# Complex and Exploratory Web Search

Anne Aula
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
1-650-253-8975
anneaula@google.com

Daniel M. Russell
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
1-650-253-2682
drussell@google.com

## ABSTRACT
We suggest that there are two undiscussed dimensions to information searches: *complexity of the information need* and *clarity of what is needed to satisfy the information need*. Among the factors that affect the complexity of the information need is the amount of information that is needed and the number of steps that are required to collect the needed information. In addition, storing found information becomes increasingly important as the complexity of the task increases (too much information to hold in memory). Clarity of the goal refers to the searcher's understanding on the pieces of information that are needed to satisfy the information need. We show how *exploratory search* may sometimes be complex, but is not necessarily so, and is characterized more accurately by the degree of clarity the searcher has about the goal. *Complex search* tasks often include exploring the topic, but do not necessarily require exploration or may require exploration only in certain phases of the search process. We suggest that complex search tasks – especially those where the goal-state is initially unclear - are the ones where the current search tools do not offer enough support for the users.

## Keywords
web search, complex search, exploratory search

## 1. INTRODUCTION
Over the years, a number of different taxonomies for classifying information seeking tasks have been proposed. For example, Rose and Levinson [6] proposed a taxonomy with three different search goals (navigational, informational, resource). For informational tasks, they included five sub-categories: directed (closed vs. open), undirected, advice, locate, and list. Similarly, Kellar et al. [2] proposed a web information classification with six different information tasks (fact finding, information gathering, browsing, transactions, communication, maintenance). In Rose and Levinson's taxonomy, undirected informational tasks and in Kellar's classification, information gathering are close to what other researchers sometimes call exploratory search tasks. [4][1]

Marchioni [4] grouped search activities into lookup, learn and investigate activities. According to Marchioni, exploratory search is especially pertinent to the learn and investigate activities. White et al. [1] propose that exploratory search tasks often require the users to submit a tentative query and use the results to find cues about the next steps. "Defining what constitutes an exploratory search is challenging…in exploratory search, users generally combine querying and browsing strategies to foster learning and investigation" [p.38].

The term "exploratory search" is generally used to refer to the challenging search tasks where the user's goal is to learn something by collecting information from multiple sources. We propose a novel classification whereby we differentiate between "exploratory search" and "complex search" by their relative levels of search goal complexity and search goal clarity. In our classification, exploration occurs when the searcher is unsure about what to do next, or has to learn more about the domain of interest. Exploratory search is often an intrinsic part of complex search tasks—but exploration is usually not the end goal—it is a property of the state of knowledge on the part of the searcher.

## 2. WHAT MAKES A SEARCH TASK COMPLEX?
We suggest *complex search task* as a term to refer to cases

- where the searcher needs to conduct multiple searches to locate the information sources needed,
- where completing the search task may span over multiple sessions (task is interrupted by other things),
- where the searcher needs to consult multiple sources of information (all the information is not available from one source, e.g., a book, a webpage, a friend),
- that often require a combination of exploration and more directed information finding activities,
- which often require note-taking (cannot hold all the information that is needed to satisfy the final goal in memory), and
- where the specificity of the goal tends to vary during the search process (often starts with exploration).

Earlier research has suggested that the current search tools are optimized for simple lookup or directed search tasks [3] and thus, these tasks are relatively easy to accomplish with the current tools. However, when we look at these simple tasks as parts of a complex search task, using the easy-to-find information to fulfill the overall search goal (that might require tens of little fact-finding tasks) may be very challenging. We have to think of search *tasks*, not individual searches or even search sessions. Individual searches might be relatively more-or-less sophisticated, but we need to analyze search behavior at the task level, not the individual query level. And we know that tasks might be broken up across multiple sessions spanning hours or days. [3]

## 3. EXPLORATION vs. COMPLEXITY
We propose two measures of search tasks that describe dimensions of user behavior during search.

1. *Goal abstraction level (GAL)*—this is a measure of how close the goal is to actionable behavior. That is, GAL is the distance

from the current level of understanding about the goal to a point where the goal can be executed (in terms of the fundamental domain operators). For example: an information goal of "find the root causes of the US Civil War" requires a decomposition to many discrete actions that can be taken by the searcher. Ultimately, a search is deemed exploratory when the searcher has a very abstract search goal—*i.e.*, when the GAL level is high. When observing the users when they are searching, an abstract (unspecified) goal can often be identified by the users' statements like "I'm interested in learning more about…"

2. *Search moves (SM)*—is an estimate measure of the number of web search steps it will take achieve the goal (i.e., how complex the search task is). That is, SM is a heuristic that can be used to estimate the number of search "moves" it takes to achieve the goal, given the searcher's current state and background knowledge. A move could be a search, or an information extraction subtask, but the idea is that SM is a factor that the searcher takes in to account when considering the next move to make in their search task.

For example, an overall task goal of "Find information that is needed to make the move to Helsinki successful" is too abstract to act on—thus, the searcher needs to decompose the goal into actionable sub-goals. For example, one such sub-goal could be: "find the best neighborhood to live in Helsinki". This sub-goal, in turn, has to be broken down into smaller goals (e.g., "find a listing of Helsinki neighborhoods and reviews/comments on each of them"). For the task of listing the neighborhoods, the goal is concrete and clear (list the neighborhood names and their reviews side-by-side); however the SM estimate (task complexity) could be quite high, including finding sites that list the neighborhoods (possibly on multiple sites), then collecting comments and reviews on each (again, potentially multiple sites, potentially requiring reconciliation between different opinions and reviews).

Clearly, both GAL and SM are heavily influenced by searcher expertise and domain knowledge. A search task such as "Find the root causes of the US Civil War" is a very high-level goal that will require a great deal of work to concretize—for a recent immigrant to the US. But a searcher with the background knowledge of a history undergraduate degree would see this as a much different kind of task.

## 4. MEMORY LOAD SUPPORT

Task goals that are both complex and exploratory cause large memory and cognitive demands on the searcher As the searcher discovers multiple exploration options and strategies, an effective tactic is to capture those options in some kind note. Similarly, in multi-session searches notetaking often bridges the gap of carrying information learned in one session to the next. This is especially true in search tasks that span multiple days.

In Figure 1 we've tried to capture this as the "Notetaking boundary," when exploration and complexity impose sufficiently large memory loads to require external capture.

We note that many systems have attempted to support searcher memory demands. Bookmarks are common, but flawed mechanism for storing information [1]. A number of notetaking systems (e.g., Google's Notebook product) have been attempted with varying degrees of success. Methods for capturing previously searched queries, context and history (and then making them available to the searcher) have been tried as well. [7]

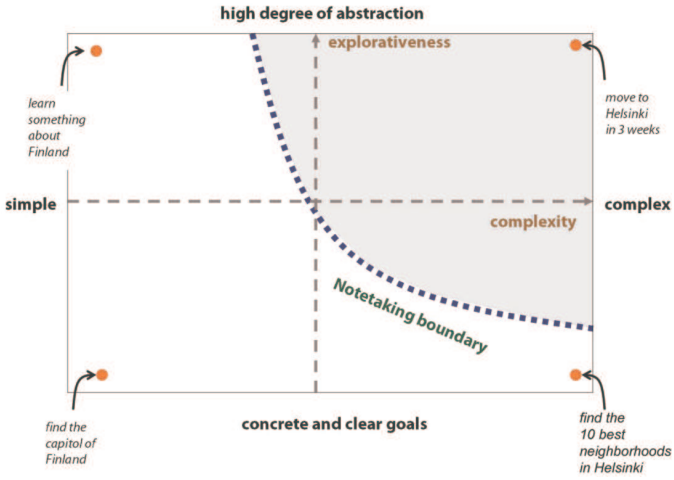


**Figure 1:** As a search task can vary from simple to complex, depending on the intrinsic complexity of the information being sought and the amount of time it takes to satisfy that information task. But as the searcher works, they may be more or less clear about what they need to do to satisfy their information needs. Exploration occurs when the searcher is unsure about what to do next, or has to learn more about the domain of interest. The Notetaking boundary marks the place where the memory loads of the task (due to both complexity or exploratiVeness) become so large as to require external memory aids to allow the searcher to succeed.

It seems clear to us that we have yet to identify the effective and powerful methods to support the memory demands that are especially pertinent to searches that are high both in complexity and in the degree of abstraction.

## 5. REFERENCES


[1] Aula, A., Jhaveri, N., Käki, M. 2005. Information search and re-access strategies of experienced web users. In Proceedings of the 14th international conference on World Wide Web, 583-592.

[2] Kellar, M., Watters, C., & Shepherd, M. (2006). A Goal-based Classification of Web Information Tasks. In the Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T 2006), Austin, TX. (CD-ROM Proceedings)

[3] MacKay, B. & Watters, B. (2008) Exploring multi-session web tasks. Proceedings of CHI'2008, 1187-1196.

[4] Marchioni, G. (2006) Exploratory search: from finding to understanding. Communications of the ACM, 49, 4, 41-46.

[5] Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

[6] Rose, D.E. & Levinson, D. (2004) Understanding user goals in Web search. Proceedings of the WWW 2004, 13-19.

[7] Rose, D. E. and Raju, S. (2007). Encouraging Exploration with Elroy: A New User Experience for Web Search. Proc. of ACM SIGCHI 2007 Workshop on Exploratory Search and HCI (ESI 2007), 52-55.

[8] White, R.W., Kules, B., Drucker, S.M., and Schraefel, M.C. (2006) Supporting exploratory search. Communications of the ACM, 49, 4, 37-39.