From Web Search to Exploratory Search: Can we get there from here?

Peter Anick
Yahoo! Inc.
panick@yahoo-inc.com

1. INTRODUCTION

A common complaint of users of search engines is that they deliver too many results. In a recent study of biology teachers, for example, 34% of respondents indicated "too many results" as a primary reason for not using search engines (Perrault 2007). At first blush, this seems like an odd criticism. After all, the results are ranked by relevance and one can always narrow the results by refining the query. It is likely, therefore, that this complaint really reflects the frustration of users when they are engaging in exploratory searches. When a user knows the domain, it's easy enough to assess the relevance of documents via titles and snippets or refine a search that is too broad in scope. But when the user is less familiar with the subject matter, such assessments and refinements can be considerably more difficult. As Marchionini (2006) has noted, exploratory search is more about learning and investigating than lookup.

For this reason, developers of exploratory search interfaces have looked to devices that help users organize and scan documents. Kules and Shneiderman (2006), for example, advocate organizing web search results by "meaningful and stable categories" while clustering interfaces (e.g., clusty.com) partition result sets into categories constructed on the fly. In this position paper, we consider how some of the tools already developed for general web search, such as terminological search assistance and snippet generation, can support the learning and investigative functions of exploratory search, given that a user engage with them from the proper perspective. Terminological search assistance, such as that offered in Yahoo's Search Assist tool (described in the next section), does not provide stable categories, nor are the terms necessarily "meaningful" to a user unfamiliar with the domain. However, term suggestions can provide a convenient way to navigate through a dense and unfamiliar result set when used in conjunction with document snippets, the short fragments of web pages generated to provide a succinct summary of each of the top matched results. Snippets are finely tuned to best express how each document is relevant to the query. Ideally, they contain fragments of text that capture the query terms in close proximity and span just enough lexical context to allow the user to glean the sense of the terms.[1] These aspects of snippets make them quite useful in the "learning" phase of an exploratory search.

[1] Many users have noted that queries posed as explicit questions often produce some snippet containing the answer.

The paper is organized as follows. First, we'll introduce the features of Yahoo's Search Assist tool that make it particularly amenable to exploratory search. Then we'll review two example scenarios. Finally, we'll consider some of the issues and next steps if we were to evolve these web search features into a tool more tuned to supporting exploratory search.

2. SEARCH ASSIST

Yahoo's Search Assist interface (version 1.0) combines two refinement tools within a single display tray. We will refer to these tools as *Gossip* and *Viewpoint*. *Gossip* offers suggest-as-you-type expansions for a user's query string on a character by character basis (as displayed in the left column in figure 1). Its suggested query completions are derived from frequently occurring queries mined offline from search logs. Clicking on a *gossip* suggestion has the effect of replacing the query in the search box and executing the search.

Once a search has been run, *Viewpoint* offers up related terms which are derived from an analysis of the top search results. Clicking on a *viewpoint* suggestion effectively AND's the clicked term with the current query and generates a new result set. However, in this case, the interface keeps the query and refinement terms separate. Rather than merge the refinement into the query box, the interface instead places it into the left hand side of the tray (under the caption "Showing results containing" in figure 2) and re-populates the right-hand-side with new related terms generated from the revised query. Selecting a new suggested term replaces the previous selection in the composite query, i.e., it ANDs the new selection with the original query. The interface thus aims to help the user deal with the problem of "too many results" by allowing him to pick one item at a time to focus on within those results. Note that this behavior differs from most clustering interfaces, which create a single document partition based on the initial query. By refreshing the terms after each term selection, the viewpoint model effectively enables the user to move about within the domain of the initial query.

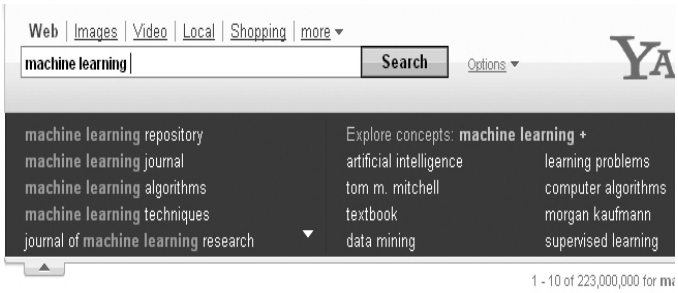

Figure 1. Search Assist tray showing *gossip* suggestions on left and *viewpoint* concepts on right after running a query.

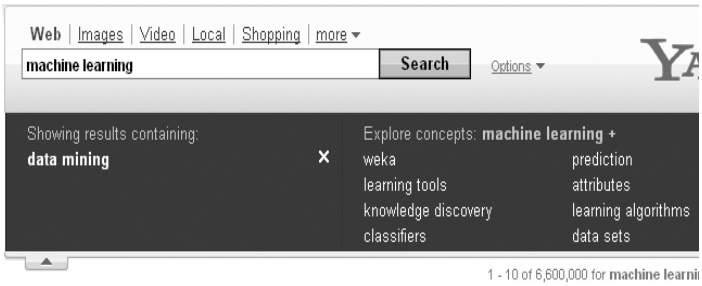

Figure 2. Search Assist tray after clicking on *viewpoint* suggestion "data mining" in right hand side of figure 1.

3. FOCUS AND SKIM

In this section, we present two short hypothetical scenarios to show how the tool may be used to support an exploratory search strategy of focusing on one term suggestion at a time and skimming snippets to quickly glean domain knowledge prior to (or even in lieu of) visiting any actual documents. For each example, we provide several *viewpoint* "focus" terms that might logically be selected and show samples of the snippets returned as a result of each search.

Example 1. Having heard a Moroccan *rebab* player at a recent conference, the user is curious to learn more about the instrument and the contexts in which it is played. He knows next to nothing about Moroccan music.

Initial query: **rebab morocco**

Focus term 1: "**Moroccan music**"

Snippets:

> **Moroccan music** is of many types; it includes Arab, Berber, classical, and **...** instruments - the **rebab**, played by the leader and considered **...**
>
> **...** master musicians performing on the **rebab**, oud, violin, viola, tar, and darbouka. **...** illustrates the Arabic, Judaic, and sub-Saharan roots of **Moroccan music**. **...**
>
> **Moroccan music** reflects the country's hybrid culture, blending Arabic, African, **...**

Focus term 2: "**Berber Music**"

Snippets:

> **...** instruments - the **rebab**, played by the leader and considered the most important; **... Berber music**, even more closely linked to poetry than Arab music, is usually **...**
>
> **Morocco's** Berber villages are rich with music. **... Berber Music** Today **...** Raiss instruments typically include the **rebab**, a one-stringed fiddle, not same **...**

Focus term 3: "**ahwash**" (selected because it is an unfamiliar term)

Snippets:

> The Berber **rebab** should not be confused with the **rebab** andalusi, a two **...**
>
> **... Ahwash** is exclusively village music, probably unchanged for centuries or longer. **...**

Example 2: This example is based on the query used in Michael Levi's position paper, "Musings on Information Seeking Support Systems". We assume the user has little background in economics

Initial query: "**Is the US in a recession?**"

Focus term 1: "**definition of recession"**

Snippets:

> A period of general economic decline; specifically, a decline in GDP for two or more consecutive quarters.

Focus term 2: "**GDP**"

Snippets:

> But what **is GDP**? Gross domestic product is the value of all final goods and services produced in **...**
>
> Imports, which are a subtraction in the calculation of **GDP**, decreased. **...**

Focus term 3: "**real GDP**"

Snippets:

> **...** the "two quarters" method indicates **the US is** not **in a recession** technically, **...**
>
> **...** A second point is that **real GDP is** not reported in a timely fashion. **...**

Focus term 4: "**NBER**" (unknown term for this user)

Snippets:

> **...** official judges from **the National Bureau of Economic Research** (**NBER**) pour through various **...**
>
> **The National Bureau of Economic Research**, an official panel of senior economists, has declared that **the US** entered **...**

Focus term 5: "**economists**"

> 60.8% of **economists** believe **a recession is** here or approaching
>
> **Economists** Disagree on Whether the Country Actually **Is in a Recession**

4. OBSERVATIONS

These two short scenarios illustrate how a web search user might go about learning domain vocabulary or identifying key facts through interaction solely with term suggestions and snippets. While not all suggested terms lead to relevant information, it is usually possible to judge the importance of unknown terms from skimming the snippets returned, some of which are likely to include either explicit definitions or contextual information that serves to define them.

What might a more sophisticated interface, specifically designed for exploratory search, do to better support a "focus and skim" interaction model?

- Provide longer snippets or a mechanism to expand snippets for more context. As proposed in White et al (2005), it may even be reasonable to organize the display as a collection of top ranked sentences rather than as per-document abstracts.
- Weight term suggestions so as to prefer those that are more domain specific. This is less useful for general query refinement where terms like "images" and "reviews" might make perfectly appropriate refiners, but for exploratory search purposes domain specific terminology should in principle offer the user more opportunities to learn.
- Provide tools for marking and remembering useful snippets. This would allow the user to organize snippets (and hence documents) by their focus terms, generating a dynamic "table of contents" through which to visit the documents themselves in an orderly fashion.
- Help the user choose a suitable initial query. The initial query should be broad enough to circumscribe the domain of interest while narrow enough to eliminate ambiguities or irrelevant aspects that might lead to poor term suggestions.
- Offer more term suggestions on demand, sorted into categories where appropriate.

5. CAN WE GET THERE FROM HERE?

We have argued that some key elements of exploratory search support may already exist in web search engines as a side effect of efforts to support query refinement and search result assessment. It remains to be seen if users themselves will recognize that such features can be exploited in the manner we have suggested. Log analysis reveals some cases of iteration over viewpoint terms (there is a 20% chance that a session containing one viewpoint term selection will contain a second selection) but real life examples such as the scenarios we presented here are difficult to validate from logs alone since it requires determining whether the lack of a click on a document reflects sufficient knowledge gleaned from snippets vs. dissatisfaction with the search results.

What is known is that users have such deeply ingrained habits for their conduct of web searches that it is difficult for them to integrate, or even notice, new features on a search results page. Naïve users are typically reluctant to interact with features whose consequences are uncertain while sophisticated users tend to construct erroneous interpretations of how a feature works and then use it sub-optimally. The greatest challenge in the development of systems to support exploratory search will be to bring the mental model of the user in line with the actual behavior of the tool.

6. REFERENCES

Levi, M. D. (2008). Musings on Information Seeking Support Systems, ISSS 2008 position paper.

Marchionini, G. (2006). Exploratory Search: From finding to understanding. Communication of ACM, 49(4): 41--46, 2006.

Perrault, A. M. (2007). An exploratory study of biology teachers' online information seeking practices. School Library Media Research, 10. www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume10/biology.cfm

Kules, B and Shneiderman, B. (2006). Improving the search experience by organizing Web search results with meaningful and stable categories. HCIL Symposium, University of Maryland, College Park, 2006

White, R. W., Jose, J. M. and Ruthven, I. (2005). Using top-ranking sentences to facilitate effective information access. *Journal of the American Society for Information Science and Technology*, 56 (10), pp. 1113-1125.