eCognition and changes of search terms and tactics

during task performance: A longitudinal case study

Proceedings of the RIAO 2000 Conference. Paris: C.I.D. 2000, 894-907

University of Tampere

Department of Information Studies

FIN-33014 University of Tampere, Finland

Pertti.Vakkari@uta.fi

Abstract

The objective of this study is to analyse how changes in users' problem stages during task performance are related to changes in search tactics and term choice. It is analysed how students' growing understanding of the topic is related to their choice of search tactics and terms during accomplishing a research proposal for a master's theses. The participants of the study were 11 students who attended a seminar during which they were to prepare a research proposal. They made a search in LISA data-base in the beginning, middle and end of the seminar. Data for describing their understanding of the work task, search goals and tactics as well as term choices were collected during the search sessions. A pre- and post-search interview was conducted during each session. The students were asked to think aloud during the search session. The transaction logs were captured and the think alouds were recorded. The results show that the students' problem stages during the task performance were connected to their choice of search terms and tactics. The differentiating conceptual representation of the task by the students lead them to use more and increasingly specified search terms, more and varied operators as well as more tactics in the course of their project.

Introduction

Information retrieval (IR) is a part of a broader process of information seeking which aims at finding relevant information for solving a problem or accomplishing a task. (Bates 1989; Belkin 1980; Belkin 1993; Hert 1996; Ingwersen 1996, Marchionini 1995; Vakkari 1999). Thus, actors' articulation of their tasks and problems, and the interaction of that changing understanding with IR-systems is a vital part of information searching. Our knowledge of task-oriented IR-interaction is based on some theoretical outlines (Bates 1989; Belkin 1980, Belkin 1993; Sutcliffe & Ennis 1998; Ingwersen 1996; Vakkari 1999) and empirical studies on the search process (e.g. Ellis 1989; Hert 1996; Kuhlthau 1993; Yang 1997), on search strategies and tactics (e.g. Fidel 1991; Wildemuth & al 1991; Xie 1997) and term choices (e.g. Hsieh-Yee 1993; Wang 1997).

Although the theoretical notions imply that IR should be studied as a process generated by a task, empirical studies typically concentrate on analysing elements such as terms, moves and tactics within a search session. These studies have identified, categorised and described those elements. Their contribution has been crucial in creating basic concepts and categorisations for analysing these features in IR. However, there are very few studies which have analysed IR as a process including shifts of search tactics within a search session (Wildemuth & al 1991; Xie 1997). Studies that connect IR with the task it supports and analyze successive searches are even rare.

The aim of this study is to analyse successive IR searches generated by real life tasks. This study concentrates on analysing how the growth in students' understanding of their research topic during writing a research proposal for a master's thesis is connected to changes in search tactics and terms. It is a longitudinal case study. To our knowledge this is the first attempt to empirically study connections between changes in an individual´s problem stages and the variation in the use of search terms and tactics during a task performance process.

The framework for this study is constructed by using Kuhlthau's (1993) model on the information search process and ideas from cognitive psychology. Kuhlthau's (1993) model is a tool for differentiating the task performance process into separate stages that generate different information needs and information search strategies. Ideas from cognitive psychology are used for describing the mental representations of the tasks by the subjects.

Framework

Taking subjects' prior knowledge of the task as a point of departure for analysing IR is proposed by the advocates of the cognitive view point. Belkin (1980, 1993) and his colleagues (Belkin & Oddy & Brooks 1982; Belkin & Seeger & Wersig 1983) have continuously argued that users' prior knowledge is crucial for understanding IR. Belkin (1980) has proposed that IR ought to be considered from the point of view of the user's anomalous state of knowledge. It is argued by Belkin and his colleagues (Belkin & Oddy & Brooks 1982) that this approach recognises that a fundamental element in the IR situation is the development of an information

need out of an inadequate state of knowledge. Moreover, for IR to be successful, that information need must be represented in terms appropriate for just that task, with the remaining elements of the system represented or constructed on the basis of that representation. These studies have made an important contribution to our understanding of the anomalous states of knowledge. However, they leave open the question of how users' conceptual structure representing information needs is related to actual search activities.

Kuhlthau's model

Kuhlthau (1993) has shown in a series of empirical studies that learning tasks and problem solving by students and library users consists of several stages. Her theory holds that people search for and use information differently depending on the stage of the process.

Kuhlthau (1993) differentiates the task performance process into six phases. At *initiation*, people become aware of the lack of knowledge and understanding. Thoughts centre on understanding the task, and relating the problem to prior knowledge. During *selection*, the task is to identify and select a topic to be investigated. In *exploration*, the task is to investigate information on the general topic in order to extend personal understanding. Thoughts centre on becoming oriented and sufficiently informed about the topic to form a focus. At these stages an inability to precisely express what information is needed makes communication between the user and the system awkward. The information encountered rarely fits smoothly with previously-held constructs.

In *formulation*, a focused perspective on the topic is formed. A focus is comparable to a hypothesis. This is a crucial phase in the task completion because it helps a person to focus on relevant information. At this point, the task is to gather information related to the focused topic. Thoughts centre on defining, extending and supporting the focus. *Collection* is the stage of the process when the interaction between the user and the information system functions most efficiently. The user, with a clear sense of direction, can specify the need for relevant, focused information to systems (Kuhlthau 1993). In the *presentation* stage, the task is to complete the search and use the findings. Actions involve a summary search for rechecking sources (Kuhlthau 1991).

To summarise, in the pre-focus phases the searcher is unable to construct the task and unable to express specifically what kind of information is needed for it. We can assume that the conceptual structure of the searcher is vague, lacking discriminatory power and thus, it is undifferentiated. The subject is able to express search terms only on a general level. Specific terms are not commonly used. Subjects tend to maximise recall, because they are not acquainted with the topic (Vakkari & Hakala 2000). In the post focus phase, searches become more specific and focused. The conceptual structure of the subject is more differentiated and integrated. This implies that search terms are more specific and the searcher is using more terms than in the beginning of the process. Subjects are more acquainted with the topic and they aim at maximising the precision of searches (Vakkari & Hakala 2000).

Prior knowledge

A subject's prior knowledge about the task considerably regulates how much and what kind of information is required and assessed as useful (Patel & Ramoni 1997). Human perception and the learning of new categories is dependent on our knowledge and models about the world. (Hahn & Chater 1997; Heit 1997) In learning new categories, people act as if these categories will be consistent with previous knowledge. People act economically, so that previous knowledge structures are reused when possible. Thus, we learn new categories and acquire information about new tasks based on our current understanding of the phenomenon at hand. Basically, we observe and shape new phenomena in terms of what we already know. (Hahn & Chater 1997; Heit 1997) People select the relevant features and categories of the problem by ignoring others that do not seem to fit with their prior knowledge. Moreover, they chunk the provided information into schemas representing their current conceptual structure of the task (Patel & Ramoni 1997). Thus, prior understanding orients the subjects to categorise the unknown parts of the task in terms familiar to them.

Cognitive structures both in texts and human minds can be understood to consist of concepts and their relationships. They can also be called mental models or schemata (Gavin 1998). If a subject has insufficient knowledge of his task, he does not have the necessary concepts and links for the phenomena he intends to understand. We can say that insufficient knowledge refers to the degree to which a person is able to connect a task to his prior knowledge. (Vakkari 1999) Moreover, if a person has an anomalous state of knowledge, the discriminatory power of his concepts is weak, and the concepts are vague. A person with a clear understanding of the task has a differentiated conceptual structure in which discriminatory power is strong.

Earlier empirical results

Due to the lack of empirical research on how the choice of search terms and tactics change in a task performance process, the following relevant results are presented from studies which analyse the search tactics or change of term choices in general as well as from studies on the connections between domain knowledge and the search process.

Kuhlthau (1993) has shown that people search for and use information differently depending on the stage of their information search process. Her results are presented in detail in the earlier section. The findings by Yang (1997) corroborate Kuhlthau's results in a hypertext environment.

Hsieh-Yee (1993) compared the search tactics of librarians and educational administration students when they searched their own or others' subject domain. She found that subject knowledge becomes a factor only after searchers have had a certain amount of search experience. According to her results, experienced searchers used more of their own terms on a familiar topic, but included more synonyms and combined more search terms when searching on an unfamiliar topic.

Wildemuth and her colleagues (Wildemuth &al 1995) studied how the subject knowledge of medical students was related to their searching proficiency. They found that there is no strong relationship between a searcher's domain knowledge and their search results and term selection.

Wang (1997) studied how users' information needs change during the stages of a research process by analysing their document selection from retrieved documents. She analysed the vocabulary of users in request, document selection and in the post project stages. She demonstrated that the individuals introduced narrower and related terms as the research proceeded (Wang 1997). The introduction of narrower terms refers to the specification of the research problem and the construction of a focus in the research process. Wang (1997) also found that the actual vocabulary in each later search stage was substantially larger in size than in the previous one, broader and deeper in hierarchy, and wider in breadth.

Wildemuth and her colleagues (Wildemuth &al 1991) studied medical students' search tactics in a factual database. They found that the simplest tactics were the most common, with single-move tactics accounting for over a half of those used. Students used almost always AND- and very seldom OR-operators.

To summarise, Kuhlthau (1993) and Yang (1997) demonstrate that subjects' search strategies change as they proceed in their task. However, these studies do not include a detailed analysis of search tactics and terms. Wang's (1997) study supports the idea that subjects use more and specific terms as they proceed in the research process. Studies on the relationship between domain knowledge and search tactics and term choice are inconclusive.

Research design and research problems

The aim of this study is to explore the choice of search terms and tactics generated by natural tasks. We study the actual search behavior of users during their task performance: how their use of search terms and tactics change in this process. This implies that our research design is not a controlled experiment, but a case study in a natural setting. The results reflect the features of the searches carried out by the users, whether advanced or simple.

-

The research design does not include variables describing IR techniques (i.e. document indexing, matching methods, relevance feedback, provision of vocabularies). The choice is conscious. We do not yet know enough about how subjects search for information in databases while performing their tasks. We have first to understand the central elements of their search activities generated by their tasks before it is reasonable to explore which IR techniques might be efficient tools to support their searching. We intend to include IR techniques in the research design at the next stage of our project.

The aim of this study is to analyse how students' problem stages are connected to their use of search tactics and terms in preparing a research proposal for a master's thesis. The participants of the study were 11 students from the Department of the Information Studies at the University of Tampere that attended a seminar on preparing a research proposal. The seminar lasted for four months during the 1999 spring term. At the beginning of the seminar they selected a topic and were expected to come up with a proposal. Writing a research proposal can be classified as a complex task. The students had attended classes on IR at the Department. Thus, they had some search expertise. Its variation within the group was considerably small.

Data

Data for describing the students' understanding of the task, their problem stages and search tactics and terms was collected in several ways. They were asked to make an IR search three times during the seminar: at the beginning, and in the middle of the seminar, as well as when they were finishing or had completed the proposal. The aim was to collect data in the pre-focus, focus formation and post-focus stages of the students. A pre- and post-search interview was conducted in each case. The pre-interview consisted of Kuhlthau's (1993) process survey questionnaire and a semi-structured interview. Both measured feelings, thoughts and actions in the respective problem stages. The latter concentrated on measuring participants' state of knowledge and experience of the topic and their goals and intended actions. They were asked what kind of information they were looking for and what they expected to do with the search results. After the interview they made a search in the Dialog's LISA database. They thought aloud during the search session, which was recorded. The transaction logs were also recorded.

In the post-session, interview data was collected on their relevance assessments of the references found. The scale was “relevant, partially relevant and not relevant“. They were also asked to assess if the session or references helped them to structure their problem. The results of the relevance assessments have been reported in Vakkari & Hakala (2000).

Concepts and operationalizations

The students' problem stages in the process were identified by using Kuhlthau's (1993) model. They were divided into pre-focus, focus formation and post-focus stages. The first stage includes initiation, selection and exploration phases, the second is the formulation phase, and the last are the collection and presentation phases. The stages were operationalised as answers to questions in Kuhlthau's (1993) process survey questionnaire.

A query is understood to be a representation of a user's information need, which consists of search terms and of possible operators connecting them. A facet is an aspect of a query, which may contain one or more search terms. The terms within a facet are combined by OR-operators (Kekäläinen 1999).

Definitions*)	Operationalizations“)
Strategies to begin a session	
Select: To break down complex search queries into subproblems and work one problem at a time	At most two thirds of all search terms entered in the beginning of the search
Exhaust: To include most or all elements of the query in the initial search formulation	More than two thirds of all search terms entered in the beginning of the search
Search formulation tactics	
Intersect: Intersect a set with a set representing another query component	Terms added to the query using an AND operator
Vary: To alter or substitute one's search terms in any of the several ways	At least one new term was substituted for one of the terms in the preceding move so that the number of terms remained the same
Parallel: To make the search formulation broad by introducing synonyms or conceptually parallel terms	At least one synonym or conceptually parallel term was added
Reduce: To subtract one or more of the query elements from an already-prepared search formulation	Set of terms was repeated, minus at least one term
Negate: To eliminate unwanted elements by using AND NOT operator	At least one AND NOT operation was used
Union: To replace an AND operator with an OR operator	Search terms were identical to the preceding set except for a change of operators from AND to OR

Other tactics

Focus: To look at a query more narrowly	To move from a narrow to broader conceptualisation of the query
Limit/de: Use free-text terms as descriptor	Use free-text terms as descriptor
Limit/la: To limit the search by language	To limit the search by language
Limit/py: To limit the search by publication year	To limit the search by publication year

*) Definitions of tactics are from Bates (1990) except Negate is from Fidel (1991) and Union from Wildemuth &al (1991). Definitions for Other tactics are given by author of this article except Focus which is from Bates (1990). ") Operationalisations are from Wildemuth &al (1991) except Monitoring and Other tactics which are defined by author of this article.

Table 1. Definitions and operationalisations of search tactics

A move in a search is the basic unit of analysis. A move is understood as an identifiable thought or action that is a part of information searching (Bates 1990) for improving search results (Fidel 1991). In our study a move was a change made in a query in order to attain the goals of the search. A move was operationalised as a step or steps that were necessary to improve the search for making the next move. Tactics consist of a set of moves. Tactics represent the first level at which strategic considerations are primary (Bates 1990). The categories of tactics used in this study are created by combining categorisations by Bates (1990), Fidel (1991) and Wildemuth &al (1991). They are described in Table 1.

The students used to some extent truncation and proximity operators in the searches. Due to the relatively limited utilization of these features the analysis was not focused on their use. However, the truncated terms as well as terms combined by a proximity operator were taken into account in the analysis of search terms. A typical way for students to use proximity operators was to form meaningful phrases like "information(w)need". These were calculated as one term.

In a field study like this it is impossible to estimate the recall of the searches, because we do not know how many relevant items the data-base contains. The precision is calculated based on the relevance assessments of the students. If the final set was large, the students were asked to assess the first twenty references. The precision is the share of the partially relevant and relevant references of all references found per student in a search session. Due to the small number of the participants in this study it is not always possible to estimate the precision figures in a meaningful way. The results will be discussed, where possible.

Research questions and hypotheses

The major research question is: How did the search tactics and search terms change during the preparation of the research proposal by the students? It can be divided into following sub-problems: In the three successive search sessions, 1) how many search terms were used by the students; 2) what kind of new terms were introduced; 3) what kind of operator types were used; 4) what tactics were used; and 5) how were the tactics patterned.

Based on the framework of the study, we can infer the following hypotheses: The less prior knowledge, the more undifferentiated the conceptual structure and the lower the discriminatory power of the concepts, and the less relationships between the concepts. As a corollary, we can hypothesize: the less prior knowledge, 1) the more difficult it is to express search facets and terms; 2) the fewer facets and terms are used in a search; 3) the more general (broad) the facets and terms are; 4) the less synonyms are used; and 5) the fewer types of operators are used.

Results

Stages in the task performance

In general, all the participants proceeded in their tasks according to Kuhlthau's (1993) model at a varying pace. In the first round, the students were moving from the topic selection to exploring it. In the middle of their task they were typically exploring the topic and trying to formulate a research problem. In the end of the project they were logically in the presentation stage, but only half had been able to construct a focus, and the other half was struggling with it. A detailed analysis of students' problem states can be found in Vakkari & Hakala (2000).

Search terms and facets

The number of search terms and facets increased when students proceeded in their project (Table 2). _In the first session the number of terms varied from 2 to 5, in the second session from 2 to 9 and in the third session from 3 to 11. The size of the vocabulary grew among all students with one exception. The students started the searches with 3 terms and in the final search they used 5,5 terms. The increase in terms and facets was also steady between the sessions. The findings suggest that the students' conceptual structure representing their topic becomes more differentiated in the process which is reflected in the growth of the number of search terms and

facets.

	I session (n=11)	II session (n=11)	III session (n=10)
Terms	3,0	4,2	5,5
Facets	2,4	3,0	3,7

Table 2. Number of facets and search terms per student in successive search sessions.

The number of terms used did not differentiate the precision (the combined share of partially relevant and relevant references) of the searches in the first and second search sessions. However, in the last search session the precision was 42 % for those who used more than average number of search terms compared with 30 % precision for those who used fewer search terms than average.

The new terms students introduced in the searches reflect their changing mental model. The new terms were classified into four categories which were adopted from Wang (1997). A synonym (ST) is a term that is interchangeable with another term. A broader term (BT) means a term which is broader in hierarchy. A narrower term (NT) refers to a term narrower in hierarchy. A related term (RT) is a term which is associated to another term.

In analysing term relations it was difficult to differentiate between NTs and RTs. In several cases, an RT introduced a new aspect in the search. However, in many cases, RTs were conceptually very close to their predecessors. In some cases they specified an aspect of the original term or were in some other way partly overlapping conceptually. For example, a student interested in computer assisted learning introduced this phrase in the first session, and in the second one intersected it by the terms "programme" and "game". These terms were classified into RTs. Thus, the role of the RTs in the vocabulary of the students, in some cases, approached that of NTs.

	II session (n=11)				III session (n=10)			
Term	ST	BT	NT	RT	ST	BT	NT	RT
Sum	7	2	8	11	7	2	10	20
Mean	0,6	0,2	0,7	1,0	0,7	0,2	1,0	2,0

Legend: ST = synonym BT= a broader term NT= a narrower term RT= a related term

Table 3. Total number and mean value per student of types of new terms in the second and third session.

In the second round, students introduced almost as many RTs, NTs and STs (Table 3). Seven out of the eleven students introduced either STs or NTs. The remaining four brought RTs. Only two new BTs were used in the queries. In the final round, students introduced into their search vocabulary the most RTs. They account for about half of the new terms. Almost all of the students used them. The share of NTs was about one quarter and STs about a fifth of all fresh terms. Also in this stage only a few BTs were introduced.

The results show that from the topic selection to focus formulation stage the vocabulary growth consisted quite evenly of RTs, NTs and STs. When the students passed the focus formulation, they utilised the most RTs, but NTs and STs had also an important role in their changed vocabulary. New BTs were very rare in the search vocabulary and there was a tendency to drop them when the focus was crystallised.

The results suggest that the differentiating conceptual structure of the students is reflected in the change patterns of the used search terms. The introduction of STs and NTs and discarding of BTs were a reflection of their narrowing and differentiating focus and of the growing mastery of terminology. The growing number of new RTs in their evolving vocabulary was a further indication of this. RTs brought either new or specifying aspects into their queries. All refer to the fact that their queries were developed terminologically and became more specific during the process.

Operators

In calculating the number of introduced operators each was counted only ones. Table 4 shows that as the students were proceeding in their project, they began to use the operators in a more multi-faceted way. In the beginning, they used the AND operator in 9 cases of the 11. Typically, they used two ANDs in the query. In the final round they mostly combined the terms by AND operators but their utilisation of the OR operator was increased to one fourth of all the operators. The increased and varied use of operators is a reflection of the growth and structure of the search vocabulary. In the course of their project students learn synonyms for their terms. OR operators are used to combine the synonyms within a facet (Harter 1986).

	I session (n=11)			II session (n=11)			III session (n=10)		
Operator types	AND	OR	NOT	AND	OR	NOT	AND	OR	NOT
Sum	26	3	0	35	9	2	36	14	2
Mean	2,4	0,2	0,0	3,2	0,8	0,2	3,6	1,4	0,2
Percent	90	10	0	76	20	4	67	26	7

Table 4. Total number and mean value of operator types per student in successive search sessions.

The precision of the searches related to the number of operator types could not be calculated in the first session, because only two out of the 11 students used more than one operator type. In the last two sessions those who utilized OR operators in addition to AND operators found more relevant or partially relevant references than those who used only ANDs. The precision of the latter group was 24 % both in the middle and at the end of the process whereas in the former group it grew from 33 % to 41 %. The growth was based on the increase in the share of partially relevant items. It seems that the utilization of the OR operator together with AND operator leads to search results with a higher precision than mere intersecting. We will return to this finding when analyzing search tactics.

Search tactics

The total amount of the used tactics increased when the students advanced in their project (Table 5). In the beginning session they applied approximately four tactics, in the interim session they utilised over five and in the final session almost eight tactics.

Tactics to begin a session

The students began their search session either by Select or Exhaust tactics. In Select, they started the search by introducing less than two thirds of the terms they used in the whole search. In Exhaust, the students entered practically all the terms they used in the initial search formulation. In the first round, 5 of the 11 students began the session by Exhaust. They included all the terms they used in the initial formulation. One of them stopped after it and the rest narrowed the query by limiting the language and printing year. Thus, students always continued Exhaust by two operational moves. An operational move uses the system’s features in order to modify a query without changing its conceptual meaning (Fidel 1991).

The students who began with Exhaust typically included two terms and one AND operator in their searches during the first round. The average for all the subjects was three terms combined with two AND operators (cf. Tables 2 and 4) . The same pattern was observed in the second round. Students who chose Exhaust in the first round were on average more at the beginning stages of Kuhlthau's (1993) model. Those who adapted Select tactics were further in the process. Thus, it seems that students with a more vague understanding of their topic tend to empty all the elements of their prior knowledge at once in the initial query and use operational moves for furthering the search.

In general, the use of Exhaust decreased and that of Select increased when the students moved towards the end of the project.

Tactics	I session	II session	III session
Select	6	9	10

Exhaust	5	2	-
Intersect	5	13	14
Vary	5	6	7
Parallel	1	3	9
Reduce	1	2	2
Browse	2	6	14
Monitor	1	4	5
Limit/la	7	4	3
Limit/py	5	3	2
Other	3	7	12
Total	41	58	77
Mean	3,7	5,3	7,7

Table 5. Number of search tactics in successive search sessions

In the first search session the use of Exhaust resulted in a precision of 34 % and the use of Select in a precision of 43 %. Those who were able to represent their topic with more terms were able to generate more differentiating searches, producing a higher number of relevant items than those with a more vague understanding of the problem. It was interesting to note that at this stage of the process the students endeavored to maximize the recall of the searches. A typical expression by the students in judging the relevance of the found items was the following: "At this stage, when I do not know much about the topic, I have to consider this reference as relevant".

Search formulation tactics

Intersect was the most common tactics in each session. Its use increased heavily after the first round. Vary was also a very common tactic in all the sessions. It was used when the precision of a set was low and its size so small that it was not reasonable to intersect. In a query with two terms, students typically kept one of the terms and substituted the other with varying terms one by one.

The number of Parallel tactics increased heavily when the students had constructed a focus for their study. In the final session, it was the second most frequently utilised tool in query formulation. In these tactics, the students increased the size of the set by introducing synonyms and parallel search terms combined with the OR operator. The use of this means goes hand in hand with their increasing understanding of the different aspects of the topic and of their various terminological expressions.

Other tactics

Students browsed the search results increasingly in the successive searches for assessing their relevance. They also monitored the moves more frequently as the project progressed. These features are consequences of the more comprehensive searches at the end of the project. They have to follow the references and the moves more often to keep themselves on track during the session.

The decrease in the use of Limit commands reflects the fact that the students were able to represent their information need in more specific terms which led to better search results. They did not need to use operational moves for reducing the size of the set as often.

Patterns of search tactics

In the analysis, we will focus on tactics to begin a session and search formulation tactics. Monitoring and other tactics were excluded for obtaining a clearer picture of the transformations. We analyse only the search process of those students who started with Select. The patterning of tactics by students who began with Exhaust has been analysed in the previous section.

As Table 6 shows, independent of the stage of the process, students used similar chains of tactics in similar retrieval conditions. In all of the rounds, if the retrieved set was large enough, they narrowed the query by Intersect. If Select retrieved plenty of references, they intersected the set as many times as necessary to receive a small enough set. A typical combination was Select-Intersect continuously throughout the seminar.

I Session

01	SELECT (229) – INTERSECT (0) * - VARY (157)
03	SELECT (372) – INTERSECT (181) – INTERSECT (2) * - REDUCE (118)
04	SELECT (307) – INTERSECT (16)
05	SELECT (157) – INTERSECT (77)
09	SELECT (16) * - VARY/PARALLEL (52) – VARY (33)
10	SELECT (1) * - VARY (12) – VARY (116)

II Session

02	SELECT (212) – INTERSECT (37)
03	SELECT (381) – INTERSECT (16) – INTERSECT (4)
05	SELECT (100) – INTERSECT (0) * - VARY (11) – VARY (4)
06	SELECT (17) – INTERSECT/PARALLEL (0)
07	SELECT (354) – VARY (1421) – INTERSECT (2) * - REDUCE (963)
08	SELECT (82) – INTERSECT (4) * - VARY (2) – VARY (17)
09	SELECT (8743) – INTERSECT/PARALLEL (1679) – NEGATE (26) – INTERSECT (1)
10	SELECT (1) * - VARY (0)

III Session

01	SELECT (173) – INTERSECT/PARALLEL (59) – INTERSECT (7) – INTERSECT (0) – REDUCE (23)
02	SELECT (232) – INTERSECT/PARALLEL (15)
03	SELECT (479) – INTERSECT/PARALLEL (80)
04	SELECT (380) – INTERSECT/PARALLEL (7)
05	SELECT (118) – INTERSECT (1) * - VARY (4) – VARY (14) – VARY (15) - VARY (15)
06	SELECT (24) – NEGATE (17) – FOCUS (18)
07	SELECT (1452) – INTERSECT (31) – INTERSECT (0) * - VARY (1)
08	SELECT (933) – INTERSECT/PARALLEL/NEGATE (95) – INTERSECT (0) * - REDUCE (1)
10	SELECT (10) * - VARY (6)
11	SELECT (10) * - VARY (20)

Legend: * = A tactics resulting a small set; Size of the sets in the brackets

Table 6. Transitions in search tactics in the first, second and third session

Throughout the process, if the opening set in Select was small or if Intersect in the middle of the search produced a minor set, the typical way to enlarge the size of the set was to use Vary tactics. From a query of two or three terms combined with an AND operator, students retained one or two terms constantly and varied one term by substituting it with a new term. The new term was in many cases a broader one to increase the recall. Students tried to map a certain conceptual field systematically by dividing it into subsets. These tactics resemble Parallel tactics that are accomplished made step-by-step.

As an alternative to Vary, Reduce was seldom used to enlarge the size of the set in all the phases. However, Reduce was used instead of Vary, if the search was elaborated at least with two tactical moves after Select. The reformulations before Reduce included more than an average number of terms and moves. This made it possible to exclude some elements for enlarging the size of the set. Vary was typically used either after Select or Select-Intersect sequence.

When the students advanced in their task, they began to use synonyms and parallel terms combined by OR operators for expressing the dimensions of facets. It was reflected in the increasing use of Parallel tactics from the beginning to the final search. Parallel was always associated with Intersect and preceded by Select. The more frequent use of Parallel tactics was made possible by students' differentiating vocabulary of the topic.

Those search strategies that were based on Parallel tactics produced higher precision rates in the second and third search sessions than those that consisted mostly of Intersections. In the second session students who used Parallel judged 33 % of the references to be relevant or partially relevant where as Intersecting resulted in a precision estimation of 23 %. In the third round students utilizing Parallel tactics found 41 % of the items and students using mostly a combination of Intersect and Vary tactics 33 % of the items to be relevant or partially relevant.

-

It seems that the students who were most knowledgeable in their research topic had the richest vocabulary at each stage of the process. This resulted in an increasing use of OR operators and consequently Parallel tactics in their searches. The results demonstrate that both the use of OR operators and Parallel tactics led to search results with higher precision in terms of relevance assessments by the students. It is evident that the increase in precision of search terms used by the students produced more relevant items in a search. When this was combined with the growing use of Parallel tactics, which aims at increasing recall by adding parallel terms in a query, it is no wonder that it resulted in a higher proportion of relevant items than other tactics.

To conclude, the patterning of certain tactics did not usually depend on the problem stage of the students, but on the size of the retrieved set. If the set was large, students use after Select Intersect tactics as many times as the number of retrieved references allowed. Select was followed by Vary if the beginning set was small. Moreover Vary was typically used after Intersect if the set was minimal. Reduce was used to reduce the size of the set instead of Vary if the query reformulation included over three tactical moves. Only the choice of Parallel increased with the pace of the students' growing understanding. It was typically connected with Intersect and preceded by Select.

Discussion and conclusions

The findings of this study support our main hypothesis, namely that the students' problem stage during task performance is systematically related to their choice of search terms and tactics. The growing and focusing understanding of the task by the students led them to use more and specified search terms, more and varied operators as well as more tactics in the course of their project. The results deepen our understanding of the nature of successive searches generated by natural tasks. The results are based on a longitudinal case study with a small number of participants. The study is exploratory, and the results tentative.

Terms

When students progress through their proposal writing, their mental construct of the research topic becomes more focused and discriminating. When selecting their topic, they represented it with few search terms. As their mental models differentiate, the vocabulary in the searches grows in each round with synonyms, narrower and related terms. Broader terms were dropped especially in the last round. Students were able to express the topic with a larger and more specific vocabulary in the successive searches.

The results by Wang (1997) on researchers' vocabulary changes generated by information needs at request, document selection and post project stages are similar to our findings. She showed that their actual vocabulary was larger in size at later stages. The individuals introduced narrower and related terms as the research process proceeded.

Tactics

Only some tactics and patterns of tactics depend on the stage of the task performance and subjects' state of knowledge. Others are related to the size of the retrieved set, which is independent of the phase of the process.

The students began the search either by introducing all the search terms or entering only a fraction of the terms in the initial query. The former tactics was called Exhaust and the latter Select. As the task performance proceeded, the students' use of Exhaust decreased and Select increased. Those who used Exhaust were not as far along in their process, and their conceptual construct was less developed than those who started with Select. The previous group were able to represent only a couple, in the first round only two, search terms. Exhaust tactics was typically followed by operational moves.

In all stages of the project Select was followed by Intersect or Vary tactics depending on the size of the retrieved set. If the set was large, Intersect was a typical move. If the set was small, Vary was the preferred option. In Vary, the number of terms was similar than in the preceding move, but one term was substituted with a new one. It was typically used when the students tried to expand a small set. An alternative for Vary was Reduce in enlarging the size of a set. It consists of reducing the number of terms in a query. It was used more seldom. It was selected when the search reformulation was more complex than on average. It was introduced after three tactical moves. Vary was used in earlier phases of the search reformulation.

The students began to use more synonyms and parallel terms when their knowledge of the topic increased. This was reflected in the increasing use of Parallel tactics search by search. It was the second most common tactic at the end of the process.

Those features in the searches yielding high precision were the use of more precise and parallel terms and their combination by OR operators in facets producing Parallel tactics. The mere use of AND operators (Intersect) with few facets containing only one term led to search results with a lower precision. The use of precise terms in a query results naturally in a precise search result. The findings also demonstrate that an active utilization of synonyms and parallel terms for increasing recall seemed to lead to a higher precision compared to other tactics like Intersect. The explanation for this is that queries consisting of facets with several and perhaps specific terms have more discriminatory power than queries consisting of an intersection of facets expressed with a single and perhaps vague term each, given that the number of facets is the same. As a result the former queries retrieve sets containing proportionately more relevant references. The results also demonstrate that students with a more structured domain knowledge tend to formulate queries with more facets (with greater complexity) and with more specific terms (with greater specificity and coverage) resulting in a higher share of relevant items.

This study has shown a systematic and many-sided connection between the students' growing domain knowledge and their IR interaction in successive search sessions. The growth in domain knowledge seems to predict users' IR interaction in successive search sessions. This is inconsistent with the findings of Wildemuth &al (1995). Their results suggest that there is no strong relationship between a searcher´s personal knowledge in a domain and his/her ability to search effectively in that domain. The difference might be explained by the differences in research design of these two studies.

Supporting searching

We found that the degree of students' knowledge of the topic predicts their ability to express search terms and formulate tactics. The less they know, the fewer, broader and more vague term they use and the shorter queries and simpler tactics they formulate. The evident conclusion is that people with scarce domain knowledge need support for expanding and differentiating their conceptual model of the topic. This would help them to develop ideas on how to structure the topic and how to express their vague information needs more in detail. Equipped with synonyms and the narrower terms provided by the system they could reformulate their query using terms with stronger differentiation power. This would result in finding more relevant information items.

Our findings support the previous claim. We found that queries containing facets with terms combined by OR operators produced a higher share of relevant references than queries containing facets represented by one term and combined by AND operators. This result suggests that, in addition to support in finding appropriate vocabulary, the users also need help in formulating the query, especially in using OR operators. Interfaces that suggest users should utilize Parallel tactics would support them in retrieving a higher proportion of relevant information items.

Our findings suggest, however, that provision of separate terms is not enough. In people´s conceptual frameworks, the meaning of a term is regulated by its connections to other related terms (Hahn & Chater 1997; Vakkari 1999). In presenting the terms to the users, they should be linked to relevant terms on different levels of hierarchy as well as to synonyms. In an ideal case, the central features of the conceptual structure provided by the system should resemble the mental model of the users (Bates 1990; Marchionini 1995; Vakkari 1999). However, in the topic selection and exploration stages when the user´s conceptual model is undifferentiated, any kind of conceptual map would help them to find new dimensions and alternative expressions for their information need.

Scholars have argued for tools like a searching thesaurus for counseling a user in search term selection (Bates 1990; Fidel & Efthimiadis 1995) or thematically organized terminological feedback (Anick & Tiperneni 1999). Our study suggests that for novices in a domain, structured terminological support would not only to improve search results by the use of more narrow terms, but it might also to support the user with ideas on how to differentiate the topic and interrelate its parts.

To conclude, our results have empirically specified some situations in IR during task performance where the use of varied query expansion tools would support the user to achieve a better search result.

References

Anick, P. & Tiperneni, S. (1999), The Paraphrase Search Assistant: Terminological Feedback for Iterative Information Seeking. In: *Proceedings of the SIGIR´99*. ACM, New York, 153-159.

Bates, M. (1989), The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13(5):407-424.

Bates, M. (1990), Where Should the Person Stop and Information Interface Start? *Information Processing & Management* 26(5):575-591.

Belkin, N. (1980), Anomalous States of Knowledge as Basis for IR. *Canadian Journal of Information Science* 5: 133-143.

Belkin, N. (1993), Interaction with Texts: IR as Information Seeking Behavior. *In: IR ´93.* Konstanz:

Universitetsverlag Konstanz, 55-66.

Belkin & Oddy & Brooks (1982), ASK for IR I & II. *Journal of Documentation* 38 (2&3): 61-71 & 145-164.

Belkin, N. & Seeger, R. & Wersig, G. (1983), Distributed Expert Problem Treatment as a Model for Information System Analysis and Design. *Journal of Information Science* 5(5): 153-167.

Brooks, H. & Daniels, P. & Belkin, N. (1985), Problem Descriptions and User Models. In: *Informatics* 8. London, Aslib, 191-214.

Ellis, D. (1989), A Behavioral Approach to IR System Design. *Journal of Documentation* 45:171-212.

Fidel, R. (1991), Searchers' Selection of Search Keys: III Searching Styles. *JASIS* 42(7):515-527.

Fidel, R. & Efthimiadis, E. (1995), Terminological Knowledge Structure for Intermediary Expert System. *Information Processing & Management* 31(1):15-27.

Gavin, H. (1998), *The Essence of Cognitive Psychology*. London: Prentice Hall Europe.

Hahn, U. & Chater, N. (1997), Concepts and Similarity. In: Lamberts, K. & Shanks, D. (eds.), *Knowledge, Concepts and Categories.* Hove: Psychology Press, 43-92.

Harter, S. (1986), *Online Information Retrieval*. Orlando: Academic Press.

Harter, S. & Hert, C. (1998) Evaluation of IR Systems: Approaches, Issues, and Methods. In: *Annual Review of Information Science and Technology* 32 (M. Williams, Ed.) Information Today, Medford, NJ. 3-94.

Heit, E. (1997), Knowledge and Concept Learning. In: K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories*. Psychology Press, Hove, 7-41.

Hert, C. (1996), User Goals on an Online Public Access Catalog. *JASIS* 47(7):504-518.

Hsieh-Yee, I. (1993), Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experiences Searchers. *JASIS* 44(3): 161-174.

Ingwersen, P. (1996), Cognitive Perspectives of IR Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation* 51(1): 3-50.

Kekäläinen, J. (1999), *The Effects of Query Complexity, Expansion and Structure on Retrieval Performance in Probabilistic Text Retrieval*. Doctoral Diss. Tampere: Tampere University Press.

Kuhlthau, C. (1991), Inside the Search Process: Information Seeking from the Users' Perspective. *JASIS* 42 (5): 361-371.

Kuhlthau, C. (1993), *Seeking Meaning*. Norwood, N.J., Ablex.

Marchionini, G. (1995), *Information Seeking in Electronic Environments*. Cambridge U.P.

Patel, V. & Ramoni, M. (1997), Cognitive Models of Directional Inference in Expert Medical Reasoning. In: Feltovich, P. & Ford, K. & Hoffman, R. (Eds.), *Expertise in Context: Human and Machine*. Menlo Park (Calif.): AAAI Press, 67-99.

Spink, A. & Wilson, T. & Ellis, D. & Ford, N. (1998), Modelling User's Successive Searches in Digital Environments. *D-Lib Magazine*, March 1998. hdl:cnri.dlib/april98-spink.

Sutcliffe, A. & Ennis, M. (1998), Towards a Cognitive Theory of IR. *Interacting with Computers* 10: 321-351.

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. & Hakala, N. (2000), Changes in Relevance Criteria and Problem Stages in Task Performance. *Journal of Documentation*, Sebtember 2000 (To be published).

Wang, P. (1997), User's Information Needs at Different Stages of a Research Project: A Cognitive View. In: Vakkari, P. & Savolainen, R. & Dervin, B. (Eds.), *Information Seeking in Context*. Graham Taylor, London & Los Angeles, 307-318.

Wildemuth, B. & Jacob, E. & Fullington, A. & de Blieck, R. & Friedman, C. (1991), A Detailed Analysis of End-User Search Behaviors. *Proc. of the 54 ASIS Annual Meeting*, 302-312.

Wildemuth, B. & de Blieck, R. & Friedman, C. & File, D. (1995), Medical Students' Personal Knowledge, Searching Proficiency, and Database Use in Problem Solving. *JASIS* 46(8): 590-607.

Xie, H. (1997), Planned and Situated Aspects in Interactive IR: Patterns of User Interactive Intentions and Information Seeking Strategies. In: *ASIS '97. Proceedings of the Annual Meeting*, Vol. 34. Medford, NJ: Information Today, 101-110.

Yang, S. (1997), Information Seeking as Problem-Solving Using a Qualitative Approach to Uncover the Novice Learners' Information-Seeking Process in a Perseus Hypertext System. *Library & Information Science Research* 19(1): 71-92.