



COGNITIVE STRATEGIES IN WEB SEARCHING

Raquel Navarro-Prieto

Mike Scaife

Yvonne Rogers

School of Cognitive and Computer Sciences, University of Sussex, Brighton, BN1 9QH, UK.

+email: raqueln@cogs.susx.ac.uk

Abstract

Usability tests have shown that users often get lost very easily on the Internet when looking for information. However, we still know very little about why this is so and how it can be avoided. The goal of our research is to develop an empirically-based model of web searching, to help explain how people search for information on the Web and to develop guidelines for supporting Web searching. Towards this goal we have developed a framework which characterises the users' characteristics, the task and the information presented, and the interaction between them. We have also conducted a study addressing some of the research questions emerging from our framework. The analysis of our data from this study focused on the cognitive strategies followed by the users, their level of experience and the type of searching task. To analyse the dependencies between these factors we applied the External Cognition framework (Scaife and Rogers, 1996). Using this framework we also analysed how external representations presented to users could explain some of the main problems that they experienced during the searching task.

INTRODUCTION

One of the main claims of the Web community is that the Web allows you to move around the EBworld' freely, giving you access to an endless amount of information, that can be accessed using hypertext navigation. In contrast, the emerging literature about Web usability has highlighted that this is often not the case. Usability studies have shown that users often get lost very easily on the Internet; even in a particular site, they sometimes assume that the information that they want is in the wrong sub-site (Nielsen, 1997). Nielsen (1999) argues that a dilemma of the Web is the difficulty in finding what you need among the abundant sources of information. Why this is so and how can these navigation and search problems be avoided?

Recent research has moved towards developing search models aimed at helping Web designers provide a more consistent framework for structuring information on the Web (Shneiderman, et. al., 1997,

Shneiderman, 1997). However, current models are limited in that they do not account for the interaction between the users searching and the way the Web is structured. Interactivity has been identified as one of the distinctive characteristics of the Web (Buckingham, 1996, Nielsen 1999). Other recent attempts of understanding the process of Web searching, like Pejtersen and Fidel (1998) and Nielsen (1997), have described several cognitive strategies developed by Web users. Under which circumstances, or why, several users develop different strategies? Previous research had not addressed these questions. This suggests that in order to understand the complex task of web searching we need expand the concept of interactivity as the interaction between the user and the system. We claim that it is necessary to consider the interaction between the users, the task and the information presented by the Web. Towards this goal, we have begun developing a theoretical framework, called the Interactivity Framework, that attempt to describe these three elements and the interactions between them.

1. - THE INTERACTIVITY FRAMEWORK

Our objective in developing the interactivity framework is to specify the units of analysis that needs to be considered to study the complex task of information searching within the Web context. Our review of the emerging literature on Web searching and about interactivity in hypermedia systems suggests three factors. These are: the users' experience and cognitive strategies, the type of searching task, and how the information is presented and interacted with the users. We would like to emphasise that the aim of this model is to help us to investigate the interdependencies among these three aspects, highlighting the interaction among them (Figure 1), and not the exhaustive description of each of them.

At the user level we should consider all the variables concerning the users like web experience, cognitive processes, cognitive style and their knowledge. For instance, we know that the needs of web users depend upon their experience and upon how frequently the use the Web (Kellogg and Richards, 1995). Also, various usability studies have highlighted the importance of users' cognitive strategies. Nielsen (1997) has shown in this studies that more than half of the users are search-dominant (i.e. go directly to a search button) , about a fifth are link-dominant (i.e. follow the links around one page), and the rest exhibit mixed strategies. In a recent study, Pejtersen and Fidel (1998) identified six different cognitive strategies used by secondary school children when they were looking for information for their class homework. The most popular were the "browsing strategy" (follow leads by association without much planning ahead) and the "empirical strategy" (use of rules and tactics that were successful in the past). Based on these findings, we are interested in exploring further the kinds of strategies different users adopted. Furthermore, we are interested in how the users plan their searching tasks and how they integrate the information that they receive during this to interpret the situation and change their behaviour.

The nature of searching task is also important to consider in relation to the user's strategies. For instance, Shneiderman (1997) varied the searching task from specific fact-finding to more unstructured open-ended browsing of known databases and exploration of availability of information on a topic. He claimed that identify users' tasks should guides designers in shaping a website. Several guides designed to teach student to look for information in the web also recognise the importance of differentiating between the situation when looking for general information as opposed to looking for specific details (Braham, 1997).

The way information is structured on the Web is also important in relation to the kind of task and the user's strategies. Most research on this has focused on the technical aspects of the interacting with the Web. Technical advances include improved reliability, speed and new tools and techniques for multidimensional and hypermedia presentation. Very little systematic research has been conducted to study how these technical improvements influence the user's interaction with multimedia systems in general (Alty, 1991; Marmollin, 1991), or the Web. However, in order to understand the cognitive processing involved in the

searching task it is critical to study the interaction between the information presented to the users and their internal representations. Previous works on graphical representation processing has emphasised the importance of studying the interaction between the internal/external structures and the cognitive benefits of different graphical representations (Scaife and Rogers, 1996). Since graphical representation are a special case of external representation our approach will be to apply the External Cognition framework to help us understand the interaction in which we are interested. External Cognition refers to the cognitive interplay between internal and external representations (see Scaife and Rogers, 1996). By this we mean the process by which people integrate representations. For example, reading and abstracting knowledge from a web page requires making connections between different elements of the display in a temporal sequence, using both internal and external representations in concert. The framework allows us to identify the properties of external representations in terms of their 'computational offloading'. This refers to the extent to which different external representations reduce or increase the amount of cognitive effort required to understand or reason about what is being represented. High computational offloading is where much of the effort is offloaded onto the representation, requiring minimal effort on behalf of the user for a given task. In contrast, low computational offloading is where much cognitive effort is required by the user to perform their task. In our analysis we have identified three main forms of computational offloading (Scaife and Rogers, 1996). These are:

* *re-representation* - This refers to how different external representations, that have the same abstract structure, make problem-solving easier or more difficult. It also refers to how different strategies and representations, varying in their efficiency for solving a problem, are selected and used by individuals.

* *graphical constraining* - This refers to the way graphical elements in a graphical representation are able to constrain the kinds of inferences that can be made about the underlying represented concept.

* *temporal and spatial constraining* - This refers to the way different representations can make relevant aspects of processes and events more salient when distributed over time and space.

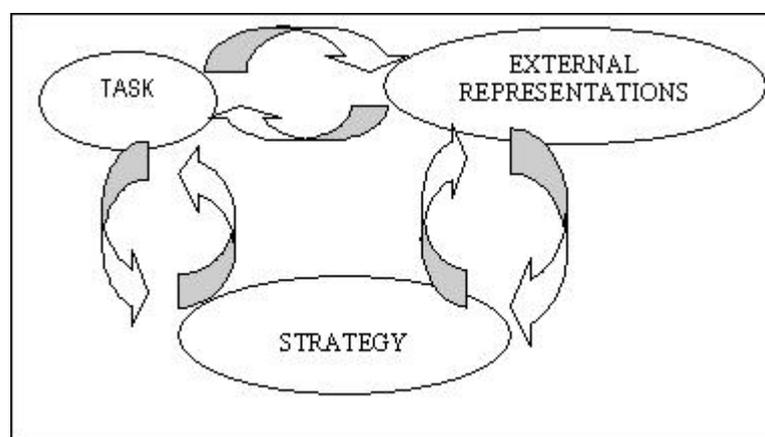


Figure 1. - Model of the interaction between the users, their task and the external representations during the process of searching information in the web.

2. - STUDY

The aim of this study was to identify more precisely the variables involved in the searching process and their importance. We investigated the interaction of several variables of searching, including user's experience and their cognitive strategies. We manipulated the type of searching task among participants, who had different levels of web expertise (novice and more experienced).

Tasks: Four types of tasks were chosen for this study (see Table 1). To study the effect of the type of information, we defined two different task scenarios, based on Shneiderman's (1997) definition: one specific fact-finding (e.g., for Computer Science students, to look for database algorithms in Java), and another exploration of availability (e.g., find all the available jobs for a specific profession). We were also interested in exploring the effects of how information is structured in the Web on user's searching behaviour. We identified two different tasks for each of these scenarios. In one of them, the information is dispersed through out the Web and cannot easily be found in any category or general resource site (e.g. find all the information available about the Nobel Prize 1997 for Literature). In the other task, the information was structured in categories that are easy to identify from the main search engines (e.g. look for definitions of several words). All the search tasks were performed in the Netscape Communicator 4.5 browser. The participants could use any search engine that they wanted to perform their searches.

SEARCHING CONDITIONS	FACT FINDING	EXPLORATORY
DISPERSED STRUCTURE	<ul style="list-style-type: none">● Look for data base algorithm in Java● Look for criteria for the diagnosis of diseases	<ul style="list-style-type: none">● Find all the available jobs for profession
CATEGORY STRUCTURE	<ul style="list-style-type: none">● Look for word definition	<ul style="list-style-type: none">● Find all information about 1997 Nobel Prize for Literature

Table 1. The four searching conditions of the study.

Participants: Twenty-three volunteers participated in the study. All of them were students at the School of Cognitive and Computer Science, University of Sussex, U.K. Ten participants were Computer Science students, and thirteen were Psychology students. This mixture allowed us to compare the results between participants with different knowledge and experience about Web and computers in general.

Measures: Because of the exploratory nature of this study, we used observational methods with interviews. During the 30 minutes in which the participants were performing the search task, the experimenter took

notes of their searching steps. At the end of the searching session the experimenter asked the participants to verbalise why they had performed each of these steps and the main problems that they experienced. Both the searching session and the interviews were video recorded. This approach is effective for providing descriptive information about the participants' strategies in web searching (Pejtersen and Fidel, 1998). We also asked the participants to fill in a questionnaire to get information about: (1) experience with computers, web and information databases, (2) what they remembered about their search paths, (3) knowledge about how web search works, and knowledge about the searching domain, (4) level of satisfaction with the search and any comments or problems that they wanted to specify.

3. - FINDINGS

We collected data from: 1) questionnaires about web searching, 2) observational studies about participants' performance, 3) post-task interviews. First we summarise the information from the questionnaire. Following this, we will explain the main findings regarding the participants' cognitive strategies. Then, we will present our model for Web searching both for novice and experienced participants. Finally, we will highlight some of the problems and interpret them from the perspective of the external cognition approach.

3.1. - QUESTIONNAIRE RESULTS

1. - Experience with computers and the Web.

Participants experience with the Web: all the Computer Science students have more web experience (on average 2 years), and use it for more complex searches, in comparison with the Psychology students (who have been using the Web, on average, for one year, and then only for course work). Some of the Psychology students were also found to have used the Web only in the last 3 months.

2. - Knowledge about how search engines work with the web.

Again there was a big difference between the two groups. Most of the Computer Science participants describe quite well how the search engines develop their databases (normally in terms of collecting web pages and keywords), and how they look for the information in the database during the search. On the other hand, only one of the Psychology students knew quite well how search engines work. In contrast, neither of these groups have a clear idea of how the search engines use the queries to look for information and only two participants refer to the functionality of the engines.

3. - Level of satisfaction

In two questions the participants were asked about their level of satisfaction with their results in the search and with their performance in the search. They had to rate their satisfaction on a scale of 5 points (Very good, good, ok, bad, very bad). Most of the participants, 17, considered their level of satisfaction in both questions to be "good" or "ok".

4. - What the participants remember/forget about their searches.

In general most of the participants were not very accurate in remembering their searches. Only two of the participants remembered all the engines and queries used and the results found. Interestingly, participants tended to forget search engines and queries that did not give any successful results and some participants even falsely remembered a systematic pattern in the queries that they had used which did not correspond with their actual behaviour when searching. This suggests that participants organise their memory about their searches in logical steps even though they don't follow them. There was also a recency effect: several of them remembered only the last search, or remembered better the last search.

3.2. - SEARCHING STRATEGIES

Combining the observational data about participant behaviour through the Web with the information that they provided us in the interviews (about what they were looking for and why), we identified three different general patterns of searching. We were specifically interested in these patterns because they reflect the kind of cognitive strategies used by the participants. Interestingly, the use of these strategies is associated with the kind of search task, especially with how the information was structured in the Web, and with the participant's experience with Web searching.

First, we describe the strategies and their relationship with the other variables. Examples of participants' searches, which illustrate each of these strategies can be found in Appendix 1.

1. - *Top-down strategy:*

A top-down strategy is when users search in a general area and then narrow down their search from the links provided until they find what they are looking for. Typically, participants using this strategy are looking for a very general site, which contains a list of facts organised in meaningful categories. For instance, a participant looking for Data Structure Algorithms in Java looked inside Sun home pages for a site with general resources of algorithms in Java. Another example is that when the participants were asked to find in which context they would use some very unusual English words, they looked for an English dictionary or thesaurus. They started clicking in a category of the browser or introducing a very general query and following the links from there, trying to narrow down until they found the specific information that they were looking for.

2. - *Bottom-up strategy:*

In contrast with the top-down strategy, the bottom-up strategy is when users look for the specific keyword that they were provided with in the instructions. Using this strategy, participants directly typed the very specific keywords in the search engine and scrolled through the results, opening one link and coming back to the list of results until they found the desired information. This strategy was most often used by experienced participants, for the specific fact-finding searches.

3. - *Mixed strategy:*

Many of the participants used both of the above strategies in parallel, searching for required information at the same time in multiple windows. Some of them alternated strategies, having 'both in mind' during their search. This strategy was only used by the experienced participants.

To give a clear overview of our main findings, we have summarised them in Table 2 in terms of the main strategies of the participants depending upon the searching task and their experience with the web. Interestingly, the kind of searching tasks (fact-finding vs. exploration) had a stronger influence with the experienced participants than with the novices. Therefore, it seems that some knowledge about Web searching is needed before participants can identify the differences between tasks. On the other hand, experience seems to facilitate the participants' knowledge about how to start the search and about how to select the most appropriate strategy for each situation.

SEARCHING TASK	EXPERIENCED	NOVICE WEB- PARTICIPANTS
-------------------	-------------	-----------------------------

	WEB-PARTICIPANTS	
INFORMATION IN WEB DISPERSED STRUCTURE (e.g. find criteria for a psychological disease)	SPECIFIC FACT FINDING: <ul style="list-style-type: none"> • Bottom-up • Mixed strategy at the beginning and selecting Bottom-up 	<ul style="list-style-type: none"> • Start with top-down and change at the end to bottom-up • Start typing without knowing why
	EXPLORATORY: <ul style="list-style-type: none"> • Top-down 	
INFORMATION IN WEB CATEGORY STRUCTURE (e.g. find a job opening)	<ul style="list-style-type: none"> • Mixed strategy at the beginning and the selecting top-down • Top-down 	<ul style="list-style-type: none"> • Top-down following browser categories • Start with bottom-up and change to top-down

Table 2. - Interactions between participants' level of experience, the searching task and the predominant strategies in each of these groups.

The next stage of our analysis is to examine the interactions between the different aspects of our Interactivity Model (i.e. user, task, and environment).

To understand these interactions in more detail, we will summarise the general strategies of the participants under each of the four task conditions, paying special attention to the effects of the experience in each condition.

3.3 - WHEN THE DIFFERENT SEARCHING STRATEGIES ARE USED

1. - INFORMATION IN WEB DISPERSED STRUCTURE/FACT-FINDING:

(Searching task: Looking for psychological diseases or data structure algorithms)

In this task we found a clear difference in strategy depending upon the experience of the participants. On the one hand, most of the experienced participants either directly started typing the keywords or names of the algorithm or diseases they were looking for, or chose a mixed strategy. In the interviews, these participants pointed out that they were trying to find the more successful way of looking for that material. Therefore, they developed a plan about how they were going to search and were flexible, choosing the more successful strategy. On the contrary, novice participants typically started with very general queries, for instance "Psychology" or "Diseases", and gradually narrowed down the search, adding the words suggested from the

search engines. Other times they followed the links and categories suggested. This finding suggests that the external representations presented in the web pages by the search engines influenced more the novice participants.

2. - INFORMATION IN WEB DISPERSED STRUCTURE/EXPLORING:

(Searching task: Looking for all the information available in the web about the 1997 Nobel Prize for Literature)

In performing this task, several differences were raised again between the overall searching of the experienced participants and the novice participants. On the one hand, the novice participants started looking with queries, which brought back thousands of results (like "Nobel Prize"). When they were asked why they searched using that specific query, all of them reported that they did not know why, and they were not following any planning or strategy. On the other hand, searching behaviour developed by experienced programmers was more complex, diverse, following a top-down approach. Experienced searchers, therefore, tended to search in a more structured way, and planned in advance more than the novice participants.

3. - INFORMATION IN WEB CATEGORY STRUCTURE/FACT-FINDING

(Searching task: Looking for the context in which you would use some very unusual English words)

In the case of this searching condition, the experience level of the participants did not seem to have as strong influence as it had in the previous tasks. Most of the participants in the "English words" condition showed a clear top-down strategy, looking directly or after only one try of typing a specific keyword, for a dictionary or a thesaurus.

4. - INFORMATION IN WEB CATEGORY STRUCTURE/EXPLORING

(Searching task: Looking for job openings in a specific area):

Under the 'Job' condition also most of the participants started with a clear top-down approach, the searches were very different from each other. While some participants went to a general category of Jobs (some variations were Job hunting, or Careers) and from there narrowed down the search to a specific area, others started looking for a very general area and then inside this area introduced 'jobs'. The way in which they tried to narrow down the search was also very different amongst participants. Some of them preferred to follow the subcategories suggested by the search engines and some others used more specific queries.

4. - FROM THE DATA TO A MODEL OF SEARCHING

From our analysis of the results of the study we have identified interactions among the three dimensions described in our Interactivity Framework. These are: the task, the user's strategies and the external representations provided to the users. Our next step is to conceptualise these interactions in a Model that could allow us to make predictions about the participants' searches. Following our results, we have constructed a model for the experienced participants and another for the novice participants.

Figure 3 shows the model for the experienced participants. As we see in that model, participants first start with a plan for their searches. In this plan they take into account how the information that they are looking for is organised in the Web. They also consider their goal for the search. These steps should not be considered necessarily as serial processing and experts seem to evaluate both variables to direct their searches.

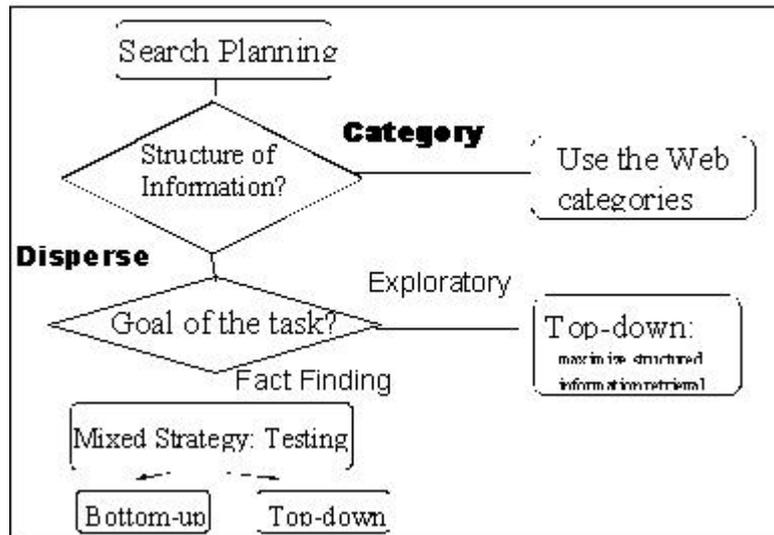


Figure 3. - Web searching Model for experienced participants.

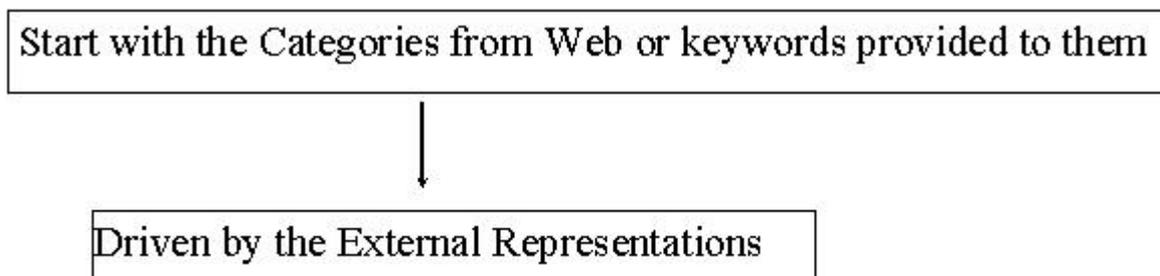


Figure 4. - Web searching Model for novice participants.

On the other hand, as we can see in Figure 4, novice Web participants do not seem to start with any kind of planning. Novices have shown themselves to be highly influenced by the External Representations presented to them. Therefore, our focus of analysis should be on the specific characteristics of the relationship between the internal representations and the external representations, and the cognitive processing involved. This is exactly the focus of the External Cognition framework. We claim that, in order to understand the Web searching tasks, we need to analyse how the information presented to the participants (External representations) interact with the dimensions defined by this framework. In our study we have found data supporting that the representations currently used by the main search systems in the Web are the cause of multiple problems regarding each of these dimensions.

First, we need to consider how external representations that have the same abstract structure, but different surface structures, could make the distinction between the relevant and the irrelevant information easier or more difficult (Re-representation dimension). In our case we found that the external representations

presented to the participants from the diverse search engines that they use made them very hard to recognise the relevant information. Many subjects either save irrelevant information or erroneous information or do not save the relevant information required from the task.

We also need to evaluate how these external representations constrain the kind of inferences made by the participants about the underlying represented world (Symbol constraining dimension). For instance, some participants got lost in their searches because they made erroneous inferences about the meaning of opening a link of a subcategory.

In addition to the low computational offloading, we should also recognise that other forms of cognitive overload can occur. For instance, our participants have trouble in remember the content of each window when they had more than three windows open. These problems could be avoided if the external representations would make more visible the correspondence between each window and the results display in that window through the whole search session (Temporal and spatial constraining dimension).

GENERAL CONCLUSION

The following conclusions of this paper are related to the objectives for this study. First, we wanted to develop a theoretical framework that could explain web-searching behaviour. We found that our proposed three dimensional model has been useful in analysing the interaction between participants, their task and the external representations. These data support our claim about the necessity of expanding the concept of interactivity as is commonly used now to account for the interaction between multiple factors. Specifically we have found that the cognitive strategies developed by participants depend on the way in which the information they are looking for is structured, as well as their level of experience. These interactions were used as an empirical base for modelling the searching behaviour of web participants. Further research is needed to investigate in more detail these cognitive strategies, in order to be able to develop a complete model of this searching process. On the other hand, the analysis guided by the External Cognition approach has proved to be useful in the analysis of the interaction between the participants' internal representations and the external representations. We claim that this approach could be complementary to the development of a search model in the analysis of the interaction at the level of the representations (e.g. to analyse why users made some erroneous inferences).

ACKNOWLEDGEMENTS:

The authors gratefully acknowledge the support from the EPSRC Cooperative Technologies for Complex Work Settings project (TRM number ERBFMRXCT960014).

BIBLIOGRAPHY

Alty, J. L. (1991) Multimedia - What is it and how do we exploit it? In. D. Diaper and N. Hammond (ED.) *People and Computers VI*. CUP: Cambridge. 29-44.

Branham, C. (1997) A student's Guide to Research with the WWW. <http://www.slu.edu/departments/english/research/>

Buckingham Shum, S. (1996) The missing link: Hypermedia usability research & the Web. *Interfaces, British HCI Group Magazine*, Summer, 1996.

Kellogg, W. A.; Richards, J. T. (1995). The human factors of information on the internet, In Nielsen, j. (Editor), *Advances in Human-Computer Interaction: Volume 5*, Ablex Publ., Norwood, NJ, 1-36.

Marmollin, H. (1991) Multimedia from the perspectives of psychology. Proceeding of the Eurographics Workshop on Multimedia. Stockholm.

Nielsen (1997) Search and You *may* find

<http://www.useit.com/alertbox/9707b.html>

Nielsen (1999) Differences between print design and web design. <http://www.useit.com/alertbox/990124.html>

Pejtersen, A.M.; Fidel, R. A framework for work centred evaluation and design: A case study of Information Retrieval on the Web. Working Paper for MIRA workshop.

<http://www.dcs.gla.ac.uk/mira/workshops/grenoble/fp.pdf>

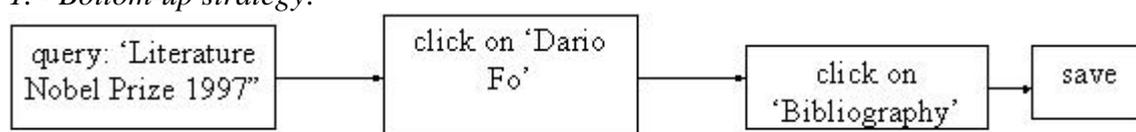
Scaife, M. and Rogers, Y. (1996) External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*.**45**, 185-213.

Shneiderman, B., Nyrd, D., Croft, B. (1997) Clarifying search: A user interface framework for text searches. *Dlib Magazine* (January, 1997).

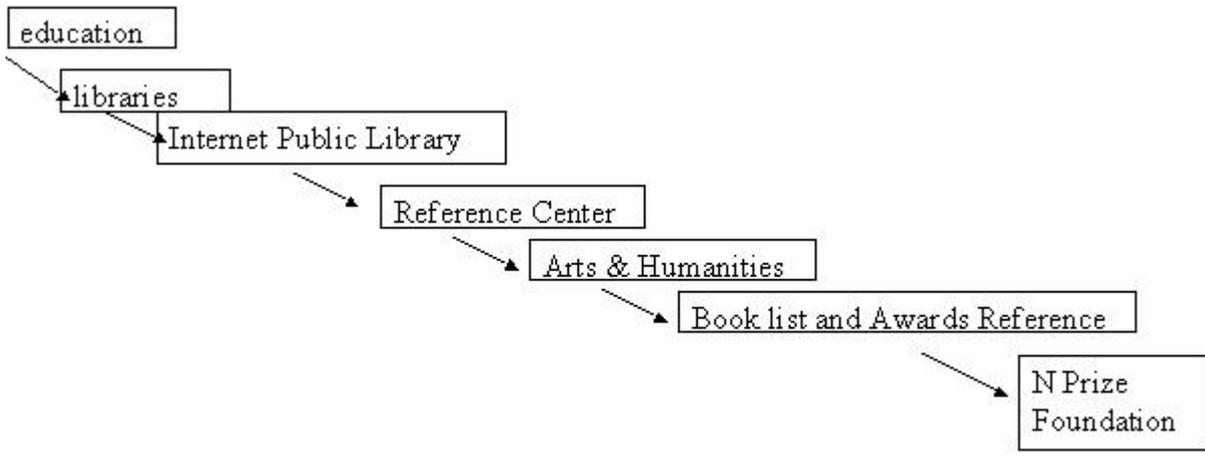
Shneiderman, B. (1997) Designing information-abundant Web sites: issues and recommendations. In S. Buckingham Shum and C. McKnight, Eds. "Web Usability" (special issue) *International Journal of Human-Computer Studies*, **46**.

APPENDIX 1: examples of users' searches for 1997 Nobel Prize in Literature

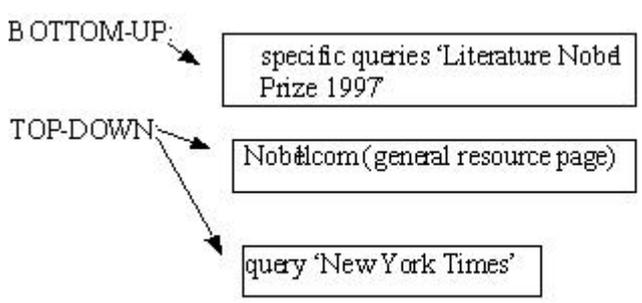
1. - Bottom-up strategy:



2. - Top-down strategy:



3. - Mixed strategy:



"Cognitive strategies in Web searching"

[<- Back](#)

Thanks to our conference sponsors:



Thanks to our conference event sponsor:



Site Created: Dec. 12, 1998
Last Updated: June 10, 1999
Contact hfw@nsl.gov with corrections.