Collaborative Exploratory Search

Jeremy Pickens and Gene Golovchinsky

HCIR 2007
Workshop Boston

INTRODUCTION

Most modern information retrieval (search) systems are geared toward helping a user quickly and effectively find a particular item. That item may be a document, a geographic location, a factoid, etc. This approach is good when a single piece of information can fulfill the user's information need.

In many situations, however, multiple items and/or richer overviews of the entire information space are necessary. In support of this, exploratory search systems have been developed. Exploratory search systems typically blend a variety of information seeking tools and tactics (e.g., querying, browsing, document clustering, etc.) to help the user better understand the range of available information. Such tactics are combined in opportunistic ways as users' understanding of their information needs evolves [Bates, 1989],

Currently, there is also much work around mass social collaboration, or aggregation of large-scale user intention and information seeking behaviors. Recommender systems such as Amazon [cite, Glinden], personalization systems based on behavioral or topical clustering such as Kaltix (Google) [cite, 2003], community-mediated search collaboration such as iSpy [cite Barry Smyth] and bookmarking and voting sites such as del.icio.us and Digg all make use of the "wisdom of crowds" approach to collaboration. The underlying commonality among these systems is the notion of using correlated aggregate behavior to steer the individual searcher toward the most relevant pieces of information based on results obtained by others. And individual, working alone, is implicitly informed by prior search behaviors of the crowd.

But what if a searcher is looking for novel information, or for information that does not have a large peer group from which to draw recommendations? What if the searcher's goals are different from others who formulated similar search requests? The problem with the crowd-based approaches is two-fold: there may be large numbers of documents in a system with no prior user attention, and the information need of the crowd might not match the need of the current searcher. These disparities may limit the effectiveness of the otherwise desirable strategy of using multiple people's input to determine the outcome of a search session.

We propose to mitigate the deficiencies of correlated search with collaborative search, that is, search in which a small group of people shares a common information need and actively (and synchronously) collaborates to achieve it. Furthermore, we propose a system architecture that mediates search activity of multiple people by combining their inputs and by specializing results delivered to them to take advantage of their skills and knowledge.

For example, the goal of previous crowd-based systems is to help the current searcher discover relevant items that others have already discovered. Interesting technological challenges involve discovering useful crowd clusters or latent spaces, figuring out which prior users' actions are relevant the current user's information need, etc. In an explicitly collaborative, small team-based environment, on the other hand, no guesswork is needed for user search behavior clustering; one's collaborators are known. The focus of the system shifts from helping individuals re-find information that has already been found by someone else to helping a team member find information that no other member has yet found, but that is relevant to the overall information needs. Explicit collaboration therefore allows a fundamental conceptual shift in system design, away from algorithms and interfaces that support re-finding and re-discovering to algorithms and interfaces that support new discovery and exploration.

	Explicit	Implicit
synchronous	Collaborative exploratory search (FXPAL)	Realtime awareness and continual-update context systems (e.g. Imity, or)
asynchronous	"search Trails" [1]	Web 2.0, (recommender systems, Google's Kaltix, wisdom of crowds)

Table 1. Collaborative Exploration Search design space.

THE DESIGN SPACE

We can characterize the design space along two dimensions, as shown in Table 1. Search can be explicit or implicit, asynchronous or synchronous. The main difference between explicit and implict collaboration is

whether people to work toward the same goal (explicit) or similar but independent goals (implicit). The key difference between synchronous and asynchronous collaboration is that synchronous collaboration occurs in a tight real (or near-real) time feedback loop, whereas asynchronous collaboration lacks that immediate interaction.
These key differences imply the need for different user interfaces and interaction styles, and pose interesting design challenges not only for the underlying architecture but also for the interfaces with which people interact.
One key advantage of small teams is the breadth of experience they bring to the search task. The nature of this experience may be expressed as roles. In some cases, each team member plays a similar role, whereas in other cases, the roles are different.

ROLES

There are a number of areas in which collaborative exploratory search of this nature is useful. Here are two examples.

- Domain expert/domain expert: two doctors from different domains need to collaborate on a complex diagnostic question, involving esoteric information from each domain. Each generates domain-specific search terms that the system integrates into joined queries. Results are allocated to each user based on domain expertise.
- Domain expert/search expert collaboration: the collaborators are a domain expert (aerospace engineer) and a search expert (librarian). The system allows the librarian to select sources and to formulate queries, while the engineer suggests terms and provides relevance feedback.

HUMAN-COMPUTER INTERACTION

A system that supports synchronous collaborative search must be able to collect data from two or more people and to use that data to synthesize queries and other operations on a search engine, and it must allocate search results to participants as appropriate to their roles.
Such a system must include interfaces for individual exploration, interfaces that support implicit awareness of others' activities, and interfaces for explicit collaboration. A middleware layer for coordinating the activities of the group, and an algorithmic engine optimized for collaborative exploratory search must exist to support this rich information sharing.
This system architecture can allow a small group of focused information seekers to search collections in concert. The system provides feedback based not only on an individual's search behavior, but also on the current, active search behavior of one's search allies.
User interface design in such an environment must draw on findings from CSCW, HCI, and IR fields.

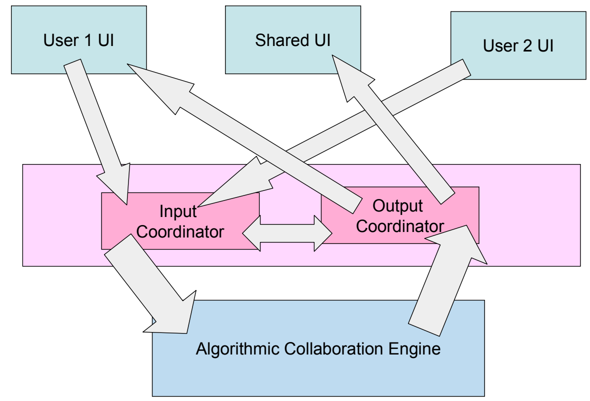

Figure 1. System architecture

EVALUATION

It is desirable to evaluate collaborative information seeking in traditional IR terms, but care must be taken when comparing search results of systems with different numbers of users working synchronously. One possible approach is to produce a combined ranked list based on contributions of all searchers, and then compute mean average precision (MAP) scores that are adjusted to account for the number of people contributing to the results.
One possible way to adjust MAP scores is to penalize highly-ranked non-relevant documents based on the number of searchers. The logic is based on the assumption that identical results produced by a smaller team are better. Thus the more people fail to make a correct judgment about a document, the worse that team performs.

$$modifiedprecision = \frac{|relevant \wedge retrieved|}{|relevant \wedge retrieved| + \left(|\overline{relevant} \wedge retrieved| * f(S)\right)}$$

Where f(S) is the penality function, such that f(1) = 1 and f(S>1) > 1. While a number of such functions are possible, we suggest that f(S)=S is a good place to start. The exact choice of function will depend on a variety of factors, and we need much more experience with these functions to determine which are appropriate when.

CONCLUSIONS

REFERENCES

1. Chi, E.H., Pirolli, P., Chen, K. and Pitkow, J, Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (Seattle, WA, United States, 2001), ACM Press, 490-497.
2. Morris, M.R., Interfaces for Collaborative Exploratory Web Search: Motivations and Directions for Multi-User Designs. In Proceedings of the CHI 2007 Workshop on Exploratory Search and HCI., (2007).