concepts	Exhaustivity = the number of concepts articulated	An increase in the number of concepts articulated in *ASK* An increase in the proportion of concepts articulated in *given*

		search tasks
	Specificity = the proportion of sub-concepts of all concepts articulated	An increase in the proportion of sub-concepts of all concepts articulated
Terms	Exhaustivity and specificity like in concepts	
	$Specificity_2$ = the number of hyponyms per concept	An increase in the number of hyponyms per concept
	Synonyms = the number of synonyms	An increase in the number of synonyms (per concept)
	Query length = the number of terms	An increase in the number of terms

So far algorithms have not been very reliable at identifying concepts and their expressions either in queries or in texts. It is demanding to automatically recognize which terms belong to the extension of a certain concept. Therefore, knowledge structures are difficult to identify both on conceptual level and term level. However, there have been attempts to cluster terms for query expansion by their semantic similarity and call these clusters facets, aspects or concepts. E.g. Kong & Allan (2014) have generated query facets (i.e. concepts) from search results. Terms belonging to query facets are generated from top-ranked results. A query facet consists of coordinate terms, i.e. terms that share a semantic relationship by being grouped under a more general hypernym. Thus, they are hyponyms of the facet. These query facets can be used to reformulate queries and re-rank search results.

If it is possible to distinguish terms belonging to query facets (concepts), then it would be possible to calculate the number of specific terms (hyponyms) per concept or per query. An increase in the number of specific terms per concept or per query would indicate an increase in the specificity of knowledge structure.

The exploration of the conceptual space in documents occurs often by using vary tactics, i.e. replacing existing concepts by new ones in a query (Bates 1979, Vakkari et al. 2003). One may be interested e.g. in the relations of information searching and tasks, and explores this conceptual space first by adding to the query "information searching, task" new concepts like "process" and then replacing it by the concept "support tools".

Pennanen & Vakkari (2003) showed that in the beginning of a writing task, an increase in the number of query terms and in the number of terms extracted from results significantly increased the use of vary tactics, which in its turn increased the number of useful references found. On the other hand, the more students were able to express the concepts in their knowledge structure concerning the topic by query terms, the more useful references they found. The coverage in query terms had a direct effect on the number of useful references regardless the use of terms or tactics. This suggests that when restructuring knowledge for formulating a focus, searchers ability to articulate their information needs as query terms is crucial for search success regardless how

they use search tactics or terms. Thus, for exploring the conceptual space in documents, searchers need support in formulating vary tactics, i.e. finding good candidates for replacing existing query concepts by new, interesting ones, and also help in recognizing potential search terms in search results.

Vakkari et al. (2004) have shown that the number of relevant documents retrieved increased, when the share of facets in a topic articulated by query terms (i.e. exhaustivity) increased. Query exhaustivity was the strongest predictor of search success. This enhances the conclusion that users' ability to express their knowledge structure representing information need by query terms is a crucial condition for search success. This conclusion is repeated over and over again in the textbooks of information retrieval (e.g. Lancaster & Warner 1993).

Exploring by vary tactics may reveal new concepts and conceptual relations which facilitate articulating and restructuring the searcher's knowledge structure. New concepts are potentially found, when document surrogates in the result list and full documents are explored (Pennanen & Vakkari 2003). Titles, abstracts, and text passages browsed include possible ideas for restructuring conceptual understanding of the topic. Restructuring refers to changes in concepts and in relations between concepts, in particular.

This behavior resembles intrinsically diverse search tasks, in which a user requires information about multiple, different aspects of the same topical information need. These often lead to longer and more complex search sessions, which can be characterized as exploratory (Raman et al. 2013). By detecting intrinsically diverse search sessions and identifying initiator queries in these sessions, it is possible to produce related queries representing new aspects of the topic, which the user is likely to issue later in the session (Raman et al. 2013). The new queries are specifications of the original query. These query suggestions resembling vary tactics likely help searchers to explore the conceptual space of interest and even to structure it to some extent.

When a person has revised her conceptual structure, formulated a focus for her topic according to ISP model, it includes the major concepts and their relations (Vakkari 2016). Kuhlthau (1993) remarks that after this stage the interaction between user and system functions most efficiently due to user's ability to articulate more focused what information is needed. The conceptual structure however, likely requires tuning to better fit with the aim of her task. Tuning typically includes mostly limiting the scope of the concepts in the knowledge structure or its applicability area. In search formulation this means using more specific search terms to present the concepts in queries (Vakkari 2001, Vakkari et al. 2003, Wang 1997). It is likely that in tuning also factual information is sought to support the arguments developed (Vakkari 2000). Tuning leads to assimilation, i.e. to populating concepts with new instances or facts. When the conceptual structure is stable, synonyms are typically used in query formulation to represent concepts (Vakkari 2001).

When users' conceptual structure concerning the topic is stabilized, it is evident that they abandon vary tactics and use exhaust tactics, i.e. seek to express the conceptual structure in its totality as a query. The concepts are expressed by more specific keywords and by synonyms (Vakkari 2001, Vakkari et al. 2003; Wang 1997). The number of query reformulations as well as time use in sessions decreases. In this phase it is likely that query suggestions for diversifying search results attract users less and result also in less successful searches. Instead, typical relevance feedback evidently functions best. In multi-session searches, a decrease in the number of query reformulations and in session time combined with longer and more specific queries indicate that users' knowledge structure, i.e. information need has stabilized.

Source selection

In source selection user's behavior typically varies with the changes in knowledge structures. In the beginning user's conceptual structure concerning a topic is vague. She is not able to articulate clearly what kind of information is needed due to undifferentiated relevance criteria (e.g. Kuhlthau 1993). She has difficulties in recognizing useful sources because her lack of firm ideas how to structure her topic. In many cases she is not able to say whether a source would contribute to her task or not. It depends on how she will shape the focus of her task. Therefore, in addition to useful sources, also those that may be of use will be selected for further exploration (Vakkari & Hakala 2000; Serola & Vakkari 2005). In this stage, SERPs are likely inspected relative systematically in an ascending order of relevance due to uncertainty of what is contributing. Searchers aim at maximizing recall implying a thorough examination of SERP. Backward skips on the result list are likely also common. Exploration continues to additional SERPs, because several links may provide useful information due to the inability to restrict the scope of topic. Dwell time in each snippet is on average long (cf. Gdwizka 2014, Smucker & Jethani 2013. The click trough rate of links is high and the proportion of selected sources of clicked ones is great. Due to the uncertainty concerning the structuring of topic a lot of sources is selected for further exploration to construct a focus.

When users' conceptual structure stabilizes, they are more certain to distinguish useful sources from useless ones. This reflects also in SERP behavior. Result list is explored in an ascending order with less backward skips. Dwell time in each document surrogate is on average shorter compared to earlier stages. Fewer SERPs are explored (cf. Zhang et al. 2015) and fewer results are clicked on a SERP. The click trough rate is lower and the proportion of selected sources of clicked ones is also higher compared to pre-focus stages (Kelly & Cool 2002). However, SERP behavior is also influenced by users improved ability to formulate pertinent queries, which produce high quality result lists. For instance, high precision may result in more frequent clicks and higher click trough rate compared to lower precision. This likely somewhat decreases the difference in SERP behavior compared to pre-focus stage. On the other hand, the marginal utility of documents on SERP evidently decreases during task performance, because it is likely that some useful documents occur repeatedly on the lists. This strengthens the decreasing click trough rate.

It is likely that the assessment of the expected usefulness of sources resembles relevance assessment of documents. Gdwizka (2014) has shown that it takes more cognitive effort, and consequently time, to judge partially relevant documents compared to relevant or non-relevant ones. The results in Smucker & Jethani (2013) correspond the previous finding, when they found that the more difficult it was to assess the relevance of a document, the more time it took from the assessors. Borderline cases took most time, while non-relevant cases required less time compared to relevant ones. Thus, we may suppose that decision time in opened sources varies according their expected usefulness. Clearly useful and useless cases do not require as much time as borderline cases to make a decision about the selection. The relation between decision time and usefulness is curvilinear. The documents with the longest decision time are not the most useful ones, but somewhat or partially useful. In general, decision time conceived as linear is not a reliable indicator of document usefulness or relevance.

Dwell time on documents clicked is considered as a major indicator of their worth be it relevance or usefulness (Fox et al. 2005; Liu et al 2010; Liu et al. 2012). The longer the dwell time, the more useful the document. In studies of information retrieval users are typically asked to assess topical relevance of documents for an information gathering task, i.e. subject search, which do not require the use of information, but just judging its topicality. Therefore, dwell time in those cases refers to the time needed for deciding about topicality. In some cases, search tasks require also using information in documents retrieved e.g. for answering a question or writing a text. Then dwell time includes both the decision time to use the document and the time for the actual use of information in the document.

The allocated time for a search task in most cases is very limited implying that information is used during a search session. Answering questions or writing a short text means that during the session users tend to keep useful documents open so long that is needed to perform the action instructed (cf. Liu and Belkin 2010). There is no time to close and reopen documents repeatedly. Liu and Belkin (2010) have shown, that in search sessions for writing a text within limited time, increase both in display time and dwell time was positively associated with usefulness.

If there is plenty of time to perform a search task requiring information use, the role of dwell time indicating document usefulness likely depends on the extent of task outcome like the volume of text to be created. If task outcome is limited e.g. to answering a couple of questions or writing a short text, dwell time as usefulness indicator behaves like in tasks with time limits. It is supposed that people perform the task in one session without breaks. Therefore, an increase in dwell time reflects increasing usefulness. However, if task product is extensive like writing a text of 5000 words based on search results, it is likely that task is performed during several sessions (cf. Hagen et al. 2016). It is also likely that users first gather sources and information, and then begin to work with sources, or that this pattern repeats until the outcome is finished (Hagen et al 2016; He et

al. 2008). If this holds, users assess first the usefulness of documents, and save the useful ones for the coming use. When they have gathered enough information for the current sub-task, they begin to work with the sources for the outcome. Vakkari et al. (2019) have shown that in writing a long text, the usefulness of documents is negatively associated with dwell time. In this case, judging usefulness resembles relevance assessment as claimed earlier. It is shown that deciding the relevance of borderline cases (Smucker & Jethani 2013) or partially relevant items (Gwidska 2014,) takes significantly more time and effort compared to relevant and non-relevant items.

If information is selected from documents for later use, the volume of text obtained is larger in pre-focus stage than post-focus stage due to users limited ability to differentiate between useful and not useful information (e.g. Kuhlthau 1993, Wang & White 1999). Uncertainty about the final focus of the topic increases the amount of collected information per document. With the stabilization of conceptual structure, the volume of information extracted per document decreases, while the proportion of information used from extracts in the end product increases. Users become increasingly selective in the use of information when their knowledge about the topic grows.

When users have selected sources to work with, it is likely that the more time they devote to a document, the more useful it is. The more and the more important information a document provides to be transformed into the evolving text, the more time is allocated to it (Vakkari et al. 2019). The degree of usefulness depends to what extent a document contributes to the expected outcome. The nature of contribution varies with the changes in the knowledge structure concerning the task (Wang & Soergel 1998, Vakkari 2000). Vakkari (2000) has shown that in the beginning of writing a research proposal, students were looking for general background information, theories and conceptualizations, while later they sought to find more specific information types like facts and empirical results. When users are shaping the structure of task topic, documents, which provide ideas for this modification are more useful than those which contain information e.g. about particular facts. Ideas for forming a focus decrease uncertainty by reducing the number of alternative conceptualizations, while facts reduce uncertainty concerning specific information. The former conveys notably more information compared to the latter. It is also likely that forming a conceptual structure from information in documents requires much more cognitive processing and consequently time compared to identifying a fact for a specific issue. E.g. Russell et al. (1993) suggest that the major cost of sense-making is associated to forming a structural representation of a topic. Consequently, we may suppose that activities and information that support in shaping a structure produce the greatest gain as well.

The usefulness of documents

When task outcome, e.g. a text, changes due to the new information from the documents selected, the usefulness of documents (or the text passages obtained from documents) may be estimated by each contribution to the changes in text at

that point in time. It is assumed that the evolving text reflects users' knowledge concerning task topic. The growth of knowledge consists of an increasing number of concepts and their relations, and of the specificity of conceptual structure (Vakkari 2010). Thus, an increase in the number of concepts and in the number of interrelations between these concepts, as well as in the specificity of concepts can be used the criteria for the advancement of text. By comparing the conceptual structure of the evolving text before and after information from a document is added to the text, it is possible to describe the occurred conceptual change. An increase in the number and specificity of concepts reflects the advancement of text. The extent to which these new concepts can be identified from a document used represents its usefulness. It is possible to give weights to the concepts based on their role. A new concept introduced in the conceptual structure typically limits the extension (i.e. specifies) of the topic more than specifying an existing concept. This is comparable to an increase in the exhaustivity of a query, i.e. adding a facet. Therefore, one may assign a greater weight to a new concept compared to a specified old one. In addition, one may weight the new concepts (actually, the expressions of those concepts) by their occurrence frequency in the evolving text. The more occurrences, the more important the concept is for the topic presented in the text. The sum of weights of new concepts from the document in the text represents the usefulness of that document.

Currently, automatic identification of concepts in texts does not work well. Concepts are represented in queries and documents by terms. The idea above to measure the usefulness of a document can be applied also on the level of terms. It is based on the idea that task can be represented as a collection of terms from the text a user is currently working with (Budzik & Hammond 2000). The new terms in the evolving text that can be identified in the document used reflect its usefulness in construing the text. The more frequently a new term from a document occurs in text in progress, the greater its contribution. This importance of a term can be weighted e.g. by its frequency or emphasis in the text (Budzik & Hammond 2000). The usefulness of a document could be represented simply by summing up the weights of new terms in the text. This is a rough measure of information that is derived from a document to the text. Vakkari et al. (2019) have developed this idea into an equation for measurement.

Information about the new terms in the text from a document can be applied for usefulness feedback to re-rank the results by giving a greater weight to these new terms in the documents. In the same way, it can be applied in query reformulation by giving a greater weight to the new terms. This procedure can also be used for query diversification by seeking to identify queries that are associated to the new terms, in particular.

The idea of using task profiles for personalization based on texts obtained from documents opened is not new (Budzig and Hammond 2000, Ahn et al. 2008, He et al. 2008). Budzig and Hammond (2000) introduced Information Management Assistant (IMA), which anticipates users' information needs while they are working with text. It extracts words from the evolving text for forming a task model, which is used for automatic querying triggered by an anticipator

component, or enriching users' explicit queries by words in task model. The evaluation of IMA showed that it produces more useful documents compared to expert searchers or a baseline system.

The influence of search process on task outcome

So far, I have dealt with how users' cognitive representation of a progressing task is associated to various parts of search process from querying to working with sources found. However, I have not yet discussed how search process affects task outcome. If the ultimate goal of information search is to contribute to task outcome, it would be interesting to know to what extent search process has an effect of task outcome.

There are only a limited number of empirical studies focusing on the associations between search process variables and task outcome (Järvelin et al. 2015). The topic is demanding, not least due to the longitudinal nature of performance in complex tasks and associated searching. Among the first to empirically study relations between searching and task performance were Kuhlthau (1993), Hersh (Hersh et al. 1996) and Wildmuth (Wildemuth et al. 1995).

The findings on the relationship between searching and task outcome are highly varied, in part due to the differences in tasks like question answering or essay writing: precision and recall either have no impact (Hersh 2003), have a positive effect (Wildemuth et al. 1995}, or a negative effect (Vakkari & Huuskonen 2012} on task outcome. Results on the contribution of search process variables to task outcome vary as well: effort in the search process either improves task outcome (Vakkari & Huuskonen 2012}, or has no effect on it (Bron et al. 2012, Liu et al. 2012). However, some studies seem to hint that the more effort is invested in reading documents (Collins-Thompson et al. 2016), or in writing compared to searching, the better the resulting essay (Liu & Belkin 2012, Vakkari & Huuskonen 2012). Thus, while most of the results on the relations between search process and task outcome are contradictory, there is an emerging trend hinting that the more users were able to focus on results inspection instead of querying and on working with information in documents found instead of result inspection, the better this seemed to be to the process of construction for proceeding in the task (cf. Liu & Belkin 2012, Butcher et al. 2011, Vakkari & Huuskonen 2012).

Conclusions

I have analyzed patterns of information search processes as user's knowledge structure develops in complex tasks. I have treated searching as a part of a larger task performance and conceptualized it as changes in knowledge structures. This notion implies that in addition of querying and result examination, search process consists of working with sources for creating a task outcome. This is consistent with Kuhlthau's (1993) suggestion that in addition to accessing information, actors need help also in making sense of search results for progressing in their task. Even if we would restrict the aim of search systems

only to support querying and search result examination, knowing also how users work with information when interacting with sources is beneficial for system design. It is evident that users' understanding of their task changes when they work with sources found, and this affects their querying and result examination behavior (Vakkari 2001, Liu et al. 2013). Therefore, broadening the scope of our studies to include also information use in documents found would deepen our knowledge about querying and result examination, not to speak about the whole search process. There have also been some interesting attempts to develop integrated tools for both accessing and making sense of information (e.g. Budzig and Hammond 2000, Ahn et al. 2008, He et al. 2008), but they seem to form a minor side-path alongside the main road of studies on information retrieval.

I also suggested hypotheses about the associations between changes in knowledge structures and search behaviors during task performance process. From these hypotheses it is also possible to derive criteria for success in various parts of search process like I have done in a study concerning learning while searching (Vakkari 2016). The empirical grounding of these hypotheses however, is scanty. It consists of findings of a few longitudinal studies on information searching (mostly studies by Kuhlthau, Vakkari, Liu and Belkin, Wang, and Wildemuth) supported with findings of some studies in pedagogics (e.g. Butcher et al. 2011, Cho et al. 2017) and on the role of topical knowledge in searching (e.g. Zhang et al. 2015). All this is framed with ideas and findings mostly in cognitive psychology (e.g. Robertson 2001) or organization science (e.g. Campbell 1988). Therefore, I would hope for more studies seeking to test the validity of these hypotheses. They would show to what extent these generalizations hold true and should they be modified. These results would also provide information for refining the criteria of search success.

Exploring information searching as a part of a larger task typically presupposes taking into account information use in documents retrieved in some form. Longitudinal experiments observing participants at several points in time, i.e. during several search sessions, are demanding and costly, although they provide lots of important information not available in experiments lasting one search session. Solutions for how to deal with information use related to larger tasks within a search session would be valuable.

When using the most complex of these search tasks for evaluating search systems, in addition to perceived usefulness of documents, one has to create new measures for assessing the value of information in documents to solve the task given. It would be interesting to know in which way or which elements in the documents contributed solving the task. One direction to look for ideas is studies on information problem solving in pedagogics (Cho et al. 2017, Walhout et al. 2017), while they seem to use search tasks comparable to those based on the framework presented.

A trend called searching as learning seeks to analyze how people learn while they search or how search process is associated with learning outcomes (Vakkari 2016). Two recent special issues of this theme, one in Journal of Information Science (vo. 42:1, 2016) and other in Information Retrieval Journal (vol. 20:5,

2017) indicates an increasing interest in the role of information search in learning. Many papers in these issues deal with searching and using information for learning purposes and seek to measure learning gain from the use of information (e.g. Syed & Collins-Thompson 2017, Zhang & Soergel 2016,). They lack indicators of usefulness, which are objective not subjective.

I proposed a measure of document usefulness based on its contribution to the task at hand. A task can be presented as an evolving text for which documents are retrieved. The contribution of a document to the text can be determined by 1) identifying the new terms in the text from the document used, 2) by weighting each new term by its frequency and emphasis in the text, and 3) by normalizing this term density by the number of words in the source, and finally by summing up the normalized weights of each term. This measure indicates the contribution of the document to the evolving text. It can be called usefulness, but I would prefer information gain, because it represents approximately how much information is derived from the document to the text. The measure is objective, and it can be generated from the log data without interfering with user activities given that the system collects the required data. Compared to traditional effectiveness measures based on perceived topicality or usefulness, it represents closer the contribution an information item provides to the text in progress. This measure could be applied also when systems are evaluated by using complex search tasks, which require processing and manipulating information in documents for solving the search tasks. Vakkari et al. (2019; 2021) have developed objective measures for the usefulness of information in documents and applied these measures for analyzing how search process is associated to to the usefulness of information and task outcome.

In all, I believe that treating information search as a part of larger tasks, and conceptualizing task performance as changes in knowledge structures provide new fertile ideas and research problems, which advance the growth of knowledge in our field of research.

References

J-W. Ahn, P. Brusilovsky, D. He, J. Grady, and Q. Li. (2008). Personalized web exploration with task models. *In Proc. WWW 2008*. ACM, New York, NY, 1-10.

Bates M.J. (1979). Information search tactics. Journal of the American Society for Information Science 30(4): 205-214.

Belkin N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information and Library Science 5: 133-143.

N.J. Belkin, M. Cole, and J. Liu. (2009). A model for evaluating interactive information retrieval. In *SIGIR Workshop on the Future of IR Evaluation, July 23, 2009, Boston*.

Belkin, N. et al. (2017). 2nd Workshop on Supporting Complex Search Tasks. In *Proceedings of CHIIR'17*.

Bron M, van Gorp J, Nack F, de Rijke M, Vishneuski A and de Leeuw S. (, 2012). A subjunctive exploratory search interface to support media studies

researchers. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12). New York, NY: AC, 425-434.

Brookes, B. (1980). The foundation of information science. Part I: Philosophical aspects. Journal of Information Science 2:125-133.

Budzik, J. & Hammond, K. (2000). User interactions with everyday applications as context for just-in-time information access. In Proc. UIU'00. New York, NY: ACM, 44-51.

Butcher KR, Davies S, Crockett A, Dewald A and Zheng R. (2011). Do graphical search interfaces support effective search for and evaluation of digital library resources? In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries (JCDL '11). New York, NY: ACM, 315-324.

Campbell, D. J. (1988). Task complexity: a review and analysis. *Academy of Management Review,* 13, 40-52.

Cho, B-Y., Woodward, L., Li, D. and Barlow, W. (2017). Examining adolescents' strategic processing during online reading with a question-generating task. American Educational Research Journal 54(4): 691-724.

K. Collins-Thompson, S.Y. Rieh, C.C. Haynes, R. Syed. (2016). Assessing learning outcomes in web search: A comparison of tasks and query strategies. *Proc. of the 1st CHIIR Conf.*, pages 163-172.

S. Fox, K. Karnawat, M. Mydland, S. Dumais, , and T. White. (2005). Evaluating implicit measures to improve web search. *ACM TOIS* 23, 2, 147–168.

J. Gwizdka. (2014). Characterizing relevance with eye-tracking measures. In *Proceedings of the 5th Information Interaction in Context Symposium*. ACM, 58– 67.

M. Hagen, M. Potthast, M. Völske, J. Gomoll, and B. Stein. (2016). How Writers Search: Analyzing the Search and Writing Logs of Non-fictional Essays. In *Proc CHIIR'16*, Diane Kelly, Rob Capra, Nick Belkin, Jaime Teevan, and Pertti Vakkari (Eds.). ACM, 193–202.

D. He, P. Brusilovsky, J. Ahn, J. Grady, R. Farzan, Y. Peng, Y. Yang, and M. Rogati. (2008). An evaluation of adaptive filtering in the context of realistic task-based information exploration. *IP&M 44*: 511-533.

W. Hersh. *Information retrieval: A health and biomedical perspective* (2nd ed.). (2003). New York, Springer.

Hersh W, Pentecost J, and Hickam D. (1996). A task-oriented approch to information retrieval. Journal of the American Society for Information Science 47(1): 50-56.

Ingwersen, P. (1992) *Information Retrieval Interaction*. Taylor Graham: London.

K. Järvelin, P. Vakkari, P. Arvola, F. Baskaya, A. Järvelin, J. Kekäläinen, H. Keskustalo, S. Kumpulainen, M. Saastamoinen, R. Savolainen, E. Sormunen. (2015). Task-based information interaction evaluation: The viewpoint of program theory. *ACM Trans. Inf. Syst.* 33(1):3:1–3:30.

Kelly, D. & Cool, C. (2002). The effects of topic familiarity on information search behavior. In Proc. JCDL'02. New York, NY: ACM, 74-75.

Kong, W. & Allan, J. (2014). Extending faceted search to the general web. In Proc. CIKM'14. New York, NY: ACM, 839-848.

Kuhlthau C. Seeking Meaning. Norwood, N.J.: Ablex, 1993.
Lancaster W. and Warner A. (1993). Information retrieval today. Information Resources Press: Arlington, VA.
C. Liu, J. Gwizdka, and J. Liu. (2010). Helping identify when users find useful documents: examination of query reformulation intervals. In *Proceedings of IIiX'10*. 215–224.
C. Liu, N.J. Belkin, and M.J. Cole. (2012). Personalization of search results using interaction behaviors in search sessions. In *Proc. SIGIR'12*. ACM, 205–214.
J. Liu and N. J. Belkin. (2010). Personalizing information retrieval for multi-session tasks: the roles of task stage and task type. In *Proc. SIGIR'10*. ACM, 26–33.
Liu, J and Belkin, N.J. (2012). Searching vs. writing: Factors affecting information use task performance. In Proceedings of the American Society for Information Science and Technology Vol. 49, 1, 1–10.
Liu, J., Belkin, N.J. Zhang, X. and Yuan, X. (2013). Examining users' knowledge change in the task completion process. IP&M 49:1058-1074.

Marchionini G. (1995). Information seeking in electronic environments. Cambridge University Press.
Pennanen M. and Vakkari P. (2003). Students' conceptual structure, search process and outcome while preparing a research proposal. Journal of the American Society for Information Science and Technology 54(8): 759-770.
Qu Y and Furnas G. (2008). Model-driven evaluation of exploratory search: A study under sensemaking framework. Information Processing & Management 44(2): 534-555.
Raman, K., Bennett, P. and Collins-Thompson, K. (2013). Toward whole-session relevance: Exploring intrinsic diversity in web search. In Proc. SIGIR'13. New York, NY: ACM, 463-472.
Robertson S. (2001). Problem solving. Hove: Psychology Press.
Russell, D., Stefik, M., Pirolli, P. and Card, S. (1993). The cost structure of sensemaking. In Proc. CHI '93. New York, NY: ACM, 269-276.
Serola, S. & Vakkari, P. (2005). The anticipated and assessed contribution of information types in references retrieved for preparing a research proposal. *Journal of the American Society for Information Science* 56(4): 373-381.
M.D. Smucker and C. Jethani. (2012). Time to judge relevance as an indicator of assessor error. In *Proc. SIGIR'12*. ACM, 1153–1154.
Syed, R. & Collins-Thompson, K. (2017). Retrieval algorithms optimized for human learning. In Proc. SIGIR'17. New York, NY: ACM, 555-564.
Vakkari P. (1999). Task complexity, problem structure and information actions: integrating studies on information seeking and retrieval. Information Processing & Management 35(6): 819-837.
Vakkari, P. (2003) Task-based information searching. *ARIST* 2003, vol. 37. Information Today: Medford, NJ, 413-464.
Vakkari P. (2010). Exploratory searching as conceptual exploration. In: Proceedings of 4th workshop on Human-computer interaction and information retrieval. New Brunswick, NJ, 24-27.

Vakkari, P. (2016). Searching as Learning. A Systematization based on Literature. *Journal of Information Science* 42(1): 7-18.

Vakkari P. (2000). Relevance and contributing information types of searched documents in task performance. In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '00*).* New York, NY: ACM, 2-9.

Vakkari P. (2001). A theory of the task-based information retrieval process: a summary and generalization of a longitudinal study. Journal of Documentation 57(1): 44-60.

Vakkari, P and Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. Journal of Documentation 56: 540-562.

Vakkari P and Huuskonen S. (2012). Search effort degrades search output but improves task outcome. Journal of the American Society for Information Science & Technology 63(4): 657-670.

Vakkari, P. & Jones, S. & MacFarlane, A. & Sormunen, E. (2004). Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion. *Journal of Documentation* 60(2): 109-127.

Vakkari, P. & Kuokkanen, M. (1997).Theory growth in information science: applications of the theory of science to a theory of information seeking. Journal of Documentation 53(5): 497-519.

Vakkari P, Pennanen M and Serola S. (2003). Changes of search terms and tactics while writing a research proposal. Information Processing & Management 39(3): 445-463.

Vakkari, P., Völske, M., Potthast, M., Hagen, M. & Stein, B. (2019). Modelling the usefulness of search results as measured by information use. *Information Processing & Management* 56(3): 879-894.

Vakkari, P., Völske, M., Potthast, M., Hagen, M. & Stein, B. (2021). Predicting essay quality from search and writing behavior. *JASIST* 72(7): 839-852.

Walhout, J., Oomen, P., Halszka, J. and Brand-Gruwel, S. (2017). Effects of task complexity on online search behavior of adolescents. Journal of the Association for Information Science and Technology, 68(6): 1449-1461.

Wang P. (1997). User's information needs at different stages of a research project: a cognitive view. In: Vakkari P, Savolainen R and Dervin B. (eds.), Information Seeking in Context. London & Los Angeles: Taylor Graham, 307-318.

Wang P. and Soergel D. (1998). A cognitive model of document use during a research project: Study I: Document selection. Journal of the American Society for Information Science & Technology 49(2): 115-133.

Wang, P., & White, M. D. (1999). A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *Journal of the American Society for Information Science,* 50, 98-114

Wildemuth, B. M., de Bliek, R., Friedman, C. P., and File, D. D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. *J. Am. Soc. Inf. Sci. Technol*. 46, 8 (1995), 590-607.

Zhang P. and Soergel D. (2014). Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. Journal of the Association for Information Science and Technology 65(9): 1733-1758.

Zhang P. & Soergel D. (2016) Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science* 42(1): 59-78.

Zhang, X., Liu,J. Cole, M. and Belkin, N. (2015). Predicting users' domain knowledge in information retrieval using multiple regression analysis of search behaviors. Journal of the Association for Information Science and Technology, 66(5): 980-1000.