ESI 2007

Proceedings of the

ACM SIGCHI 2007 Workshop on "Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction"

http://research.microsoft.com/~ryenw/esi

Organizers:

Ryen W. White, Microsoft Research
Steven M. Drucker, Microsoft Live Labs
Gary Marchionini, University of North Carolina
Marti Hearst, University of California at Berkeley
m.c. schraefel, University of Southampton

In conjunction with:

Preface from the Workshop Co-Chairs

Welcome to the ESI 2007 Workshop! Exploratory search represents a shift from the analytic approach of query-document matching toward direct guidance at all stages of the information-seeking process. It can be seen as a specialization of information exploration – a broader class of activities where new information is sought in a defined conceptual area. People engaged in an exploratory search require support that extends beyond the known-item retrieval well-handled by many modern search systems. They may require help discovering new associations and kinds of knowledge, resolving complex information problems, or developing an understanding of terminology and information space structure. The creation of exploratory search interfaces (ESI) to support user-system interactions during search exploration has become a topic of great interest to the research community. This workshop provided an interdisciplinary forum for researchers, practitioners, and academics, interested in designing and evaluating technology for supporting exploratory search activities. Participant interests spanned areas such as interactive design, human-computer interaction, computer science, sociology, anthropology, ethnography, and information science. The workshop facilitated discussions aimed at developing of a set of guidelines for designing and evaluating exploratory search interfaces, and was a vehicle for introducing exploratory search to the SIGCHI community. The goal of this workshop was to find common elements for understanding the next generation of search interfaces to support users' exploratory search interactions. Questions we sought to answer fell under two headings:

- **Interaction**
  - What are the important features of an interface to support exploratory search?
  - What about interfaces to support exploratory search activities beyond the desktop?
  - What skills in the CHI community can we leverage to help develop these interfaces?
  - What are the challenges surrounding searchers' adoption of new search interfaces?
- **Evaluation**
  - How is evaluating ESIs different from evaluating other interactive applications?
  - What methodologies are most appropriate to evaluate these interfaces?
  - How do we evaluate rates of learning, discovery, accretion, and the quality of planning, analysis, and decision-making, within the context of exploratory search?

Through paper and panel presenations, break-out sessions, and discussions, we identified the issues pertinent to the evaluation of ESIs in a way that benefited participants in their own endeavours. We are grateful to the following people for taking the time to serve on the program committee, and review submissions for the workshop:

Bill Kules, Catholic University of America, USA
Bonnie Nardi, University of California at Irvine, USA
David Huynh, MIT, USA
Deborah Barreau, University of North Carolina at Chapel Hill, USA
Gene Golovchinsky, FX PAL, USA
Ian Ruthven, University of Strathclyde, UK
Jim Jansen, Pennsylvania State University, USA
Matt Conway, Expedia, USA
Merrie Morris, Microsoft Research, USA
Mira Dontcheva, University of Washington, USA
Sherry Koshman, University of Pittsburgh, USA

We thank you for participating in the workshop and we hope you enjoyed being part of the event!

Ryen W. White, Steven M. Drucker, Gary Marchionini, Marti Hearst, and m.c. schraefel
ESI Workshop Co-Chairs, April 2007

Table of Contents

Presented papers

Poster papers

Background papers

[Continued over]

Information Seeking and Sensemaking for 'Personal Fit'

Ann Abraham, Marian Petre, Helen Sharp
Centre for Research in Computing, Department of Computing
Open University, Walton Hall, Milton Keynes, England MK7 6AA
Contact author: a.m.abraham@open.ac.uk

ABSTRACT

This preliminary study investigates how participants make sense of search results, web-pages and documents when they use web-based environments. From a sensemaking perspective, it bridges across previous information seeking studies, which focused on specific aspects of information interaction such as search or information relevance judgments. This study corroborates some of those previous observations and extends them, by distinguishing different phases within participant's evaluation of information sources and identifying relationships between them. Five categories of Information Interactions emerged: the results are selectively summarised and discussed with reference to the literature and within the context of 'making sense'. Finally, the paper considers the implications for interaction design and exploratory search.

INTRODUCTION

The goals of this preliminary study were to gain insight into the range of strategies that individuals might employ in their search activities in order to select, interpret and integrate information, gathered from the Web in the context of their research-related tasks, i.e. to gain insight into how they 'make sense' of the information seeking process, and the implications for interaction design. As such, the study was interested in the interactions between the end-user and the information sources (e.g. search results and web-pages/documents), and looked across different aspects of the sensemaking process:

- deciding which sources to use;
- collecting, extracting and organising relevant information from the sources;
- forming external representations, both of the collection and of the meaning being derived from it.

One of the early Information Seeking Models is Kuhlthau's Information Seeking Process (ISP) model [10]. This captured the user's constructive approach to 'finding meaning through encountering information over a space of time'. Thus, Kuhlthau considered information seeking as a sensemaking process, i.e. where the individual is actively interpreting and constructing their own meaning to fit in with what they already know. Kuhlthau's influence is seen in more recent models, including Choo's Human Information Seeking: An Integrated Model [3] and Ingwersen's Cognitive Framework of (longitudinal) IS&R [8]. Several recent empirical studies are also relevant to this project, in as much as they have informed this study's aims, design etc., i.e. each addresses one or more aspects of the search process, although none actually span the range of strategies being investigated here.

Findings from this study corroborate some of the existing literature, but also provide further insight by distinguishing different phases within participant's evaluation of information sources and identifying relationships between them. To date, no studies have been found that focus on information gathering, i.e. search/selection, and external representation strategies and their relationship. Indeed this research appears to address a call to investigate how information seeking and sensemaking construction interact [12].

RESULTS FROM THE EMPIRICAL STUDY

Nine computing PhD students, who were all experienced information processors, were individually observed conducting single session web-based searches to locate information sources for their own authentic, research-related tasks. A pre-observation questionnaire was used to collect demographic data and some typical information behaviours, but the main data collection was via 'think-aloud' protocol, audio recorded during the 1.25hrs observations. Evidence of external representations (ERs) was collected post-session, to verify any claims.

Emergent Strategies

Several emergent categories of user' information seeking strategies were identified from inductive analysis of the transcribed audio recordings. Initially the emergent strategies were broadly categorized as '*Search*', '*Evaluation*', and '*Use*'.

	Categories of Information Interaction Strategies				
	Search	**Evaluation for Selection**	**Evaluation for Use**		**Use**
			Utility	**Personal Fit**	
Purpose:	*Finding/ locating* information	*Selecting* sources to look-at	*Deciding* what source/document to read (examine in depth) & store or print	*Digesting* the content, determining which content to extract, synthesise, use	*Organising* and incorporating into a body of knowledge
Characteristics:	interaction with the *query interface*, search engine, tools, etc.	interaction with the *results lists* at screen	interaction with the *information source* at screen	interaction with the *information source* predominantly using paper-based source	interaction with the *information source(s*) and with its (their) *representation(s)*

Table 1: Information Interaction Strategies in Web-based Information Seeking: Purpose and Characteristics

The complexity of the *Evaluation* strategies, as seen in Table 1 (above), emerged from recognising the three groupings as interaction strategies, each with its own purpose and characteristics. As a direct result of this perspective, *Evaluation* was further categorised as '*Evaluation for Selection'* and '*Evaluation for Use*'. Of particular significance here, is the further elaboration of '*Evaluation for Use'* into '*Evaluation for Utility*' and '*Evaluation for Personal Fit'* strategies. Table 1 (above) summarises the purpose and characteristics of these five 'Categories of Information Interaction Strategies'. The complexity can be explained as follows:

- '*Evaluation for Selection*', involves interactions with information objects (surrogates) in results lists. These were discernable from the interactions with actual sources themselves, i.e. during '*Evaluation for Use'*;
- '*Evaluation for Use'* was elaborated into '*Evaluation for Utility'* and '*Evaluation for Personal Fit*'. The transition from '*Utility*' to '*Personal Fit*' was further distinguishable by
  - a typical change of media i.e. from screen to paper;
  - a trend towards more intense, iterative interactions, e.g. in-depth source reading and multi-passes of sources, leading to identification of new, and clarification of already identified, conceptual extractions;
  - emergence of examples of participant's '*on-source'* external representations (*ER*s), e.g. highlighter and margin notes, to emphasise or mark for further attention etc., as well as '*off-source*' ERs, e.g. hand-written/electronic descriptive notes, to provide quick recall of main points etc..
- *The Use* category signalled a change of task focus: rather than searching for sources, or searching the located sources for conceptual and physical extractions, the focus was on 'writing' or preparing a 'task deliverable'. Other characteristics of this category include
  - the move to multi source interactions e.g. comparisons across sources;
  - emergence of more complex analytic ERs e.g. multi-dimensional tables, diagrams etc., developed using high order cognitive skills, such as categorisation, comparative analysis, etc.;
  - interactions with its (their) own ERs;
  - integration of new-found information into their own body of knowledge.

The boundary between '*Evaluation for Personal Fit*' and '*Use'* categories may need further investigation. For example, participants might generate ERs and integrate them in part ('*Use'* them), whilst they are reading in-depth ('*Evaluation for Personal Fit')*; similarly they might delay generating an ER until they are ready to '*Use'* the source.

Relationship across Categories

Table 1 (above) should not imply any linear progression across the groupings. In fact the participants were opportunistic in their interaction behaviours e.g. these varied depending on task, timing and other factors. However, many stated their preference to work in 'batch-mode' e.g. iterate around the *Search*, *Select* & '*Evaluate for Utility'* categories, until they had gathered a batch of sources for further in-depth reading, i.e. '*Evaluate for Personal Fit'* and subsequent '*Use'*, sometime later. The reasons for a further search could be a new reference, new keyword/concept etc., thus in theory a further search could be instigated at any point in the interactions.

Other relationships across categories included influences on evaluation judgments. One example is the authoritative nature of the source, seen as an influence from '*Selection'* through to '*Use*'. Task influences also proved to be influential across all aspects of the seeking process: it shaped their early evaluations, for example, in terms of the volume of sources they selected from the search results, and again later during their interactions with the actual source itself.

DISCUSSION

Whilst search and evaluation strategies are fundamental elements of Information Seeking & Searching (IS) Models, e.g. Ellis [5, 6, 7] etc., the *Use* aspect is often neglected although exceptions include [15] and [3]. The sub-categories '*Evaluation for Utility*' and '*Evaluation for Personal Fit*' provide further insight into information seeking and are not known to be in any IS models, although from initial inspection these map onto Ingwersen's Cognitive Framework for (longitudinal) IS&R [8]. This mapping needs further investigation, as does a comparison of these *Interaction* categories with Cool et al's 'Classification of Interactions with Information' scheme [4].

Sensemaking can be considered to be making sense of the information around us, finding the important structure in seemingly unstructured situations [12]. Considering the five interaction categories (Table 1 above), it can be seen that the transition from '*Utility*' through '*Personal Fit*' to '*Use*', evidences the sensemaking process, with each intermediary stage showing the participant to be progressively making sense of the information obtained. For example, the utility assessments involved tacit reference to internal (cognitive) frameworks: at this stage they seem to recognise that the found information could fill their knowledge gap but not 'how'. Whereas in '*Evaluation for Personal Fit*', conceptual extractions of the source content are identified, interpreted as their own, and emerge as ERs containing new explicit knowledge; participants appear to be concerned with 'best fit' for their knowledge gaps but at this stage they might not know how it all fits together. Finally at the *Use* stage, the prime driver is to integrate the new found information or its ER(s), with previous existing structures and develop a framework or structure to externally represent their problem solution. Sensemaking is clearly evident at this stage, e.g. from the emergence of more complex structures produced using higher order cognitive skills such as analysis, synthesis etc.. It is the sensemaking aspect of these Interaction Strategies that encourages comparison with Kuhlthau's [10] original model: this range of categories would seem to map into Kuhlthau's last two stages, namely 'gathering and presenting'.

The richer picture given by this study's results is not discussed here; however, the detailed findings have been compared with other recent studies which focused on particular facets of the information seeking task including [1], [9], [13], [14], [16], [17], [12]. The ongoing task influence reflects the research from [2] and [8].

Implications for Interaction Design and Exploratory Search

This study shows that there are discrete categories of information interactions, all needing their own interaction design solutions. '*Use*' might be considered too broad and complex to be addressed as a whole, but the discrete categories of '*Evaluation for Utility*' and '*Evaluation for Personal Fit*' offer scope for further investigation into interaction design implications. '*Personal Fit*, evaluation**,** where there is emergence of the transition from internal to external mappings of knowledge (tacit-explicit), more intense interactions with source, and deliberate change of media (screen to paper), would appear to have the most interesting implications. For example,

- understanding the needs of emerging external representations (ERs), which are explicit artifacts of the user's sensemaking, could help inform designers/ developers about the functionality and support needed for visualisation and organisation;
- understanding the relationship between '*Searching*' and '*Personal Fit*' strategies, could provide insight into what specifically activates further searches;
- understanding the nature and the format of these further searches, which can be broadly considered as examples of exploratory searches [11], could help inform current research in this area.

REFERENCES

1. Aula, A., Jhaveri, N. and Kaki, M. (2005). Information Search and Re-access Strategies of Experienced Web Users. In *Proceedings of the 14th International Conference on World Wide Web 2005*, ACM Press, 583-592.
2. Byström, K. and Järvelin, K. (1995). Task Complexity Affects Information Seeking and Use. *Information Processing and Management 31* (2), 191-213.
3. Choo, C.W. Closing The Cognitive Gaps: How People Process Information. (1999). *Financial Times*, March 1999, Reprinted August, 2001, in National Post of Canada. http://128.100.159.139/FIS/respub/FThis/default.html [last accessed 08/2006].
4. Cool, C. and Belkin, N. (2002). A Classification of Interactions with Information. In Bruce, H., Fidel, R., Ingwersen, P. and Vakkari, P. (Eds.) *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (COLIS4) 2002*, Libraries Unlimited, 1-15.
5. Ellis, D. (1989). A Behavioural Model for Information Retrieval System Design. *Journal of Documentation 45*(3), 237-247.

6. Ellis, D., Cox, D. and Hall, K. (1993). A comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences. *Journal of Documentation 49* (4), 356-369.
7. Ellis, D. and Haugan, M. (1997). Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. *Journal of Documentation 53* (4), 384-403.
8. Ingwersen, P. and Järvelin, K. (2005). *The Turn: integration of information seeking and retrieval in context.* Springer/Kluwer, Berlin.
9. Ivory, M., Yu, S. and Gronemyer, K. (2004). Search Result Exploration: A Preliminary Study of Blind and Sighted Users' Decision Making and Performance. *In Proceedings CHI 2004*, ACM Press, 1453-1456.
10. Kuhlthau, C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of American Society for Information Science 42*(5), 361-371.
11. Marchionini, G. (2006). Exploratory Search: From Finding to Understanding. *Communications of The ACM 49* (4), 41-46.
12. Qu, Y. and Furnas, G. (2005). Sources of Structure in Sensemaking. *In Proceedings CHI 2005,* ACM Press, 1989-1992.
13. Rieh, S.Y. (2002). Judgment of Information Quality and Cognitive Authority in the Web. *Journal of the American Society for Information Science and Technology 53* (2), 145-161.
14. Tombros, A., Ruthven, I. and Jose, J.M. (2005). How Users Assess Web Pages for Information-Seeking. *Journal of the American Society for Information Science and Technology 56* (4), 327-344.
15. Vakkari, P. (2001). Changes in Search Tactics and Relevance Judgements when Preparing a Research Proposal: A Summary of the Findings of a Longitudinal Study. *Information Retrieval 4* (3/4), 295-310.
16. Wang, P. and Soergel, D. (1998). A Cognitive Model of Document Use during a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science and Technology 49*(2), 115-133.
17. Wang, P. and White, M.D. (1999). A Cognitive Model of Document Use during a Research Project. Study II. Decisions at the Reading and Citing Stages. *Journal of the American Society for Information Science and Technology 50* (2), 98-114.

Social bookmarking and exploratory search

David R Millen
IBM T.J Watson Research
One Rogers Street
Cambridge, MA 02142
david_r_millen@us.ibm.com

Steve Whittaker
Department of Information Studies
University of Sheffield
Sheffield S1 4DP, U.K.
s.whittaker@shef.ac.uk

Meng Yang
IBM Software Group
One Rogers Street
Cambridge, MA 02142
yangmeng@us.ibm.com

ABSTRACT
In this paper, we explore various aspects of exploratory search supported by a social bookmarking service. These bookmarking services hold great potential to powerfully combine personal tagging of information sources with interactive browsing, resulting in better social navigation. We examine use when deployed in a large organization, through logfile analysis and end-user interviews, to understand how social search is supported. We conclude that social search is supported well through a combination of better personal bookmark management, enhanced social navigation and application-specific search.

Keywords
social software, social bookmarking, tags, social navigation

1. Social Bookmarking Systems
In recent years, there has been tremendous growth in shared bookmarking applications. Introduced in 2003, the del.icio.us [3] social bookmark manager was one of the first of this kind of application, and has enjoyed an early and large base of committed users. A flurry of similar offerings has since been unveiled [see [4] for a recent review].

These internet oriented social bookmarking services have been adapted for use in large organizations. Examples include the *dogear* [7] and *onomi* [2] social bookmarking services. Both of these enterprise-ready bookmarking services support bookmarking of both internet and intranet information sources, and provide user authentication via corporate directories

These social bookmarking systems share a number of features. First, they allow individuals to create personal collections of bookmarks. These centrally stored bookmark collections bring immediate personal benefit by providing a collection that can be browsed from any web-accessible machine.

A second and significant enhancement in these systems is the use of keywords or tags that are explicitly entered by the user for each bookmark. These tags allow the individual user to organize and display their collection with labels that are meaningful to them. Furthermore, multiple tags allow bookmarks to belong to more than one category, avoiding one of the limitations of the hierarchically organized folders of bookmarks (or "favorites") found in most web browsers.

The final distinguishing characteristic of these social bookmark applications is the social nature of their use. There is a bias towards increased transparency in these tools. While bookmark collections are personally created and maintained, they are typically also visible to others. A number of user interface elements allow social browsing of the bookmark space. For example, user names are "clickable" links, and, when a name is clicked, the bookmark collection for that user is presented. This allows someone to get a sense of the topics of interest for a person. Tags are also clickable, and when selected will result in a list of all bookmarks that share that tag. This is a useful way to browse through the entire bookmark collection to see other information sources of interest. We call this ability to reorient the view by clicking on tags or user names, "pivot browsing"; it provides a lightweight mechanism to navigate the aggregated bookmark collection.

These new social bookmarking applications are a natural and powerful extension of existing social navigation tools and practices. They provide a mix of both direct (intentional) navigational advice as well as indirect (inferred) advice based on collective public behavior. They also provide support in various ways for different kinds of exploratory search.

In this paper, we present the results of a field study of the *dogear* social bookmarking service that has been deployed in a large enterprise [7]. We are generally interested in the different ways that this social bookmarking service supports exploratory search. In particular, we are interested in understanding how social bookmarking tools are used to find, refind and explore new information resources.

2. Dogear Field Study Results
To answer our research questions, we performed a field study of an enterprise bookmarking service deployed in a large multi-national company. Our general understanding of the use of the enterprise bookmarking service was based on many sources of user data, including log files, the primary bookmark data files, and in-depth interviews with end-users of the service. Included in the log files are user actions (e.g., create, delete, edit a bookmark, bookmark "clicks"), user and bookmark owner identifiers, and a time and date stamp.

The system was introduced in mid-2005 and contains over 100,000 bookmarks and 250,000 tags created by over 1600 active end-users The activity analysis presented here is based on log files covering a 12 month period from July, 2005 to July, 2006 (~450K records). The dogear research team records were omitted for these analyses.

2.1 Supporting various search activities
In this paper we are particularly interested in understanding how the bookmarking service supports various kind of exploratory search activities We examined the logfiles for a one month sample (June 06) to get a sense of how often end-users navigated through the bookmark collection using tags, names, and explicit search features. To better understand the results of the end-user actions,

we have considered four different kinds of search, which are described in Table 1.

Table 1. Different kinds of search support by dogear.

General browsing	Examining bookmarks by time (recent bookmarks,), by frequency (most popular sites bookmarked), or URL details (all bookmarks for a particular info resource)
Social search	Exploring the bookmark collection through social tags, people links, or some combination of the two. Subscriptions are included in this group as a form of a saved social search/query.
Personal search	Looking for bookmarks from one's own personal collection of bookmarks.
Explicit search	Using a traditional search box to enter a set of search terms

The results of our logfile analysis are presented in Table 2. We group the results by the different kinds of search that we described above. We show the number of times that a particular navigation path resulted in a click-through to the original information resource that was bookmarked. Finally, we present the percentage of navigation events that resulted in a click-through..

Table 2. Browsing frequency by tag and person combinations.

Action	#	# clicks	% clicked
General browsing			
View-recent	9708	1838	19%
View-popular	1287	450	35%
View all for a-URL	919	181	20%
Subtotal	11914	2469	21%
Social navigation			
Person	2595	691	27%
Tag	1899	432	23%
View Subscription	781	192	25%
Person x tag	766	205	27%
Tag x tag	310	67	22%
Person x tag x tag	32	6	19%
Subtotal	6383	1593	25%
Personal search			
mybookmarks	7280	1517	21%
Self x tag	2461	990	**40%**
Self x tag x tag	184	64	35%
Subtotal	9925	2571	26%
Explicit search			
Search- Internal dogear	4157	1769	**43%**
Search–plug-in enterprise	9841	4068	**41%**
Subtotal	13998	5837	**42%**

The click-through percentage might be considered a navigation "completion" rate, and may be suggestive of different kinds of search strategies. For example, larger click-through percentages might indicate a more goal-oriented search. At minimum, the click-though percentage provides an end-user estimate of the usefulness of the search results that were presented via each of the various search pathways

2.2 General - Browsing

As can be seen in Table 2, the single most common way to view the dogear bookmark collection is to look at a listing of the most recently added bookmarks [View recent – 9708 above]. This is not very surprising as there is a main dogear menu link labeled "all and it is recommended in the application help to add a browser toolbar link to this dogear page.

Several dogear users that we interviewed indicated that they scan this list of newly created bookmarks with some regularity, much like they would scan news of blog feeds, exploiting the fact that dogear users form a shared community of interest within IBM. For example, one informant said:

> The first thing I normally go to in Dogear is All because I'm interested in finding out what everyone is doing. And I equate Dogear to basically Slashdot for IBM plus everything else on the Internet. And, it used to be -- I find that now --- I think it's better than Slashdot. Before when Dogear was starting, Slashdot was probably I would consider it a little superior because there was still a lot of people, more and more people out there finding a lot of interest in feeds. And now that Dogear has achieved critical mass, there's a lot of people within IBM who I know and I respect, they've been picking up lots of interesting things that Slashdot folks haven't been picking up on.

This same informant indicated that he had changed his scanning the "all" list over time as the volume and variability of the bookmark content changed.

> I would scan through probably maybe three or four pages a day if I have time. When it was -- when I used to -- when I started with Dogear, I would actually scan maybe more, but the volume was less. And it was a lot easier to pick -- sift through things. But now that there's more bookmarks today, and there's a lot of people from the other parts organization that I don't necessarily care about, I reduced the amount of pages I look at.

There were fewer reports from informants about the use if the "popular" or most active bookmarks, although one person indicated that he was aware of the top three bookmarked sites. There appear to be a relatively large percentage of click-throughs on the popular bookmarks, indicating significant interest in what other people have already found. The click-through data also suggests that, compared with recency, popularity is an effective recommendation about which resources might be worth reading.

2.3 Social Search

Of greater interest, perhaps, we note that there is also considerable interest in browsing other peoples' bookmark collections. As can be seen in Table 2, the most frequent way to browse another user's bookmarks is by clicking on that person's name. The next most popular way to browse other users' bookmarks is by selecting a specific tag from the system-wide tag cloud. It is considerably less common for a user to select a tag from another user's tag cloud and there is almost no use to date of the more advanced browsing of tag intersections (in Table 2, Tag x tag, Person x tag x tag, and Self x tag x tag).

Once again, we also examined the number of times that end-users clicked through on a URL that had been bookmarked by another. . The click-through results in Table 2, show that browsing someone else's bookmarks often results in following the URL to the original information source (19-27 percent of the time). This indicates the utility of social navigation in identifying useful informational resources.

These results suggest widespread curiosity about what others are bookmarking and provide evidence of the kind of explicit social navigation that is taking place within the bookmarking service. These results are significant as they represent a novel form of information browsing within the enterprise. Comments from end-user interviews confirm that the bookmarking service is supporting social navigation. One respondent indicated that she browsed the links of others for different reasons. In one case, it was to infer topical interest for an organization thought leader:

> It's usually I'm just looking at the people I know, or like, you know, thought leaders and trying to get at, you know, what is their thinking… And that might come up just through their tags or just through their content, or a combination.

Another respondent indicated that navigating tags was a good way to informally find people with interest or expertise on various topics.

> And it just seems like a fabulous way of either finding people who either might know about something or might have bookmarked something that I would be interested in, and that's when I would use the tags because it would be about subject matter. So I would be more likely to use tags then

This same respondent indicated that another significant benefit is being able to trust the information sources based on knowledge of the people in the group.

> …there would be words that I would be thinking in my head, like key words. And if I saw them in a tag cloud, then I would click on them and … all these things would show up that have potential of being what I'm looking for. And I would do that as opposed to going to Google because in some way, it's a somewhat trusted community already. And I – mentally I'm saying I respect the people in this community and they probably know things that I would be interested in. And I would trust their sources.

And yet another respondent indicated that tag browsing was also an efficient way to keep up-to-date on a particular subject.

> And sometimes when I'm searching by topic, like if I'm working on – if I have a thought or an idea, or I'm writing something, let's say about attention management, I might want to see what's new on attention, and I'll just search that tag to see what's new, rather than searching my subscriptions or on social networking or task management.

Another informant described different using social searching to do learn about new topics.

> So, an example is, somebody was telling me, somebody that I work with, was telling me that they have suddenly gotten interested in Second Life. Second Life is a virtual reality environment, and I knew that there was a bunch of people in IBM who were tracking it. So, I thought, well let me see -- let me look and see what other people have found about Second Life because that is something that people who use Dogear are likely to have found.

The informant described looking through other bookmarks tagged with "second life" until she can " get an idea of what is this thing? What is this category?" She continues to search until she finds a reasonable description.

> Now, in this particular case, you know, so here is something that is, you know, a reasonable description. It gives you some idea and, oh, look, it's Business Week. Well, I can click on that because that is going to tell me in plain English what Second Life is.
>
> So, there is some serendipity that occurs within that, but it is really because I am looking for the beginnings of information about something. I am not looking for deep research.

2.4 Personal Search

As can be seen in Table 2, a popular way to browse the bookmark space is to look at one's own collection (labeled "mybookmarks." This is not particularly surprising as there is a top level menu button labeled "My Bookmarks" that will take the user to their own bookmark collection. There is also, however, considerable browsing of one's own collection navigating using personal tags (Self x tag). Applying tags may serve as a useful way to filter bookmarks once users build up larger collections of bookmarks.

In Table 2, there is also evidence that end-users actually use the bookmarking service to "revisit" information sources that they have previously bookmarked. In fact, when end-users navigate their bookmarks using their personal tags, they click through 40% of the time. This finding is contrary to early research suggesting that classic single-user web bookmarking is ineffective as users seldom turn to bookmarks they have created [1, 5].

Several of the end-users that we interviewed indicated that this support of personal bookmark management was an important characteristic of the service. One end-user highlighted the ability to centrally store bookmarks:

> And I find a lot of personal value, or at least personal comfort, in realizing that I saved this stuff somewhere, and it's refreshable, and it's in that location.

Another end-user said that social tags were very helpful in managing the overload of blog content:

> Because I have a lot of blogs that I read and I have found that this (bookmarking service) is just a better way, a quicker way for me to organize them. I will click the word blog, the tag, and then I'll see all the blogs.

The specific need to re-find information was highlighted by another informant

> I bookmark anything that I think I am going to want to go back to. I think that's the general reason for bookmarks, and often I will, as I am browsing stuff, not bookmark it because I say, oh, I will be able to find it, and then I realize that I am looking for it one day, and I can't find it, so then I will go and bookmark it.

2.5 Explicit search of dogear bookmarks

As described above, there is a general search capability provided within the dogear service. As can be seen in Table 2, the internal search capability is the third most used way to "explore" the dogear collection (4157 times in the log sample).

There is also an ability to search the *dogear* bookmarks as part of a more general enterprise search. The number of search events (Table 2) shows a very high usage rate of the external search plug-in (9841). Of particular note, is the high click-through rate (41% and 43%) for search actions, which indicates a more purposeful searching of the collection.

Several dogear users commented on the value of the dogear search that was integrated into the more general search tools (i.e. Google and enterprise search). One benefit was simply being reminded (inline) that related content has been bookmarked on dogear.

> Another thing that happens is that even though the Google results are more specific generally, because it's more of a general search than dog ear, I may, depending who the person was that created the bookmark, I may tend to go look at it

Another informant indicated that she frequently did enterprise searches and then decided to redo the search in dogear.

> Because a lot of times, I mean especially if it's an internal [enterprise] search, …, you search for something and you're like, "Man, I can't find it." And then you just jump over to Dogear and you get it.

Other dogear end-users have indicated that an important reason for preferring the dogear results over generic search results was the fact that very inclusion in the dogear collection meant that the search had been actively prefiltered by a community of interest or trusted IBM colleagues.

3. Discussion

In this paper we have investigated the navigational elements of a social bookmarking application that have been designed to support exploratory search. These design elements afford navigation through the bookmark collection by tags, by people, or by combinations of tags and people, and through multiple search interfaces (i.e. both internal dogear search and through the Firefox plug-in). The results of our log file analysis confirm that these navigational elements are used by end-users of the social bookmarking service, while interview comments provide support that both personal and social navigation benefits are being realized.

The results of this study suggest that social bookmarking services provide significant benefits for managing personal bookmarks. There is substantial browsing of personal bookmarks, which often result in revisits to the original content source, suggesting the superiority of our system over more traditional bookmarking methods. The interview comments suggest that much of the use of personal bookmarks is to refind information, which would support various *look-up* search activities, as described by Marchionini [6].

The results also show a fair amount of social navigation (i.e. users' in looking at another person's bookmark collection). This kind of social search may be due, in large part, to one of the unique characteristics of enterprise social bookmarking services. End-users of these kinds of services [2, 7] are required to use corporate identities (i.e. no pseudonyms allowed). This enables others in the enterprise to easily recognize a coworker in a list of people tags (see Fig 2(b)) or as a creator of a specific bookmark. Once recognized, it is easy to learn from organizational thought leaders or to learn about the interest of new team members. Furthermore, coworkers provide an important selection and filtering function for information content.

The interview comments indicated that a large portion of this social navigation is in support of what Marchionini would call *learning* search activities. For example, individuals mentioned following the name links in dogear to learn more about someone and to find out what others are interested in. Interviews also revealed cases in whish dogear users followed a tag (or multiple tags) to learn about a particular technology. This exploration of other people's bookmarks is a very promising way to support lightweight information sharing within an organization.

The general browsing activities, which included casual monitoring and clicking of *recent* and *popular* bookmarks, are also evidence of *learning* search activities. The explicit dogear search appears to be used for both *look-up* queries and to *learn* about various topics.

There are a number of implications for the future design of *dogear* (or similar bookmarking services). It seems that although people prefer to browse by recency (i.e. view recent), they actually find more useful information using popularity (as indicated by click-through results).). So the interface should be redesigned to respect this. Similarly, for refinding personal resources the "mybookmarks" user interface could better support tag searching by providing a list of recently search or most used tags for each individual.

4. REFERENCES

[1] Abrams, D., Baecker, R., and Chignell, M. 1998. Information archiving with bookmarks: personal Web space construction and organization. In *Proceedings of SIGCHI* (Los Angeles, California, April 18 - 23, 1998).

[2] Damianos, L, Griffith, J, Cuomo, D, Hirst, D. Smallwood, J. (2006) Onomi: Social Bookmarking on a Corporate Intranet. WWW 2006 Tagging Workshop, Edinburgh, May 22.

[3] del.icio.us; http://del.icio.us/

[4] Hammond, T., Hannay, T., Lund, B., Scott, J. Social Bookmarking Tools (I), A General Review in D-Lib Magazine, Volume II, Number 4, April 2005.

[5] Jones, W., Bruce, H. and Dumais, S. (2003). How Do People Get Back to Information on the Web? How Can They Do It Better? In *Proceedings of INTERACT*, Zurich, Switzerland, September 1-5, 2003.

[6] Marchionini, G. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (Apr. 2006), 41-46.

[7] Millen, D. R., Feinberg, J., and Kerr, B. (2006). Dogear: Social bookmarking in the enterprise. In *Proceedings of SIGCHI* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM Press, New York, NY, 111-120

Interfaces for Collaborative Exploratory Web Search: Motivations and Directions for Multi-User Designs

Meredith Ringel Morris
Microsoft Research
One Microsoft Way, Redmond, WA, USA
merrie@microsoft.com

ABSTRACT
In this position paper, we examine exploratory Web search as a collaborative activity and propose that such collaborations are commonplace. We present survey results that support this claim, and argue that current search interfaces provide limited support for common collaboration needs. We identify four features of the exploratory search experience (*coverage*, *confidence*, *exposure*, and *productivity*) that could be enhanced by providing explicit support for collaboration during the search and subsequent sensemaking processes.

Author Keywords
Exploratory search, computer-supported cooperative work.

ACM Classification Keywords
H5.3. Group and Organization Interfaces: Computer-supported cooperative work.

INTRODUCTION

Currently, searching the Web is considered a solitary task. The most commonly used tools for Web search, sites such as Google, Yahoo!, and Windows Live Search, are designed for single-user scenarios. However, joint and sequential activities by people collaborating on information-gathering tasks are commonplace. To date, there has been little work on providing means of enhancing collaborative exploratory search activities. Tools supporting collaboration during Web search promise to be valuable to users.

We provide motivation for focusing attention on multi-user information-gathering scenarios, review related work on this topic, and report on our own surveys regarding collaborative search experiences. Then, we discuss key features of the exploratory search experience that could be enhanced by providing explicit support for collaboration.

Motivation

Though not the target use scenario of current Web search engines, collaboration with friends, family, or colleagues during a search task is surprisingly commonplace. School-age children, for example, frequently work together at a single computer for the completion of group projects (both as a consequence of the limited supply of computing resources and to experience the pedagogical benefits of team work). Large *et al.*'s fieldwork with elementary-school students [4] found that, "Much class work, including project assignments, is undertaken by school students in groups rather than individually. Collaborative information-seeking, then, is not unusual."

Ethnographic work in higher education has also identified collaborative information-seeking practices. Twidale *et al.* [9] observed students using databases in a university library, and identified two types of collaborative information-seeking behaviors, neither of which was explicitly supported by the library's search application. The first, "joint search," involved groups of two to four students clustering around a single computer, frequently pointing to the screen while discussing ideas and planning actions. The second, "coordinated search," involved the group sitting with each member at a separate, but adjacent, workstation. The students would discuss the task, compare results, compete to find information quickly, and lean over to look at their teammates' screens.

Our own investigations, discussed in the following sections, provide evidence that collaborative searching is a strategy that is also employed in non-educational contexts, such as information-oriented workplaces, as well as in more casual settings. For example, colleagues in a research lab who are working together to write a scholarly article would benefit from a collaborative interface to facilitate a joint literature search. A group of friends planning a vacation to Hawaii might benefit from an interface that allows them to combine their efforts in identifying cheap airfares, appropriate hotels, and interesting tourist activities.

Related Work

Most research in the area of collaborative search focuses on passive forms of collaboration (*i.e.,* using data generated by large numbers of users' interactions with a system to impact current system behavior). Examples of such work include using query logs and clickthrough data to generate query substitutions or recommendations [1, 2]. Implicit data such as clickthrough actions of a large number of users can also be used as a collaborative filtering mechanism to re-rank search results [7].

There have been a few research efforts toward providing more active collaborative experiences among groups of users who know each other. However, these systems tend to be designed for very specialized domains and/or devices. TeamSearch [6] is a system that enables co-located groups of up to four people to simultaneously search collections of digital photographs, using a visual query language designed for a multi-user interactive tabletop. Maekawa *et al.* [5] describe their system for groups of co-present people who each have a small, Web-enabled mobile device – to improve the efficiency of searching for information within a Web page (since scrolling through long Web pages on small screens is time-consuming), they allow a page to be split into several parts, each of which is displayed on a different user's device to facilitate parallelization of visual search. Krishnappa [3] developed a system that supports synchronous, remote collaboration between two people searching a medical database. Users of his system perform standard, single-user searches, but have a built-in textual chat facility as well as the ability to press a "share" button that sends some metadata about what they have found to the other user.

A few commercial products also offer support for collaboration during search tasks. For example, the Cha Cha search engine[1] pairs searchers up with a live person who assists them in formulating their query and suggesting interesting websites; however, this is not intended as an interface for allowing people to collaborate on an exploratory search task, but rather an attempt to support novice searchers by providing the ability to interpret natural language queries via human cycles. The Live Messenger[2] IM client provides a "shared search" feature whereby conducting a Web search through the client allows the list of returned URLs to be displayed to both the searcher and his/her IM partner. Google Notebook[3] allows a user to store clippings from several Web sites in one document; the tool provides a facility for allowing multiple users to add content to a single notebook document. Both Live Messenger's "shared search" feature and Google's notebook software are important steps toward facilitating collaborative exploratory search; however, both technologies focus on collaboration during the sensemaking [9] portions of an exploratory search task (*i.e.,* the viewing, selecting, and organizing of search results). Tools that allow users to collaborate in the formulation and refinement of queries during an exploratory search in addition to supporting collaborative sensemaking could offer many benefits, which we discuss in the following sections.

INITIAL EXPERIMENT: THE "AUNT EDNA" TASK

We conducted an online survey of ten knowledge workers in our research lab, in which they were asked to perform a specific exploratory search task. The task they were given was:

> *Your elderly Aunt Edna was recently diagnosed with high blood pressure. She's asked you to send her a link to a good source of information on treatments for her condition. Using your favorite search engine, find a single website that you would recommend to your Aunt Edna.*

Participants were asked to keep track of and submit the following information: the URL of the site they chose to recommend, the query terms they used to discover the chosen site, the search engine used to discover the chosen site, and their level of confidence (on a 5-point scale) that the site they chose was the best one for the task.

Note that this was not a collaborative task. We began our investigation with a single-user task in order to better understand how the single-user experience might be improved through the addition of collaborative support.

Results and Discussion

A total of seven distinct Web pages were "recommended" to Aunt Edna by the ten participants. Three sites were recommended by two participants each and the other four were each recommended by one. Eight different query formulations were employed to discover these sites (with two query formulations each being used by two users, and the other six by one each). Two different search engines were used (one by 40% of users, the other by 60%).

This variation in strategies (search engines and keywords used) and results (final URLs recommended) is quite interesting, especially considering that the participants in the exercise were all skilled searchers (all work in technical fields and use search technologies on a daily basis). In particular, this variation demonstrates potential benefits of collaboration for such tasks – the "Aunt Edna" task suggests that the breadth of tools used, query terms generated, and sites considered may increase as multiple users tackle the same search task.

Users' level of confidence in the quality of their chosen website also suggests room for improvement of the exploratory search experience. On a 5-point scale (with 1 = "Very Unconfident" and 5 = "Very Confident"), the mean rating of confidence that the recommended site was the best available was 3.0 (s.d. = 1.05). Notably, no participants selected the "Very Confident" option. Again, this points out a potential benefit of facilitating collaborative exploratory search – there is significant room for improvement in increasing users' confidence that they have succeeded at a search task; allowing them to combine and verify their efforts with others may be one means of accomplishing this goal.

Follow-Up Task

The next day, we sent a follow-up task to the same ten participants. This task, which was completed by six of the ten participants, asked them to visit each of the seven websites collected during the first task, and to answer two questions about each. The first asked them to rate (on a 5-point scale) the quality of the suggested page compared to the quality of the page they had originally recommended. The second asked them whether they had encountered this suggested page during their initial search task.

---

[1] http://www.chacha.com

[2] http://get.live.com/messenger/

[3] http://www.google.com/googlenotebook/overview.html

For each of the seven sites, at least three of the six respondents reported that they had not encountered that site during the initial search. For four of the sites, at least five of the six respondents reported not encountering it previously. In the previous section, we postulated that the large number of sites suggested in response to the initial task indicated that involving more users in a search task could increase the number of websites considered during an exploratory search. This follow-up data adds additional confidence to that hypothesis by demonstrating that users hadn't even encountered many of these sites during their search (rather than encountering them but choosing not to recommend them).

All but one site had at least one person (not including the person(s) who initially recommended that site) who felt the site was at least of equal quality to the site they had recommended. Three of the sites had at least one person who felt that the site was even better than the site he/she had recommended during the previous task. Again, these results point to the potential for collaboration to assist users in discovering high-quality information on the Web that they may not encounter on their own, and perhaps thereby increasing users' confidence that they have successfully met an information need through exploratory search.

SURVEY ON WEB SEARCH HABITS

While the results of the "Aunt Edna" task suggest potential benefits in allowing users to combine their search efforts, the task itself did not involve nor ask directly about collaboration. To learn more about current and desired practices regarding collaboration during search, we conducted a survey of employees at Microsoft. The survey was sent to 740 people, and was completed by 204 (giving a response rate of 27.6%). Entry into a prize drawing served as incentive for completion of the questionnaire.

Respondents represented a variety of knowledge workers, including researchers, software developers, managers, and administrative assistants. The population surveyed had a relatively high level of expertise in Web search – 99.5% report using a Web search engine at least once per day, and no respondents self-identified as "novice" users of search technologies.

53.4% of respondents answered "Yes" to the question "Have you ever cooperated with other people to search the web?". This is a surprisingly high number, considering that current technology is not explicitly designed to support cooperative searching. Of the people who said they had cooperated on a search task, 26.6% reported engaging in cooperative search activities at least once a week, and an additional 48.6% reported such activities occurring at least once a month.

Additionally, when we described specific collaborative behaviors, even some respondents who initially said they had never cooperated to search the Web reported engaging in a variety of "workarounds" that in fact enabled cooperative searching. For example, 87.7% of all respondents said they had watched over someone's shoulder as he/she searched the Web and suggested alternate query formulations to that person. 30.4% of respondents reported using instant messaging applications to coordinate real-time information seeking with a remote partner. 18.1% reported having divided up the responsibilities for sub-portions of a search task among several people, and then shared the results afterward.

18.1% reported needing/wanting to cooperate with others to search the Web, and being unable to find an effective workaround. Respondents provided some specific examples of search tasks where they had wanted the ability to cooperate with others. For example:

- "Ever try buying an airplane ticket at the cheapest possible price with someone else? Yikes."
- "Helping less computer-savvy users search the web (*e.g.,* my parents)."
- "Vacation planning: looking at hotels, dates, etc. … Goal: agreeing on vacation details."
- "We were trying to do a lit search but we both have different strategies for how we traverse the space. It was difficult to do together (because we wanted to follow different paths) … however, when we did it separately we weren't sure how much redundant information we were gathering."

DISCUSSION

Our survey on Web search habits revealed that a large number of people already engage in collaborative Web searching activities, even though these activities are not well-supported by current systems. The fact that people employ a variety of workarounds in order to collaborate suggests that collaboration support should be a first-class design requirement in Web search systems. It is likely that the frequency and pervasiveness of cooperative searching would increase if appropriate tools were available to end-users.

As indicated by the results to our "Aunt Edna" task, users employ different strategies to locate information (*e.g.,* different search engines, different query formulations, etc.), and subsequently encounter different, but still highly-relevant, websites during the search process. Creating user interfaces that support collaboration during exploratory search has the potential to improve search experiences and outcomes in several ways:

(1) better *coverage* of the space of relevant, high-quality sites.

(2) higher user *confidence* in the completeness and/or correctness of the search.

(3) *exposure* to varying search strategies and query syntax

(4) increased *productivity* due to a decrease in redundant information-seeking

Coverage

Facilitating collaboration during exploratory search could increase the number of relevant results discovered via several means. First, the simple increase in the number of pairs of eyes exploring the data could decrease the likelihood that a useful result is overlooked, and the increase in total person-hours devoted to the task is likely to increase with the number of collaborators as well. Additionally, the increased variety of search strategies (search engines, keywords, etc.) used is likely to increase the breadth of information encountered, giving another opportunity for additional relevant results to be discovered.

Confidence

The increased coverage of the search space in and of itself may inspire increased confidence among collaborating searchers that they have encountered all of the relevant sites during an exploratory search. Additionally, the ability of collaborators to view, and perhaps rate or comment on, the results contributed by others to a shared search could improve confidence in the quality of the results found.

Exposure

Collaborative search interfaces have the potential to serve as mechanisms not only for performing searches, but for helping users improve their own searching techniques. For example, providing users with awareness of query formulations used by their teammates could expose them to previously unknown syntax, which they could then use in their own future searches.

Productivity

By allowing people to collaborate with friends or colleagues on searches of mutual interest, redundant work could be reduced or eliminated, thus improving net productivity. For example, colleagues attempting to do a shared search now (for instance, by each searching independently on their own computers and then emailing each other lists of relevant links) must either spend time explicitly dividing up the search into subtasks for each user, or potentially duplicate each others' efforts by using the same keywords, finding the same documents, etc. In contrast, a system that provided a shared workspace with awareness of keywords used and sites flagged by each collaborator could reduce the overhead of coordination and duplication of effort.

An additional productivity benefit of interfaces enabling collaboration during Web search occurs in single-user scenarios. One important component of any collaborative system for exploratory search would be a persistent representation of the current state of the search process (*i.e.,* what queries have been issued, what results have been found, etc.), so that each of the collaborators could examine and add to the search session. This persistent representation also adds value to the single-user exploratory search experience since, after all, a user often must "collaborate" with him/herself, either across time (*i.e.,* beginning a search task that is interrupted and resumed hours, days, or weeks later) or across locales (*i.e.,* beginning a search task in the office, and continuing it on a different computer at home). By maintaining a persistent search state, single-user productivity could be improved by avoiding duplication of effort and by potentially increasing the speed with which a person can regain the context of their search task when resuming after a gap in time.

CONCLUSION

We have presented an overview of the current state of collaborative search interfaces, which shows the current lack of systems supporting collaboration for exploratory search tasks. Our survey on Web search habits showed that cooperative search is already a part of current work practice, although users must develop workarounds in order to share the process and results of a search with others. Data from our sample exploratory search task demonstrated potential benefits to providing formal support for group searching. In particular, we have described four aspects of exploratory search (*coverage*, *confidence*, *exposure*, and *productivity*) that stand to benefit from collaborative search UIs.

ACKNOWLEDGMENTS

We thank Eric Horvitz and Susan Dumais for their insightful discussions and comments.

REFERENCES

1. Beeferman, D. and Berger, A. Agglomerative Clustering of a Search Engine Query Log. *Proc KDD 2000*, 407-416.
2. Jones, R., Rey, B., Madani, O., and Greiner, W. Generating Query Substitutions. *Proceedings of WWW 2006*, 387-396.
3. Krishnappa, R. Multi-User Search Engine: Supporting Collaborative Information Seeking and Retrieval. *Master's Thesis, University of Missouri-Rolla*, 2005.
4. Large, A., Beheshti, J., and Rahman, T. Gender Differences in Collaborative Web Searching Behavior: An Elementary School Study. *Information Processing and Management*, Vol. 38, 2002, 427-443.
5. Maekawa, T., Hara, T., and Nishio, S. A Collaborative Web Browsing System for Multiple Mobile Users. *Proceedings of PERCOM 2006*, 22-35.
6. Morris, M.R., Paepcke, A., and Winograd, T. TeamSearch: Comparing Techniques for Co-Present Collaborative Search of Digital Media. *Proceedings of IEEE Tabletop 2006*, 97-104.
7. Pujol, J.M., Sangüesa, R., and Bermúdez, J. Porqpine: A Distributed and Collaborative Search Engine. *Proceedings of WWW 2003 (poster)*.
8. Russell, D., Stefik, M., Pirolli, P., and Card, S. The Cost Structure of Sensemaking. *Proceedings of CHI 1993*, 269-276.
9. Twidale, M., Nichols, D., and Paice, C. Browsing is a Collaborative Process. *Information Processing & Management*, 33(6), 1997, 761-83.

Patterns of Tag Usage in Enterprise Tagging Services: A Constraint on Exploratory Search?

Michael J. Muller
IBM Research / Collaborative User Experience
One Rogers Street
Cambridge, MA, USA 02142
michael_muller@us.ibm.com

Abstract: We compare four tagging-based enterprise services, that respectively stored bookmarks to web-pages and documents, to people, to blog entries, and to hierarchically-structured activity records. Analysis of user data and tag data showed relatively small overlaps in tags used. Conventional normalization strategies produced only modest improvement. These results suggest difficulties in combining exploratory searches across multiple social-tagging services. We discuss potential solutions.

In our research on social computing within the enterprise, we are concerned to understand how people interpret and use tags in different contexts. Tagging services within an enterprise can take advantage of full authentication of users, facilitating the identification of experts [3,5], people with shared interests [6], features of a shared vocabulary [1], and colleague-based strategies of information search [6]. Some of these strategies may be available, perhaps to a lesser extent, for searches on the public internet.

Full authentication also makes it possible to compare users' tagging practices in multiple tagging-based services. This paper goes beyond the previous research, by examining tagging behaviors across four distinct tagging-based services, with partially-overlapping user populations, within a single enterprise. This paper takes a comparative view that was not possible in previously published single-service analyses.

Damianos et al. [1] described a pattern of vocabulary stabilization with increasing tag usage (see also [6]). We hope to learn if people expect each tag to have a stable, consistent meaning across services, and if their tag usage is consistent across the multiple types of objects stored in the multiple services. These studies will help us to understand the semantics of social software, which will in turn help us to understand the advantages and disadvantages to such semantics across multiple tag-based services.

Four Enterprise Tag-Based Services

We collected tag data from the four services summarized in Table 1. **Dogear** is a social-tagging service for resources such as public URLs, company-internal URLs, and other company internal documents (e.g., Wiki pages, Domino documents, etc.) [6]. Dogear bookmarks are readable by anyone in the company. **Bluepages+1** is an enhanced version of the company's online employee directory, in which one person may apply a tag directly to another person's directory page [2, 3]. Bluepages+1 data (including tags) are readable by anyone in the company. **Blog Central** is an internal blogging service, open to any employee. All entries in Blog Central are readable by anyone in the company. Finally, **Activities** [7] is a web-based version of ActivityExplorer [8], an activity-centric collaboration service in which teams may create a collections of diverse objects in a tree-like structure consisting of a root "activity" and its daughter components. Unlike the other three services, Activities restricts access to each activity to a specified access control list, which may be as small as one person, or as large as several hundred people.

Analysis of Tags

For the purposes of this paper, data consisted primarily of user-tag tuples. Note that previous studies of tagging behaviors *within a single service* have included not only user and tag (as in our study), but also the tagged object [1, 2, 3, 4, 5, 6]. By contrast, this paper is concerned with tagging *across diverse services*, where each service is concerned with different types of tagged objects (e.g., documents vs. people vs. blogs vs. activity components).

System	What is tagged	Number of Taggers	Number of Unique Tags	Potential Number of Readers
Dogear	Public URLs, internal URLs, and other internal documents	1710	21072	As many as 300,000 IBM employees
Bluepages+1	Persons	713	2992	As many as 300,000 IBM employees
Blog Central	Blog entries (tagged by author only)	2092	3322	As many as 300,000 IBM employees
Activities	Activity root object + daughter components of the activity	742	3269	From 0 to several hundred

Table 1. The four internal tagging services in this study. Data were collected from Dogear, Bluepages+1, and Activities on 21 July 2006. Data were collected from Blog Central on 9 August 2006. Usage has increased since those dates.

The same objects cannot appear in different services, because each service stores a different class of objects. Therefore, for this initial report, we omit references to the tagged objects, and focus only on the users and the tags.

Common Users

Analyses showed complex patterns of overlaps. Many of the 4987 unique users participated in more than one of the services. On a pairwise basis, the number of people writing tags into each respective *pair* of services ranged from 193 users (Bluepages+1 and Blog Central) to 613 users (Dogear and Activities). Smaller numbers of users participated in more complex intersections of the services, with only 79 of the users participating in all four services.

Tag Re-Use Analyzed by Services

When we examined tags in the four services, we found similar, complex patterns. Of the 28460 unique tags, many appeared in more than one service. On a pairwise basis, the number of tags appearing in each pair of services ranged from 522 tags (Bluepages+1 & Activities) to 2953 tags (Dogear & Blog Central). Smaller numbers of tags appeared in more complex intersections of the services, with only 396 tags appearing in all four services.

Tags occurred with different relative frequencies in different services. Using the Fisher r-to-z transformations, we calculated the mean correlation of relative tag frequencies across services as $r=.286$. While significant ($p<.001$), this correlation explains only 8% of the variance. People appear to use tags in somewhat different ways in the different services.

Tag Re-Use by Individuals

In addition to comparing lists of users and tags on a *per-service* basis (above), we also conducted analyses at the level of each person and each tag. When we focused on personal re-use of tags across services, we found very low re-use rates (mean of 1.06 tags re-used per person, or mean of 2.65% of per-user opportunities to have tags in common). When we re-focused our analysis on the tags, and asked whether each tag was associated with the same users across services (a relationship that we called "tag membership"), we again found very low rates of commonality (mean of 0.35 people associated with the same tag in different services, or mean of 4.89% of per-tag opportunities to have users in common).[1]

Discussion of Low Tag Re-Use Rates

It could be argued that these results are not very surprising: If people were describing different types of objects, then shouldn't they use different vocabularies to do so? This explanation is not supported by the data. If people used different vocabularies, then we should have found very few tags in common between services. However, we found a minimum of 522 tags in common across pairs of services. Thus, the re-use of the tags, on a *service-by-service* basis, provided ample opportunity for the re-use of tags on a *person-by-person* basis. Nonetheless, we found very little personal re-use of tags. We also note that, at a conceptual basis, all four services were designed to support people's daily business work, and therefore it appears reasonable to expect people to use similar vocabularies to describe the related aspects of their work that involved different types of objects in different tag-based services. That is, we could easily expect that a user would bookmark a website (in Dogear) using the same project-related tags that s/he used to bookmark a collaboration partner (in Bluepages+1) and that s/he used to describe that project in a posting in Blog Central or Activities. But this did not appear to take place very often.

Another possibility is that people *did* re-use their tags, but did so with variations in spelling or capitalization or punctuation. We explored the following normalization strategies applied to all of the tags from all four of the services:

- "Folding" all alphabetic characters to lower-case
- Stemming (both Porter[2] and Paice/Husk[3] stemming)
- De-prefixing (e.g., removing a prefixed symbol such as "-" or "%")

The best results (achieved with a combination of de-prefixing and Porter stemming) led to a 21.3% improvement. This is promising, but a manual inspection of the most-frequently occurring tags in each combination of services showed many additional missed opportunities to find commonalities among tags – especially at the semantic and person levels (e.g., "Kim" vs. "KimChen" vs. kimch@us.ibm.com [fictitious name]).

---

[1] We considered that these results might have been due to a sampling oversight, i.e., the mistake of including people who might have created only a very small number of tags in a service. To test this sampling-error hypothesis, we sorted each set of data in terms of the minimum number of tags created by the user in *each* of the services under analysis. We then experimented with temporarily removing the users with the lowest numbers of tags in one service or the other. We systematically tried all possible cut-off usage rates. None of the cut-off rates substantially increased the re-use of tags by persons. None of the cut-off rates substantially increased the number of people associated with each tag. The phenomena of low tag-re-use and low tag-membership appear to be robust across users with *any* number of tags.

[2] http://www.tartarus.org/~martin/PorterStemmer/, retrieved 12 January 2007.

[3] http://www.comp.lancs.ac.uk/computing/research/stemming/, retrieved 12 January 2007.

Other research within two of these services has shown that people devote considerable effort into writing tags [6], and that in some cases they use the tags as part of managing their reputations within the enterprise [3]. We are left with the apparent paradox that people work hard to write tags, but seem not to do this consistently for related aspects of their work that are stored in different services.

Conclusion

It appears reasonable to expect the ability to re-use tags from one domain to another. These results suggest that tag re-use may prove more difficult than anticipated. Various methods are under investigation at internet sites to assist users in the operation of *initial tag-entry* to re-use the tags that are already available at that site (e.g., type-ahead suggestions). However, these strategies are unlikely to assist with the problems described in this paper, because part of the vocabulary "drift" appears to have occurred from one service to another. Indeed, the largest of the enterprise systems already implements a form of type-ahead-based assistance for tag-entry, and its vocabulary continues to contain many variants on what appear to be common tags.

A different strategy could be adopted at the point of *search specification*, rather than of tag-entry. This strategy could take advantage of patterns of tag co-occurrence within systems, and could ultimately be extended to comparisons and inter-communications of patterns of tag co-occurrence between systems.

Future analyses will attempt to explain this apparent paradox through more a detailed comparison of the tags that do and that do not involve tag re-use and tag membership across services. Future analyses will also examine co-occurrence and network analyses of tags and taggers, as well as the semantics of the tags.

We anticipate combining these analyses with other examinations of tagging patterns across diverse services and resources, to understand the emerging tag-based work practices, and to clarify the value of social software data and services within the enterprise.

References

[1] Damianos, L., Griffith, J., & Cuomo, D., "Onomi: Social Bookmarking on a Corporate Intranet," *Proc WWW 2006*.

[2] Farrell, S., & Lau, T., "Fringe Contacts: People-Tagging for the Enterprise. *Proc WWW 2006*.

[3] Farrell, S., Lau, T., Muller, M.J., & Ehrlich, K., "Augmenting employee profiles with people-tagging," Submitted to CHI 2007.

[4] Golder, S.A., & Huberman, B.A., "Structure of Collaborative Tagging Systems," *J. Info. Sci. 32*, 2 (April, 2006).

[5] John, A., & Seligmann, D., "Collaborative Tagging and Expertise in the Enterprise," in *Proc WWW 2006*.

[6] Millen, D.R., Feinberg, J., & Kerr, B., "Dogear: Social Bookmarking in the Enterprise," *Proc CHI 2006*.

[7] Moore, M., Estrada, M., Finley, T., Muller, M.J., & Geyer, W., "Next generation activity-centric computing. Demo at CSCW 2006.

[8] Muller, M., Geyer, W., Brownholtz, B., Wilcox, E., & Millen, D.R., "One Hundred Days in an Activity-Centric Collaboration Environment based on Shared Objects," *Proc. CHI 2004*.

Maintaining Consistency without Stagnation during Exploratory Search

Jaime Teevan
Microsoft Research
Redmond, WA 98052 USA
teevan@microsoft.com

ABSTRACT
Many real life search tasks are complex, multi-stepped processes. As people explore a search space, they want to find the most relevant information available to them while remaining oriented in the space they have already explored. This paper argues that during exploration it is important for search engines to present relevant information to their users in a way that maintains the users' existing contexts. This means that the most relevant results should not necessarily be ranked first, but rather ranked where users expect them to occur. The paper presents a model of what people remember about search results, and shows that it is possible to invisibly merge new information into previously viewed search result lists where information has been forgotten.

INTRODUCTION
Consider as an example of an exploratory search Connie's search for breast cancer treatments. Connie was recently diagnosed with breast cancer and wants to learn more about the available treatment options. For this reason, she runs a search for "breast cancer treatments". The result list returned to her is shown in Figure 1. Several results from the National Cancer Institute are listed first, followed by a result about alternative treatments, a link to About.com's page on treatments for breast cancer, and so on. The government pages appear too technical to interest Connie, and she is not generally interested in alternative treatments, so she skips over the first couple of results in the list and decides to follow the fourth link to an About.com page.

As Connie explores treatment options, it becomes possible for the search engine to identify results she may find more relevant to her search. Connie provides implicit feedback about what she considers relevant and irrelevant in the links she chooses to follow. She may also be willing to provide explicit feedback or query refinements because this topic is important to her. Further, her information need may evolve in predictable ways as she learns more about the topic, and new timely information about the latest treatments may become available as her search extends over time.

Although new, more relevant results can benefit Connie, naively re-ranking the search results she has already seen to place the better results first is not necessarily the best way to help satisfy her information need. Connie has developed expectations about what results the search result list for "breast cancer treatments" contains during her initial interactions with the list. If, for example, the About.com page she clicked on was no longer ranked about fourth in the list, she would have trouble returning to it.

This paper explores how consistency can be maintained in search result lists during long search sessions where new

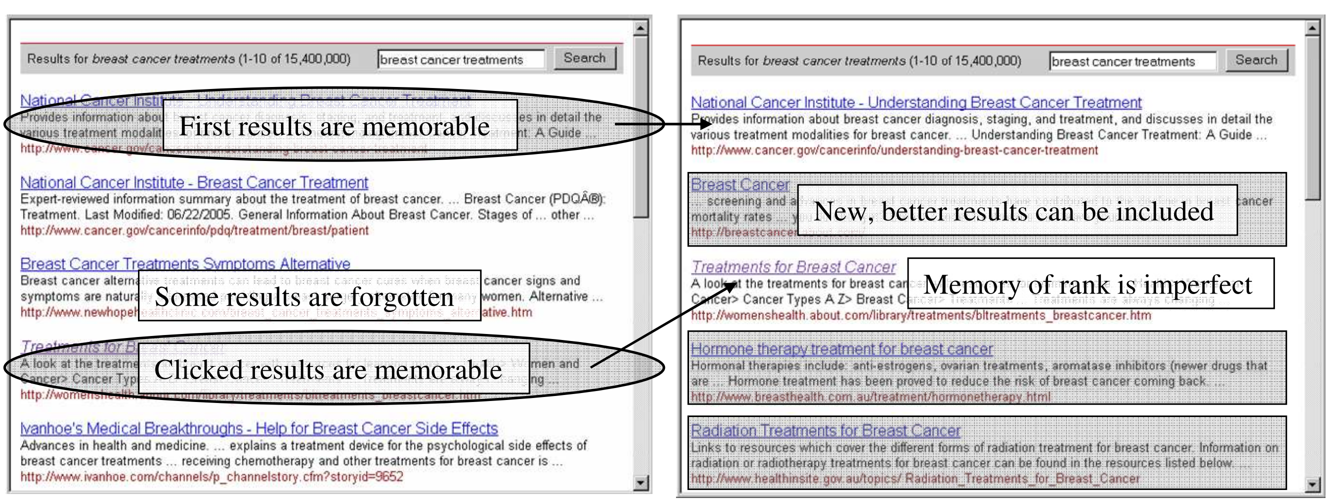

Figure 1. On the left is the search result list originally returned to Connie. On the right is a search result list that contains the results Connie remembers having seen before where she expects them that still includes new results.

information becomes available. It begins by demonstrating that changes to search result lists, even when beneficial, can cause disorientation. It then shows that changes can be made undisruptive by taking advantage of people's limited memories. Rather than keeping an entire result list static when a person returns to it, only that information that is remembered need be kept the same. New information can be snuck into the holes where results were forgotten.

Figure 1 illustrates the way that Connie's lapses in memory can be used to her advantage. Because she only remembers a few results in the original list returned for "breast cancer treatments", those results can be held constant while new more relevant results are added. The merged list is likely to look the same as the old list to her, despite containing new useful information. Sneaking new results into the result list can help Connie find new information while not interfering with her ability to re-find previously viewed information.

RESULT LIST CHANGES CAUSE PROBLEMS

Maintaining consistency is important because time and time again, changes to electronic information that should help the user get in the way. For example, dynamic menus were developed to help people access menu items faster by bubbling common items to the top of the menu. Rather than decreasing access time, research revealed that dynamic menus actually slow their users down because commonly sought items no longer appear where expected [4, 7].

Selberg and Etzioni [6] studied the rate of change of search result lists, and found result lists change rapidly. They noted that, "Unstable search engine results are counter-intuitive for the average user, leading to potential confusion and frustration when trying to reproduce the results of previous searches." Problems of this type appear in a study by White, Ruthven, and Jose [11]. In this study, the authors tried to help people search by giving them lists of relevant sentences that were dynamically re-ranked using implicit feedback gathered during the search. However, people did not enjoy the search experience as much or perform as well as they did when the sentence list was static.

The problems caused by changes to result lists were further explored in a study by Teevan, Adar, Jones and Potts [9]. They analyzed the queries issued by 114 people to the Yahoo search engine over a year, and found that when a previously clicked result changed position in the result list, users were less likely to re-click results. This suggests that changes to result ordering caused people to re-find less information and view more new information. A reduction in re-finding is not necessarily a bad thing if the new results are better than what was previously available. However, the authors observed that when the searcher clicked on a previously viewed result, the time it took to make the click was significantly longer if the result's rank had changed.

Because changes to result ordering slow re-finding, it is likely that highlighting previously viewed results, as is done by Google and A9, is not enough to support truly natural re-finding – not only because the result being re-found may have disappeared from the result list, but also because how it is re-found is not the same as how it was originally found. However, highlighting previously viewed results could be used to supplement the approach presented here.

Information management systems that permit a consistent interaction allow users to choose to interact with a cached version of their information space [2, 5]. Unfortunately, caching denies users the opportunity to discover new information. For example, Connie would not be able to revisit previously found information on breast cancer treatments while still learning about newly available treatments. People regularly find new information while re-finding. Although repeat clicks are common for repeat queries, 27% of all repeat searches involve the finding of new information as well [9].

MAKING CHANGES WITHOUT THE PROBLEMS

Fortunately, not all changes to search result lists result in a loss of context. It is possible for search tools to maintain the appearance of consistency while still presenting their users with the latest, most relevant information. This can be done by taking advantage of the fact that people do not remember all of the information they see. New information can be snuck into the holes where information has been forgotten. This section presents a model of what is memorable about search result lists, and shows how that model can be used to change unmemorable aspects of result lists to include new results without notice. A study is then presented that shows that sneaking in new results enables people to find new information quickly without destroying their interaction with previously viewed information.

Modeling What is Memorable

A study of 119 people was conducted to model what is memorable about search result lists [8]. In the study, participants were asked to interact naturally with a list of results for a self-generated query, and then later asked an hour later to recall the list without referring back to it. Two main factors emerged from the data as affecting how memorable a result was: where it was ranked and whether it was clicked. These factors were used to model which aspects of a search result list should be changed with care, and which could be changed freely. An intelligent merging algorithm was developed that merges new information into an existing result list by considering all permutations of old and new results and choosing the one with the highest benefit of both old and new information.

Model Allows New Information to be Snuck In

A second study was conducted with 165 different people to test the ability of the merge algorithm to hide change. Participants were asked to interact with a list of results, and a day later were asked to determine whether a follow-up result list was the same as or different from the original list.

Participants often noticed the follow-up list was different from the original list. When the follow-up list consisted of entirely new results, participants reported the list had changed 81% of the time. When six random results were changed, the change was noticed 62% of the time, and when the clicked results were listed first and all other results were new, change was noticed 59% of the time. The differences between these cases are not significant.

However, when changes were made intelligently merging new information into the original list, the follow-up result list appeared static to participants. In this case, differences were observed only 19% of the time – less often, in fact, than the 31% of the time that a change was noticed when the follow-up list was unchanged. The disparity between the intelligent merging and the static list is not significant, but could possibly reflect the fact that changes to the result list made according to the model may create a list that looks more like the list the participant remembers than the actual original. The result lists from these cases were significantly more likely to be considered the same as the original list than any of the three cases with naïve change (p<0.01).

Merged Lists Support Finding and Consistency

Even though the merged list looks the same as the original list, the inclusion of new and better results can satisfy the user's information needs sooner. Usability improvements do not need to be noticed to benefit the user. A classic example is the Macintosh design for cascading submenus, where some flexibility in navigating to menu items is built into the menu design. The tolerance for small errors in navigation goes unnoticed by almost all users, but leads to fewer errors overall [10]. Similarly, a study of an improvement to cascading submenus showed all users performed better even though only three out of the 18 participants actually noticed the change [1].

To understand whether people were able to both maintain their context and find new information using the merged result lists, a subsequent study of 30 people performing re-finding and new-finding tasks was conducted. The study involved two parts: 1) an initial session where participants conducted initial finding tasks, and 2) a follow-up session where participants conducted finding and re-finding tasks.

When the list used for the follow-up session was the same as the list used during the initial session, participants were able to re-find information easily. However, they were unable to find new information. In contrast, when the follow-up list was comprised of entirely new information, participants were able to find new information easily, but could not complete the re-finding tasks. The intelligent merging performed closely to the best in both cases, allowing participants to find new information as if the list were entirely comprised of new results and re-find old information as if the list were unchanged. When the merging was done randomly, rather than according to the model, people performed comparatively worse.

Sneaking new information into the holes in a result list where people have forgotten results appears to be a good compromise for exploratory search. It allows consistency to be maintained during extended search sessions, but permits the inclusion of new information.

CONCLUSION

This paper investigated the importance of consistency during exploratory search. Changes to result lists were shown to create problems for people during the finding process. Although the ability to find new information may appear at odds with the maintenance of context, a solution was presented where new relevant results were not ranked first, but rather where the user expects them.

Looking forward, the effective management of changing information will be essential to successfully supporting complex finding behavior. The growing ease of electronic communication and collaboration, the rising availability of time dependent information, and the introduction of automated agents, suggest information is becoming ever more dynamic. Even traditionally static information like a directory listing on a personal computer has begun to become dynamic; Apple, for example, has introduced "smart folders" that base their content on queries and change as new information becomes available. As Levy [3] observed, "[P]art of the social and technical work in the decades ahead will be to figure out how to provide the appropriate measure of fixity in the digital domain." The solution presented here is a good first step towards that end.

REFERENCES

1. Ahlström, D. (2005). Modeling and improving selection in cascading pull-down menus using Fitts' law, the steering law and force fields. In *Proceedings of CHI '05*, 61-70.
2. Hayashi, K., Nomura, T., Hazama, T., Takeoka, M., Hashimoto, S., and Gudmundson, S. (1998). Temporally-threaded workspace: A model for providing activity-based perspectives on document spaces. In *Proceeding of HyperText '98.*
3. Levy, D. (1994). Fixed or fluid? Document stability and new media. In *Proceedings of European Conference on Hypertext*.
4. Mitchell, J. and Shneiderman, B. (1989). Dynamic versus static menus: An exploratory comparison. *ACM SIGCHI Bulletin*, 20(4): 33-37.
5. Rekimoto, J. (1999). Time-machine computing: A time-centric approach for the information environment. In *Proceedings of UIST '99*, 45-54.
6. Selberg, E. and Etzioni, O. (2000). On the instability of Web search engines. In *Proceedings of RIAO '00.*
7. Somberg, B. L. (1986). A Comparison of Rule-Based and Positionally Constant Arrangements of Computer Menu Items. In *Proceedings of CHI/GI '86*, 255-260.
8. Teevan, J. (2006). What People Remember About Search Result Lists. In *Proceedings of CHI '06*.
9. Teevan, J., Adar, E., Jones, R., and Potts, M. (2005). History repeats itself: Repeat queries in Yahoo's query logs. In *Proceedings of SIGIR '06,* 703-704.
10. Tognazzini (1999). A quiz designed to give you Fitts. http://www.asktog.com/columns/022DesignedToGiveFitts.html.
11. White, R., Ruthven, I., and Jose, J.M. (2002). Finding relevant documents using top ranking sentences: An evaluation of two alternative schemes. In *Proceedings of SIGIR '02*, 57-64.

Bridging the Gap: Using IR Models for Evaluating Exploratory Search Interfaces

Max L. Wilson, m.c. schraefel
School of Electronics and Computer Science
University of Southampton, UK
[mlw05r, mc]@ecs.soton.ac.uk

ABSTRACT
Exploratory Search Interfaces are being designed to support increasingly high-level search activities, such as comparison and aggregation. This position paper suggests that the history of research into user search behaviour may provide grounds for developing a combined model approach that evaluates features designed to support such exploratory search activities.

Author Keywords
HCI, Searching, Search, Exploratory, Exploratory Search, Theory, Evaluation, Comparison

ACM Classification Keywords
H.5.2 User Interfaces, H.1.2 User/Machine Systems, H.3.3 Information Search and Retrieval

INTRODUCTION
Keyword searching has become the standard model of search, through the popularity of web giants like Google. Yet research has become directed at richer modes of search, known as Exploratory Search [9]. Naturally, some recent Human Computer Interaction (HCI) research has been investigating the design of Exploratory Search Interfaces (ESIs) for supporting users who have less clear or more complex needs.

There are many ways to design support for more complex modes of search and it would be useful to frame ESI development in an approach that HCI designers can apply to evaluate their work. A first approach may be to see if Information Retrieval (IR) research into user search behaviour, dating back to the 1980s [1, 5], can be used to model the success of any potential designs.

In this position paper, we propose an approach that combines two established IR models for evaluating ESIs. We show briefly how this approach might be applied to example ESIs, in this case three Faceted Browsers, which identified strengths and weakness in feature design for the support for users in different conditions. In the conclusion, we touch on the next steps for expanding this approach.

RELATED WORK
Exploratory search has been defined as more complex IR activities that may be required by users who do not have clearly defined goals, have changing complex needs or may be using a system that is poorly indexed [9]. Marchionini suggested some of these possible activities, including: Aggregation, Comparison and Evaluation [4]. In these conditions, simple keyword search may not support users effectively and convoluted user coping strategies have been recorded that involve iterative tentative guessing of keywords [6]. Clearly, what these users require is an alternative method of search that may involve more browsing and exploring activities to achieve their goals. This has been the seed for developing ESIs that support many richer modes of search.

Faceted browsers are an example class of ESIs, which present meta-data attributes as a series of selectable categorised options. Through modelling the domain of information through a faceted structure, direct manipulation can be used to construct queries. Thus, when a user is not clear on appropriate terminology or the meta-data is unpredictable, they do not have to estimate search terms, but can make selections to build their query. Through this extra support, faceted browsers can be considered a type of ESI.

There are many more features that can support Exploratory Search. For example, mSpace [8] is a faceted browser that also includes: a collection space, a multimedia preview cue and a persistent panel for contextual information. Further, each of these main features are designed to be persistently in view and available for use. As more features are developed and added to software, there must be rationale for how and when to expose features to support users effectively. This can not be done without evaluating how easily these features can be used and when they might be appropriate. Models of user search behaviour may be able to provide this information.

Traditionally, there are two overview styles of model in IR research: Holistic System models and Interactive models. Saracevic's stratified model [7] is an example of a holistic model that describes the different layers of an IR system: both computer and human. While both contain cascading and interacting levels, here we are concerned with analysing *user* search behaviour. The user has cascading levels including, from the top: Situational, Affective, Cognitive and Query Generation. That is, mental query generation is affected by their situational tasks, the affective intentions and their existing cognitive knowledge. It should be important to consider each of these aspects of a user when defining complex needs and subsequent requirements for a system.

Belkin *et al.* developed one of the more popular interactive models of IR, defining a series of typical episodes and scripts for the interaction between users and IR systems [2]. Included in this research, however, was an initial model for situational and cognitive user needs. They identified four dimensions, which together produce 16 unique Information-Seeking Strategy (ISS) conditions. *Method* describes whether a user is either searching for an information object, or scanning a set of information objects. This is easily differentiated by finding a specific paper in order to get its reference details, or by searching for an possible paper, which may not exist, that can be used to support a point. *Goal* describes whether a user is learning about something or selecting something. Using the bibliographic example differentiates these as researching a topic, or finding a reference. *Mode* is between recognising and specifying something. One might remember that there was a useful publication at CHI2005 and so is trying to identify it in the proceedings, or may have known the author, title and year and has typed them into the ACM Portal[1]. *Resource* is between wanting information items or meta data about an information item. Usually, with a bibliographic repository users are trying to find specific papers, but it is possible that the user is trying to find out first what workshops existed in a conference so that they can better define a search query at a later point in time.

In 1990, Bates described a four-level hierarchy of search activities: Move, Tactic, Stratagem and Strategy [1]. The first of these is a single action performed by the user, either physically or mentally: mental actions may be deciding or reading. A tactic is a combination of moves and there are endless combinations of moves that can be used to support a tactic, which depends on system implementations. She defines 32 specific information search tactics that a search system should support. Stratagems are a larger combination of both individual moves and tactics: some examples include performing a citation search or following a footnote. Marchionini noted a series of exploratory search activities, which could be considered as Stratagems [4]. Exploratory search activities could be considered complex combinations of tactics and moves, whereas a simple lookup could be a simple set of tactics and moves. Strategies are again higher and involve a combination of moves, tactics and stratagems: this might be finding relevant work for a paper and depends heavily on what the user is currently working on.

PROPOSED MODEL

Design

When considering users of a system, a user's strategy may have led the user to the system, as it usually represents a certain type of resource, such as a journal archive or a product collection. These strategies relate quite well to Saracevic's situational tasks. Once using a particular resource, the user may wish to employ a set of stratagems to achieve their goal, and modelling these in full should relate directly to Saracevic's affective intentions level. Similarly, tactics, may also be considered part of Saracevic's affective intentions. Finally, moves can relate to Saracevic's query generation, as each action should contribute towards exploring the information set. The missing level of Saracevic's model is the cognitive level, defining existing knowledge, for example. This can be modelled by using Belkin's user conditions, which incorporates things such as previous knowledge. This model actually touches on a few levels of Saracevic's model, including intention.

As a result a combination of Bates' and Belkin's models has been suggested. Bates' moves are used to quantify the support for tactics by interface features. Then these tactics are applied to support Belkin's dimensions, so that the support for the sixteen conditions can be calculated.

Stage 1: Feature Identification. First, the interface features and their interactions must be identified. For example, mSpace has a set of features including: browser columns, a collection space, a preview player and an information panel. The features of each design should be incorporated.

Stage 2: Measuring Support for Tactics. Each interface feature is addressed one at a time, for each design. For the current feature of the current design, the moves required to support each tactic are counted. This produces a series of tables, one for each design, where tactics are listed across the top and the interface features down the side. The count of moves is noted in the appropriate cross section between feature and tactic. No support by a feature for a tactic counts as 0. Four moves for the user to use the feature in support of a tactic counts counts as 4. Repeat and Optional moves are ignored. For example, selecting multiple items involves choosing and selecting 2+ items, selecting 3+ is considered a repeat move of selecting 2 items. Optional moves include scrolling: a desired item may be the first or last item. The optimum situation is that it is one of the items that is visible without scrolling.

Stage 3: Summarising Metrics. As no support is represented by zero, support in a single move is represented by 1 and support in ten moves by 10, all values above 0 must be inverted. Thus a feature that supports a tactic well approaches the value of 1 and a poor support approaches 0. These inverted metrics can then be summed by feature and by tactic. This calculates the support provided by a feature for all tactics and the support provided for a tactic across all features, respectively.

Stage 4: Feature Strength Analysis. A graph can be produced including the summed values for each feature in each design. An example can be seen in the following section. Strong features will score produce tall bars, and a quick comparison of user effort can indicate a strong feature design.

Stage 5: Tactic Support Analysis. A graph can be produced including the summed values for each tactic in each design. Again, tall bars indicate strong support for a tactic. This comparison may identify tactics which may require improved support through redesign.

[1]http://portal.acm.org

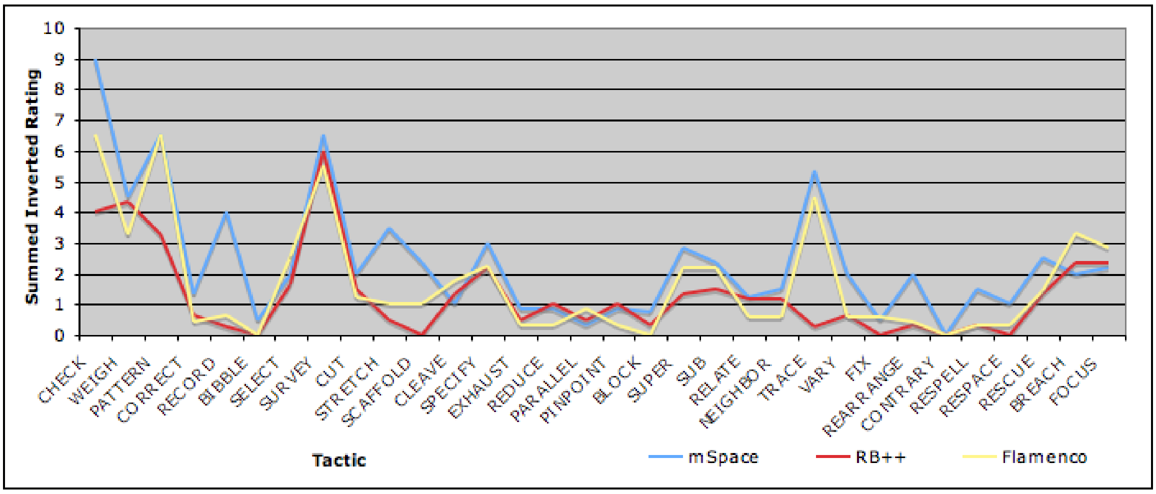

Figure 1. Graph Showing the Summed Inverted Metrics of each Tactic for the Three Browsers.

Stage 6: User Conditions Analysis. Each tactic supports particular ends of Belkin's dimensions of user conditions. CHECK, a tactic for users checking their decisions so far, supports users who are trying to Learn as their *Goal*. The support for a tactic by a design is added to the total support for a dimension. Then for each of the sixteen conditions, the sum of the total support values are calculated. This value for each condition can be graphed showing the difference in support for different user conditions.

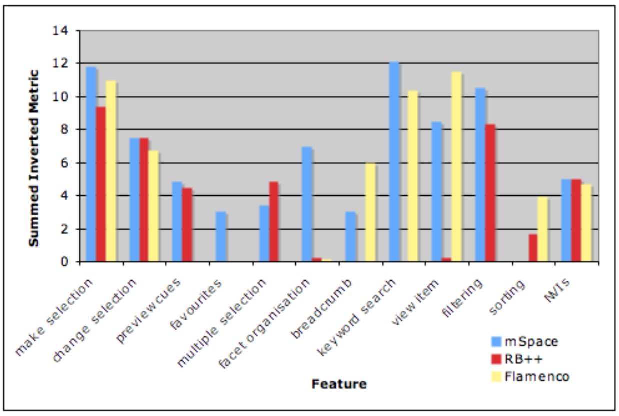

Figure 2. Graph Showing the Summed Inverted Metrics of each Feature for the Three Browsers. Taller bars are better.

Model Applied

In a recent application of the model, three faceted browsers were compared: mSpace[2], Flamenco[3] and RB++[4]. It is clear from Figure 1 that there are few tactics that are not best supported by the mSpace browser. This is explained, in Figure 2 by the number of strong contending features: there are comparably high mSpace bars for almost every feature. This is arguably representative of the focus+context design, which aims to present as many options and features to the user as possible and at all times. In line with the second observation, however, there are clearly some features of each interface that have stronger implementations of the three browsers. For example, multiple selection is easiest in RB++, yet keyword search is missing from RB++ and the implementation is strongest in mSpace. One feature missing from the mSpace implementation is the ability to sort items. The strongest implementation of this is the ability to group the results by any facet, as seen in Flamenco.

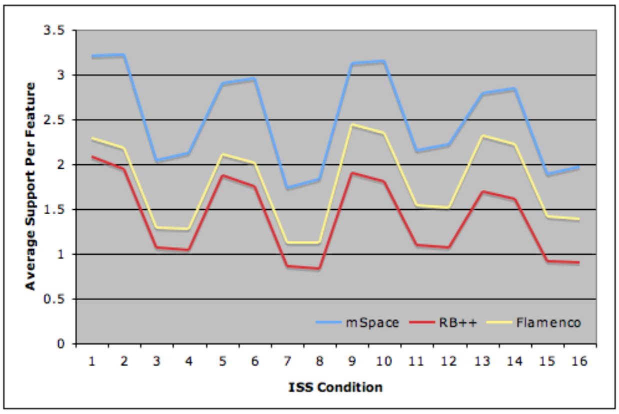

Figure 3. Graph Showing the Normalised Support for each ISS Condition by Faceted Browser

The pattern that is seen almost identically for each interface in Figure 3 is indicative of the mapping between Bates' tactics and the pattern of ISS conditions defined by Belkin *et al.* Predictably, as was shown in Figure 1, there are three distinct lines, showing that mSpace provides the widest support for search. This height difference does not show us new information. Instead what should be drawn from the graph is hidden within this pattern and shown in the differences in peaks and troughs for each interface condition. Quite clearly the graphs rise and fall in alternating pairs. This represents the alternation between recognise and specify (*Mode*) and is perhaps a predictable outcome for faceted

[2]http://mspace.fm
[3]http://flamenco.berkeley.edu/
[4]http://idl.ils.unc.edu/rave/

browsers. By including more lessons learnt from the IR work on keyword search, such as relevance feedback, we might see a balance between these two conditions. Within each of these alternating pairs, the mSpace line marginally increases where the others fall. This indicates an increased support for meta-information (*Resource*). Considering individual browser lines, while RB++ and Flamenco follow a similar pattern for the first 8 ISS conditions, Flamenco notably improves this gap in the final 8 conditions. These two halves are made unique by the *Method* dimension and indicates that Flamenco provides better support for search, which is defined by having a known target object to exist: this might be knowing that an academic paper exists and just trying to find it. This significant increase, also sharper than mSpace, may be present due to the better support for advancing selections and the lower support for changing selections.

The final pattern we draw from Figure 3 is shown every four conditions and is controlled by Belkin's *Goal* dimension. The Learn aspect of this dimension is shown by height differences between ISS1-4 and ISS5-8, and again between ISS9-12 and ISS13-16. This is characterised by the ability to see options in faceted browsers. The persistence of these options shown throughout to the user of mSpace is highlighted by the exaggerated difference in the first and third troughs compared to the second and fourth.

OPEN RESEARCH QUESTIONS

There are a few areas where this approach could be extended. First, the definition of a single move is still open to some interpretation. In the application above, choosing mentally and selecting physically are separated as different moves. Cognitive models exist that could be used to define moves in a clear manner. Further, cognitive load is an increasingly important problem for rich interfaces, and the effect of adding more features should also be modelled; an interface with 100 features may support many tactics but may not be easy to use. This is already modelled to some extent, as it may take more moves to access a feature if you have to select and open it before it can be used.

Of the four user levels of Saracevic's stratified model, there are still gaps for improvement. Belkin is one model of user intentions, but more exist. The same team later developed user and search conditions into much greater detail [3]. Further, Saracevic's affective intentions can be better modelled by trying to define Stratagems like Marchionini's exploratory search activities. These activities in particular are largely undefined, but where definition is missing, a model may be available or developed to extend this research. Also, this research is currently trying to model the user levels defined by Saracevic in 1997. There may be more recent or more specific research that models the user in more detail.

Finally, the current model is only useful when comparing designs and implementations. Ideally the model can identify a strong design or a weak design independently, so that it can be applied where there is no benchmark for comparison.

CONCLUSION

In this paper, we have described an approach for evaluating Exploratory Search Interfaces for strengths and weaknesses in terms of feature design for the support of users in different search conditions. This approach combines two established models of Information Retrieval to quantify a comparison of three Faceted Browsers, as example Exploratory Search Interfaces, and identified some clear areas for improvement on each. The approach can be applied to existing implementations, as above, or to new designs for exploratory search interfaces. Designs can be measured in their support for search tactics and strengthened before user studies begin.

These are early steps to a potentially promising approach for blending IR frameworks with HCI design. Based on these initial studies, we are investigating several refinements of the approach. In particular we are considering strategies to correlate the predictive findings presented here against experimental studies to validate and refine the approach proposed. We anticipate that with such validation, this hybrid IR/HCI approach will be a useful design tool for the development of exploratory search interfaces.

REFERENCES

1. BATES, M. J. Where should the person stop and the information search interface start? *Inf. Process. Manage. 26*, 5 (1990), 575–591.
2. BELKIN, N. J., MARCHETTI, P. G., AND COOL, C. Braque: design of an interface to support user interaction in information retrieval. *Inf. Process. Manage. 29*, 3 (1993), 325–344.
3. COOL, C., AND BELKIN, N. J. A classification of interactions with information. In *Proc. CoLIS 4* (2002), pp. 1–15.
4. MARCHIONINI, G. Exploratory search: from finding to understanding. *Commun. ACM 49*, 4 (2006), 41–46.
5. PEJTERSEN, A. M. A library system for information retrieval based on a cognitive task analysis and supported by an icon-based interface. In *Proc. ACM SIGIR* (New York, NY, USA, 1989), ACM Press, pp. 40–47.
6. PIROLLI, P., AND CARD, S. Information foraging in information access environments. In *Proc. ACM SIGCHI* (New York, NY, USA, 1995), ACM Press/Addison-Wesley Publishing Co., pp. 51–58.
7. SARACEVIC, T. The stratified mode of information retrieval interactino: Extension and applications. In *Proc. Info. Science* (1997), vol. 34, pp. 313–327.
8. SCHRAEFEL, M. C., WILSON, M. L., RUSSELL, A., AND SMITH, D. A. mspace: improving information access to multimedia domains with multimodal exploratory search. *Commun. ACM 49*, 4 (2006), 47–49.
9. WHITE, R. W., KULES, B., DRUCKER, S. M., AND M.C. SCHRAEFEL. Introduction. *Commun. ACM 49*, 4 (2006), 36–39.

Exploratory Search Using Timelines

Omar Alonso
Department of Computer Science
University of California, Davis, CA
oralonso@ucdavis.edu

Ricardo Baeza-Yates
Yahoo! Research
Barcelona, Spain
rbaeza@dcc.uchile.cl

Michael Gertz
Department of Computer Science
University of California, Davis, CA
gertz@cs.ucdavis.edu

ABSTRACT
As search applications keep gathering new and diverse information sources, presenting relevant information anchored in time becomes more important. Temporal information is available in every document either explicitly, e.g., in the form of temporal expressions, or implicitly in the form of metadata. Recognizing such temporal information and exploiting it for document retrieval and presentation purposes are important features that can significantly improve the functionality of search applications. In this paper, we present an exploratory search interface that uses timelines to present and explore search results. We also describe a prototypical implementation that illustrates the main ideas of our approach.

Author Keywords
Temporal expressions, temporal document retrieval, user interface, visualization.

ACM Classification Keywords
H.3.3 [Information Search and Retrieval]: Clustering, Retrieval Methods. H.5 [Information Interfaces and Presentation]: General

INTRODUCTION
Current interfaces to search engines typically present search results ordered by the relevance of documents from a document collection to a search query. For this, the freshness of the information, that is., documents or parts thereof, is considered an important part of the result quality. Temporal attributes in Web pages or documents such as date, however, are just viewed as some structured criteria to sort the results in descending order of relevance.

In search situations where the task requires the browsing and exploration of a search result [11], we argue that temporal information can help significantly to accomplish respective tasks. The presentation of relevant information along a well-defined and understood timeline is an important step to find, for example, the most recent document relevant to a query or the first point in time a document (based on the temporal information contained in the document) is relevant to the query.

Our approach supports the exploitation of temporal information in documents, and the usage of such information to anchor search results along a well-defined timeline. We believe that such timelines should be an essential part of every exploratory document search system.

Research in using time for retrieval and browsing activities is fairly recent. The use of tags to visualize photos taken over a period of time is a good example of how useful time can be for arranging objects [4]. Furthermore, in addition to topic detection and tracking, the discovery of bursts in a stream of content can be useful for the identification of topics [6]. In particular, in settings where a user is looking for relevant documents in a less familiar domain, we would like to show peaks of activity (in the form of documents) over time that contain information relevant to the user query.

We agree with previous work that placing search results in a timeline can facilitate the exploration of information [1], [7], [8]. Our approach, however, differs in a number of ways. We do not restrict the search space to a personal desktop environment, because we believe that timelines should be an integral part of search applications. Ringel et al. [8] concentrate on "object" timestamp (e.g., a document, email message, or presentation), whereas we use both temporal expressions and document metadata as document timestamps alone can often be misleading. Exploiting the temporal expressions embedded within a document leads to a much richer framework for search result exploration.

TEMPORAL EXPRESSIONS
A document typically has temporal metadata, such as the creation date or modification date. We also observe that the content of a document typically has *temporal references* to past and/or future points in time. These temporal references are either (1) explicitly represented, such as a date and time in a calendar ("March 12, 2004"), (2) denoted as events that have an associated time value, such as a holiday

("Christmas" or "Thanksgivings"), or (3) represented by a vague time reference ("by Friday").

The detection and extraction of time information from documents utilizes *named-entity extraction* techniques [5]. The output of this document preprocessing step is a time-annotated document, where the desired time information and expressions are represented in a specific format, for example, in the form of XML, possibly outside the original document [10]. Named-entity extraction is a very scalable approach, e.g., for the identification of holidays, and does not just rely on document specific time mappings as employed, e.g., in an email-calendar based approach [7].

We illustrate the above ideas of exploiting temporal information associated with documents using a literature search scenario based on the DBLP bibliography data set [3], which contains detailed information about journal, conference, and workshop publications, in particular the date of each publication.

TIMELINE CONSTRUCTION

Time has been a subject of study in many disciplines, and it is usually represented as a continuum line with origin and no end. We assume a time representation based on the Gregorian calendar, with a single day being an atomic time interval called *chronon*. A timeline consists of a sequence of chronons, and optionally can have a start and/or end chronon. Consecutives chronons can be grouped into more coarse-grained time *granules* such as weeks, months, years, etc.

As stated before, our goal is to present search results that are arranged in a timeline. There are several ways to accomplish this. The most obvious approach is to use the document metadata for anchoring documents in a timeline. Unfortunately, if the time range is large, i.e., the documents' metadata are widespread over a long time interval, the timeline might be too fine-grained and therefore too large. It thus makes more sense to group a search result based on a more general, coarse-grained time granule, such as year, and if needed, allow the user to explore a particular year chronons at a more fine-grained level, e.g., at months, weeks, or days.

The timeline construction is based on a clustering algorithm that uses temporal expressions extracted from documents and anchoring these expressions in a timeline [2]. The first step is to match the query terms with the document text near temporal expressions. Since our document collection consist of research articles, all of them having publication year as a temporal expression anyway, the relevancy match is reduced to whether or not the query terms occur in the document.

The search engine retrieves the results by relevancy using a traditional tf/idf metric. The resulting hit list is then clustered by year and within each year, documents are ranked by score.

The timeline is made of clusters (labeled by year), and also provides for more fine-grained cluster exploration for each year. It is important to note that if a document has more than one temporal expression, it can appear in more than one cluster, and it can also be ranked differently in different clusters.

PROTOTYPE IMPLEMENTATION

The document collection consists of approximately 297,000 journal papers records from the DBLP bibliography data set. Since the data is available in XML format, we loaded the data into an Oracle database for storage. The extensions to SQL allow us to use XPath and search features for querying the text indexes.

For the visualization part, we use the SIMILE timeline AJAX toolkit [9], for presenting time-based events. The exploratory search prototype is a Web-based application and no specific plug-in is required.

The interface is organized as follows. The main section takes half of the screen and contains the search box and the timeline. The timeline consists of two *bands* that represent different time scales: decade and year. Both bands are synchronized such that panning one band also scrolls the other. The lower band (decade) is much smaller since the goal is to show activity in a decade. The upper band shows all articles in a given year. When the user clicks on an item bullet, the bibliographical information is presented (e.g., author, title, journal, etc.). If the user clicks on the "EE" link (electronic copy), the article content is presented in a separate frame.

Figure 1 shows the exploratory search interface in action for the query results about "compiler". The system retrieves all journal articles that contain "compiler" in the title and returns a hit list clustered by year. All the search results are anchored in the timeline. If more than one article falls within a year, the order is based on its relevance to the query. In this example, we can use the lower band to pick a decade. We can move to the 50s where the early papers on compilers where published or stay in the current year to see the latest publications. Say we are interested in papers from the 60s, because they were very influential. After selecting that particular decade, we can see an interesting number of articles including the one selected about GIER ALGOL.

In summary, the user can see all search results in a timeline, observe years of high activity (i.e., many document are relevant to the query terms in respective years), explore the items using both bands and then click to select a particular article. A separate frame shows the complete article as well as other information such as related people and the latest year of a citation.

This output of the search results is an intermediate representation, which is transformed to the timeline visualization format. This is useful in case one wants to use other timeline visualizations. The toolkit is flexible enough so it is also possible to show the same information in a vertical timeline.

Compared to other literature search systems (DBLP, Google Scholar, etc.), our prototype system allows the user to see the presence of articles in time as well as years of high activity. We believe this type of representation of high activity is crucial for domains where there is a concentration on the research or analysis part.

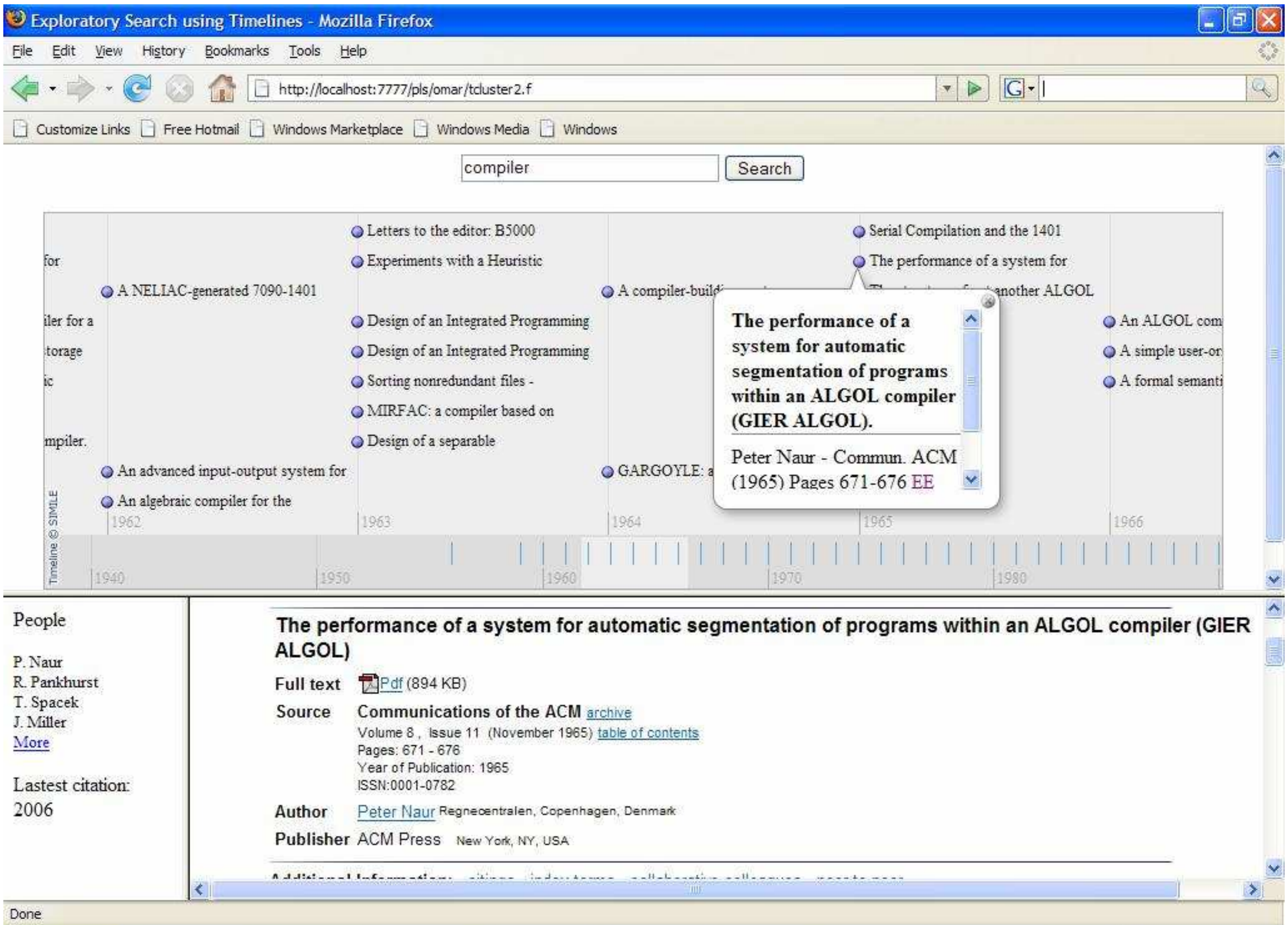

Figure 1. Exploring articles about compilers in a timeline, which is formed by two bands that represent decade and year respectively.

CONCLUSIONS

Adding time as part of exploratory search tasks can lead to interesting discoveries. We propose to use well-defined timelines as an alternative view for presenting search results that have rich temporal expressions or where the usage of time plays an important role.

Future work includes adding more implicit time information to the timeline visualization and to enhance its presentation with more metadata.

For a user study and evaluation, our initial plan is to assign a number of search tasks to participants and compare the results using our approach and existing bibliographic systems on the Web. For example, finding the first article on a topic or the period of time where most of the research for a particular sub-area has been done. We would also like to see the influence of time in search results rather than just popularity or last modified date attributes.

REFERENCES

1. Allen, R. "A Focus-Context Browser for Multiple Timelines", *JCDL* 2005, Denver, CO.

2. Alonso, O. and Gertz, M. "Clustering of Search Results using Temporal Attributes", *SIGIR* 2006, Seattle, WA.
3. DBLP Computer Science Bibliography. http://dblp.uni-trier.de
4. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. "Visualizing Tags over Time". *WWW* 2006, Edinburgh, UK.
5. Jackson, P. and Moulinier, I. *Natural Language Processing for Online Applications*. John Benjamins (2002).
6. Kleinberg, J. "Bursty and Hierarchical Structure in Streams" *8th ACM SIGKDD*, 2002.
7. Krishnan, A. and Jones, S. "TimeSpace: activity-based temporal visualization of personal information spaces", *Personal and Ubiquitous Computing* 9(1): 46-65 (2005).
8. Ringel, M., Cutrell, E., Dumais, S, and Horvitz, E. "Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores", *Proceedings of Interact 2003*, Zürich, Switzerland.
9. TimeLine http://simile.mit.edu/timeline/
10. TimeML http://timeml.org
11. White, R., Kules, K., Drucker, S., and schraefel m. (Eds). "Supporting Exploratory Search", *CACM*, Vol. 49, No. 4, April 2006.

SmartWeb: Display of Web Search Results in a Virtual City

Nicolas Bonnel[*]
IRISA / INSA
Campus universitaire de Beaulieu
35042 Rennes Cedex, France

ABSTRACT

While searching the web, the user is often confronted by a great number of results, generally displayed as an ordered list. Due to the limits of this approach, we propose to explore new visualizations of search results, as well as new types of interactions with the results to make their exploration more intuitive and efficient. Indeed, even if the relevance of Information Retrieval Systems depends on the retrieved results, the effectiveness of the results visualization represents an alternative way to improve the relevance for the user. So this paper deals more specially with 3D visualizations of web search results. We present an interface which uses the 3D metaphor of the city to display the results. This 3D virtual city is defined thanks to the X-VRML language. From a more general point of view, the main idea of this paper is the 3D visualization of web documents such as search results, news or RSS feeds. In this case, the use of 3D graphics is less conventional because users do not associate such documents (which are abstract and non-geographic data) to a 3D representation. Another important point is that the proposed approach is not space-driven because our goal is not to increase the visualization space but rather to exploit cognitive metaphors.

Author Keywords

3D Visualization of Search Results, 3D Metaphors, Web Mining

ACM Classification Keywords

H.5.2 [Information Systems]: Information Interfaces and Presentation — User Interfaces; H.3.3 [Information Systems]: Information Storage and Retrieval — Information Search and Retrieval; K.6.1 [Computing Methodologies]: Computer Graphics — Three-Dimensional Graphics and Realism

INTRODUCTION

Searching the web is one of the most frequent tasks, but often one of the most frustrating too. Search engines which are a way to represent the web to the users, are mainly used for web searches. However they are as easy to use as their results are difficult to interpret, which shows the web search contradiction. Indeed it becomes more and more difficult to extract the relevant information for a given search since available data on the World Wide Web is constantly increasing. The search engines return a number of results so great as it is necessary to search for new methods to visualize these results. These methods must be more adapted thanks to: a more relevant result organization, a richer visualization interface and an intuitive navigation in the result space.

This paper deals with the visualization of web search results. This step, still neglected in some Information Retrieval Systems, is becoming more and more important and essential. It can be considered as a solution for enriching the results. It is in fact complementary to the search process and is also a way to increase the result “relevance” for the user. If the result quality remains a major concern, the quality of the result restitution (organization and visualization) must be taken into account too. Without effective organization and visualization of the results, the user has to process manually the huge amount of results or refine the query in order to limit the results. This last solution can be compared to use a search engine for searching into the results! And these alternative solutions require efforts from the users.

Facing the increase of query results, it seems natural to want to **organize** and **visualize** them in an effective and adapted way. That explains the goal of the presented approach, which is to propose an user-friendly search interface enabling the user to quickly find the relevant information. The two main points to reach this goal are a good document organization and an effective visualization. Concerning these two aspects, our directions are a clustering method (the self-organizing maps) and a 3D visualization. The choice of a 3D visualization enables to exploit cognitive metaphors (such as spatial metaphors). 3D graphics offer new interaction possibilities too. So it enables us to bring a new point of view to the result visualization. However, considerable new problems appear, such as the navigation in such an environment.

This paper deals with 3D visualization of web search results. Indeed we present a 3D interface which displays the search results in a virtual city. And we also propose a discussion about our interface as well on the user’s perception and on the technical choices. So the paper is structured as follows. The next section presents our interface which is based on an unsupervised organization of documents and on a display of the results in a virtual city. Then, in following sections, a discussion about this interface is proposed and some related works are given. The last section allows us to conclude and gives an outlook on future work.

[*]e-mail: nicolas.bonnel@irisa.fr — This work has been carried out while working at the R&D Division of France Telecom.

3D VISUALIZATION OF ORGANIZED RESULTS

The visualization of search results has two main issues: organization (or clustering) of the results and their graphical representation. For the first one, the goal is to find an effective method which allows to group similar results together and to spatially organize the clusters. The second one is to find an effective visualization of the organized results. These two points are discussed in the following subsections but we mainly focus on the second point.

In the context of web search, documents are web pages returned by the query. In this paper, the organization of the results is only based on the textual information of documents. This information is exploited through a vectorial representation (word vectors) of the pages, which is frequently used in Information Retrieval field. The number of results to process must also be specified because it is crucial for the organization and visualization choices. A recent study [iProspect, 2006] shows that 88% of users will try a new search if they are not satisfied with the listings they find within the first 3 pages of results. However it would be too restrictive to only consider the first 30 results (10 results per page). Indeed this study has been done on search engines with linear results visualization (ordered lists) and users may want to see more results on 2D or 3D visualizations. Due to the lack of studies for 2D and 3D visualizations, the maximum number of results is currently fixed to the first hundred results, which is more than the 30 results with which the users are satisfied.

Then a mixed interface for visualizing organized results is proposed. This interface is composed of a 2D part (Java applet) and a 3D scene which represents the metaphor. A good definition of the metaphor word is: the realization of an association between graphical parameters of the presentation and information on the indexed documents.

SOM-based Organization of the Results

In this subsection, we briefly give some information about the organization of the results. For more details, see [Bonnel et al., 2006]. A good 3D visualization of search results needs to be able to organize these results in the output space according to their content. This clustering problem has been investigated in many previous works [Hearst and Pedersen, 1996] [Zamir and Etzioni, 1998]. However, this point is not discussed in this paper and we only give the solution used in our interface.

We use a self-organizing map (SOM) [Kohonen, 1995] which is an unsupervised method which enables to cluster **and** to project documents onto an output space (generally a 2D map). This method is used to carry out an on-the-fly clustering which can be considered as a post-retrieval document browsing technique. In other words, this clustering method organizes documents (or word vectors) on a map with predefined size, which guarantees a good use of space during the visualization. And the obtained organization has a neighborhood concept. Indeed two neighboring documents on the map have similar word vectors. Privileged application areas of the SOM are visualization and cluster analysis [Vesanto, 2002]. Self-organizing maps have already been used for textual data clustering such as in the WEBSOM project[1] [Kohonen et al., 2000].

Web search results are special textual documents due to their various size, content, vocabulary or reliability. So their organization implies a particular SOM application whose adaptation is described as follows. The SOM-based organization proposed can be divided in three main steps which are more precisely described in [Bonnel et al., 2006]. First the pre-processing step deals with the documents representation (word selection, word and document weights). Then the computation step corresponds to the execution of the modified batch SOM algorithm. This modified version takes the document weight into account and the input parameters are fixed to make this algorithm deterministic. The labels of the clusters are computed in this step too. Finally the post-processing step enables to group similar clusters (or map units) and then to obtain the various “topics” of the search. These abstraction levels are the result of a hierarchical agglomerative clustering applied on the map units.

Also the SOM-based method proposed is only based on word distribution and has the advantage to respect the “semantic” proximity of the data. It enables to have various levels of abstraction too. In our case, the weak number of results is not really a problem to compute the SOM, because the result organization is more important than the clustering itself. However one method which provides the best organization in all of the cases probably does not exist. That is why many organization methods must be defined in order to select the most adapted to each case.

Metaphor Context

The visualization interface proposed in this paper is based on the city metaphor. This choice is mainly justified by the cognitive aspect of this metaphor. And it seems adapted to a 3D environment contrary to the map metaphor where two dimensions are enough. A first version of this metaphor was developed and a user study was carried out on this metaphor. Upon the test results, the city metaphor has evolved. The new version will also be tested in order to know if the modifications answered what the users are waiting for and to identify new issues. Figure 1 gives an overview of the new metaphor whose explanations can be divided in four categories: basic elements, cluster representation, visualization and navigation. The last three categories are described in the following subsection. 3D graphics functionalities are poorly exploited because the visualization deals with abstract data, which explains that we do not want to have a realistic rendering of the 3D scene. Now the basic elements of the city metaphor are introduced. Each building of the city represents a web page, and the buildings are grouped by districts which are placed on the ground according to a grid. The building height represents the page relevance, which enables to quickly see the best classified pages according to this criterion. As our mapping choice for the relevance does not al-

[1] http://websom.hut.fi/websom

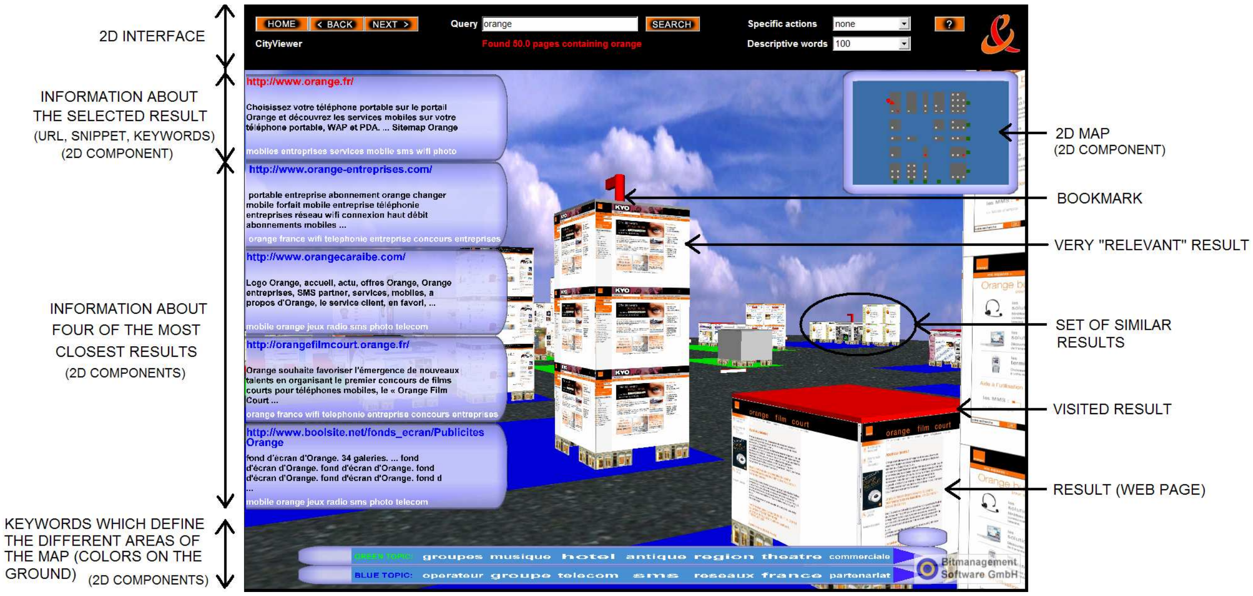

Figure 1. Visualization based on the city metaphor.

low to visually differentiate two successive ranks, an interval approach is adopted. So the first 10 ranks (*i.e.* the first result page in traditional search engines) are associated to a high building height. The following 20 ranks are represented by a medium building height. And the other ranks have a small building height. These interval choices are motivated by a study on user attitudes [iProspect, 2006].

Application of the City Metaphor on Self-organized Results

Cluster representation.

The districts are placed on the ground according to a 2D grid and each district represents a neuron (or map unit) of the self-organizing map. So this grid (which is square in our example) enables to map the search results classification obtained with the self-organizing algorithm presented in the previous section. Indeed the 2D grid on the ground is the same one as the SOM output grid. This organization has two interesting properties: the web pages (or building) unicity in the city and the “semantic” neighborhood between the different pages and between the different districts. So the documents of the same district are close to each other and two neighboring districts correspond to two topics as neighboring as possible. Colors were chosen for representing clusters defined by the hierarchical clustering on the neurons. Each district is associated with one of these colors (which are displayed on the ground). It enables us to show the main topics of the search. So the number of clusters is interface-oriented because each cluster is associated with one color.

Visualization.

The choice of a 3D interface to visualize the search results pleased the users. Broadly the 3D visualization is not a problem because it corresponds to our natural vision. Moreover this 3D metaphor enables us to give an overview of a great number of results. The building texture represents the document content, which enables us to quickly have an overview of the results when hanging around in the 3D environment. Highlighting a building allows the user to see information (URL, snippet, keywords) about the associated document and information about four neighboring documents which are obviously close to the chosen document. However, the user test reveals that the user seeks a compromise between the comfort and the effectiveness of the visualization more and more.

Navigation.

A user study was carried out and shows that the main drawback is the navigation in the city, which does not seem to be commonplace. Other navigation problems are mouse sensibility or the loss of reference marks in the city. So certain displacements toward strategic places of the 3D scene were simplified. To do that the 2D map of the scene was made interactive in order to be able to move to any district in only one click. This modification makes navigation more comfortable but it must be coupled with other approaches. So solutions must be found in order to make navigation more familiar for the user (like navigation in a 2D interface). A more constrained navigation (and thus less tiresome for the user) must be proposed to avoid the user getting lost in the 3D environment (see figure 2). Therefore the dimension increase makes navigation essential and especially more complex. And this problem is not obvious to solve. Another point concerns the used distance. In real life, to go from one district to another in a city, we follow the streets in a block like fashion, which corresponds to the Manhattan distance. This is the case in this metaphor only if the user uses the walking mode. However one interest of this metaphor is the use of the flying mode, and then the displacements are based on the Euclidean distance (the same as in the organization algorithm). So the user is not constrained to use the streets.

Figure 2. The user… lost in the 3D city?

DISCUSSION AND EVALUATION

Our visualization has the advantage to avoid supporting one topic for a query. Then the user has the possibility to choose their topic. Another point concerns data visualization which is strongly dependent on many criteria such as the results number and type, the search goal or the user category. So one single solution for data visualization probably does not exist. That is why an interesting characteristic consists of making the visualization adaptive. To do that many interfaces have to be defined in order to choose the most adapted according to the context.

The evaluation task is something very important in the visualization process. In our case, the evaluation needs to take the organization and the visualization metaphor into account. A user study was carried out and an extract of this user study can be found in [Bonnel et al., 2006]. This study was based on the well-known propositions of [Shneiderman, 1998] for evaluating graphical interfaces, and was mainly oriented to evaluate the visualization metaphor. However the interpretation of the results is hard because the users prefer the average marks and then avoid the extreme marks. The average mark for each question is generally higher than 3 (scale between 1 -bad- and 5 -good-), so it can be said that there is no drawback to use our approach for searching the web (compared to the existing ones). Another interesting remark concerns the fact that users are ready to use visualizations of clustered web pages. However a more relevant study and a comparison with other interfaces need to be carried out. For this purpose, we are currently working on the definition of evaluation criteria. Concerning the runtime evaluation, it seems that the interface needs almost 10 seconds to display 50 results. However it is important to highlight the following points: there is no code optimization, the organization is more time-consuming than the 3D visualization, and the X-VRML interpretor (*cf.* prototype section) is written in Java, which enables to be independent from the platform but does not optimize the runtime efficiency.

RELATED WORK

Many works have been done on search result visualization in the last few years. Some examples can be found in [Mann, 2002]. These visualizations as well as our approach can be located in the literature thanks to the taxonomy of search result visualization systems proposed in [Bonnel et al., 2005]. Among these various approaches, geographic metaphors (2D maps or 3D worlds) are often used because they can take advantage of the cognitive aspect.

In this paper, the proposed visualization has two main particularities : it is done on organized (or clustered) search results and it uses a 3D environment. Concerning the first point, the meta search engine KARTOO[2], the clustering engine CLUSTY[3] or the GROKKER[4] interface are examples which take inter-document similarities (or content-based links) into account. However all these examples use a linear or 2D interface and therefore they do not propose an overview of the results. Concerning the second point, there are some interfaces which propose a 3D visualization of documents. Moreover the third dimension is often used for replacing maps by 3D worlds such as landscapes [Boyack et al., 2002] or cities [Sparacino et al., 2002]. These 3D approaches are a good way to give the user an overview of the results. However these interfaces often propose a bad organization of documents or they are not user-friendly. Another work is the AVE method [Wiza et al., 2004] and its Periscope system which are the closest works to those ones presented in this paper. We have the use of mixed interfaces (3D scene and 2D interface) in common, or the use of many visualization metaphors which answer different goals. However the approach proposed in this paper takes the problem of data organization in a "semantic" point of view into account. Indeed it is not sufficient in the context of web search to only order the pages according to some low-level descriptors.

CONCLUSIONS AND OUTLOOKS

In this paper we present an effective method for organizing and visualizing search results. The organization is based on a self-organizing map which is adapted to the context of web search results. Concerning the visualization, we propose a new 3D approach based on a city metaphor which is very effective for representing organized documents. The graphical interface is dynamically generated, interactive and based on a compromise between the 3D scene and the 2D interface. With the proposed method, we provide the user with a three levels approach: low-level with document visualization, medium-level with neuron visualization (similar documents are grouped together), high-level with "topic" visualization (similar neurons are grouped together). The idea is to show the various topics of the query when going up in the hierarchy. A discussion about this interface is also proposed in this paper. It deals mainly with the interface evaluation and the 3D language used.

The information retrieval systems must be concerned with the quality of the results returned by the system (not discussed in this paper) but with their organization, visualization and interaction too. So our aim is always to enrich

[2] www.kartoo.com
[3] www.clusty.com
[4] www.grokker.com

the organization and visualization steps. Some interesting outlooks are the interaction improvement between the various hierarchical levels, the navigation simplification and the reduction of the response time. The general approach presented in this paper can be adapted to other exciting issues. For example, one idea is to compute a personalized representation of the web through web search results. The web searches of the user make the representation growing according to a criterion such as the visited pages or the first results. Another interesting issue is to use the 3D metaphor of the city for visualizing other data such as RSS feeds, bookmarks or news.

PROTOTYPE

The works presented in this paper have been integrated in a prototype which is very close to a classical search engine according to the query and the database. The aim is to provide the user with the best organization and visualization of their query results, without soliciting them during the process. A simplified overview of the prototype can be described as follows. On one hand the server side is composed by the database and the different interface models. Documents (web pages) are indexed in the database whose descriptors are those frequently used by the search engines such as URL, title, summary, snippet, word vectors. On the other hand the client side is composed by the Java applet (2D part of the interface) and the VRML browser (3D scene of the interface).

In our context, the main parts of the prototype are organization and visualization of the documents. And an interesting point concerns the prototype implementation which allows to have a certain independence between data organization and the selected metaphors for the visualization. The two input parameters of the visualization module of the prototype are: the previously organized results and an interface model (selected either by the user or automatically). The model which defines the visualization metaphor and the interactions used, is selected in a list of interface models. However it is possible for the user to easily create his own model and thus to personalize the visualization interface and the interactions. The models are expressed in X-VRML. During the query processing, the model is interpreted in order to dynamically produce the 2D interface and the VRML file which contains the 3D interface provided to the user. More details about this prototype can be found in [Bonnel et al., 2005].

The city metaphor is written in X-VRML which is an XML-based meta-language developed by France Telecom R&D and the Poznan University of Economics. This language has been created because VRML is too poor to satisfy our particular needs such as database queries, data organizations, interactions, iterations... So X-VRML enables to add many functionalities. Moreover X-VRML is extensible and it is possible to easily add new tags which have to be defined in Java. The use of X-VRML language is necessary because it seems that the current specification of X3D is not yet sufficient to answer all our problems. In our particular case, visualized data is non-geographic and abstract. So the 3D language used (X-VRML, X3D or others) has to take this point into account. This can be done with the integration of special interface-oriented functionalities (such as a good management of the text display in a 3D scene). The goal is to be able to propose 3D intuitive interfaces (or world) for visualizing abstract data.

REFERENCES

Bonnel, N., Cotarmanac'h, A., and Morin, A. (2005). Meaning Metaphor for Visualizing Search Results. In *Proc. of Int. Conf. on Information Visualisation*, pages 467–472. IEEE Computer Society.

Bonnel, N., Lemaire, V., Cotarmanac'h, A., and Morin, A. (2006). Effective Organization and Visualization of Web Search Results. In *Proc. of Int. EuroIMSA Conference*, pages 209–216. ACTA Press.

Boyack, K. W., Wylie, B. N., and Davidson, G. S. (2002). Domain Visualization Using VxInsight for Science and Technology Management. *JASIST*, 53(9):764–774.

Hearst, M. A. and Pedersen, J. O. (1996). Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. In *Proc. of ACM SIGIR*, pages 76–84. ACM.

iProspect (2006). iProspect Search Engine User Behavior Study. White paper.

Kohonen, T. (1995). *Self-Organizing Maps*. Springer.

Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., and Saarela, A. (2000). Self Organization of a Massive Document Collection. *IEEE Trans. on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery*, 11(3):574–585.

Mann, T. M. (2002). *Visualization of Search Results from the World Wide Web*. PhD Thesis, University of Konstanz, Germany.

Shneiderman, B. (1998). *Designing the User Interface*. Addison-Wesley.

Sparacino, F., Wren, C., Azarbayejani, A., and Pentland, A. (2002). Browsing 3-D spaces with 3-D vision: body-driven navigation through the Internet city. In *Proc. of Int. Symp. 3DPVT*, pages 224–233.

Vesanto, J. (2002). *Data Exploration Process Based on the Self-Organizing Map*. PhD Thesis, Helsinki University of Technology, Finland.

Wiza, W., Walczak, K., and Cellary, W. (2004). Periscope - A System for Adaptive 3D Visualization of Search Results. In *Proc. of Int. Web3D Conf.*, pages 29–40.

Zamir, O. and Etzioni, O. (1998). Web Document Clustering: A Feasibility Demonstration. In *Proc. of Int. ACM SIGIR Conf.*, pages 46–54. ACM Press.

ResultMaps: Search Result Visualization for Hierarchical Information Spaces

Edward C. Clarkson and James D. Foley
College of Computing and GVU Center, Georgia Institute of Technology
801 Atlantic Drive, Atlanta, GA 30332-0280
{edcclark, foley}@cc.gatech.edu

ABSTRACT
We present a treemap-based search result visualization system for hierarchical information spaces we call ResultMaps. ResultMaps are designed for the compact representation of search results that are embedded in a complex hierarchical classification. We also discuss our plans for using ResultMaps to investigate questions not addressed or raised by previous work in this area, such as the value of consistent representation of the information space or ResultMap's applicability to faceted browsing systems.

INTRODUCTION
People use hierarchy to abstract key concepts from groups of similar items and to help structure their thinking and reasoning. One common use of hierarchy is as a knowledge classification system, such as the Library of Congress classification scheme. As libraries have moved onto the Web, so have their use of such classifications. Moreover, many digital-only libraries have created new hierarchies for use with their contents. Naturally, these hierarchies are often a central component for browsing online repositories. But conversely, their use in repository search is generally limited. Conventional search relevancy reports allow users to limit searches to certain sections of a hierarchy and perhaps return hierarchy sections as query responses. But beyond these uses, hierarchies are underutilized in the context of repository search.

For the past three years, we have been developing digital libraries for two disciplines: Human-Centered Computing (HCC) [2] and Visual Analytics (VA). In addition to the goal of creating useful resources, we have made it a priority to use the libraries as platforms for related research [3]. Accordingly, we have developed *ResultMaps*, a treemap-based visualization system that we use in conjunction with our digital library search facilities (see Figure 1). Each node in the treemap represents a single document, and is colored by type. A ResultMap provides a glance-able overview of the entire document space, and makes both cluster and outlier detection straightforward. Other systems (e.g., [10]) have been developed with similar capabilities; we detail the features, differentiating factors, and future goals of our system after discussing motivating factors for this work.

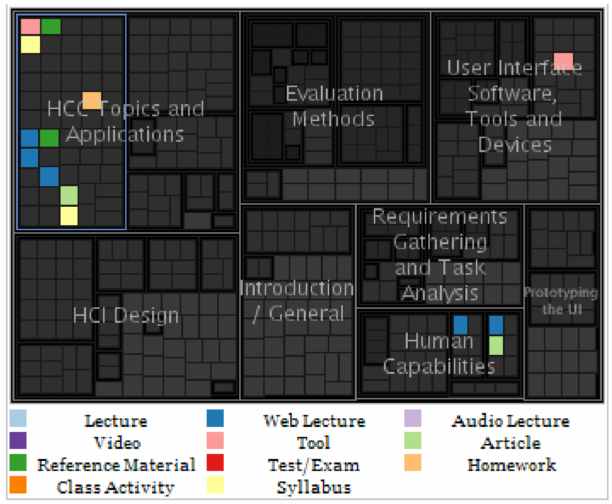

Figure 1 – A ResultMap (with key) for the query 'information visualization'. The 10 highlighted documents are in the search result set; color indicates document type.

RELATED WORK
There is a tremendous amount of work related to browsing or visualizing hierarchical data. An important class of browsing and exploration systems is that which employs *faceted metadata*, which organizes data characteristics according to a set of orthogonal attributes. The Flamenco search interface project is an elegant modern example of a faceted browsing system [13]. The traditional node-link tree structure is the most common and familiar representation for hierarchical data, but the space requirements become large even with a relatively small number of nodes. Various systems leverage its familiarity with various kinds of modifications (e.g., 3D, selective rendering) to the node-link paradigm. Another approach is to subdivide a 2D area, which is most prominently represented by the treemap technique [7].

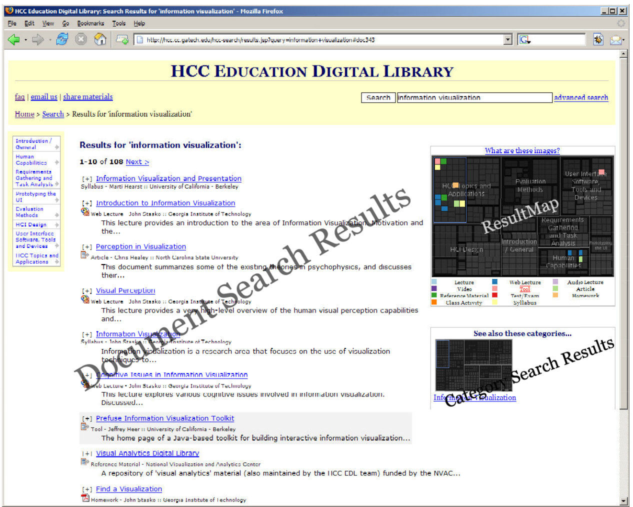

Figure 2 - A ResultMap-enhanced search result listing.

There are too many examples of digital library-specific visualization systems to mention exhaustively. We will discuss only a few relevant works: Klerx, Duval and Meire implemented a treemap-based visualization interface into the Ariadne repository [8]. However, there is no search facility within their tool and there is no application of their visualization to search results. Also, none of their systems were the subject of any form of evaluation. Lin used self-organizing maps (SOMs) to automatically associate and visualize search results [10], but technical limitations of the time made using an interactive implementation for evaluation prohibitive. Kules investigated a treemap-based overview of search results and developed design guidelines for such category-based overviews [9].

Microsoft Research's FacetMap system is also closely-related [12]. FacetMap is a treemap-inspired exploratory search tool for faceted data stores. Though conceptually similar, ResultMaps differ from the FacetMap system is several ways. First, the FacetMap layout algorithm does not show the entirety of a facet hierarchy (and trades space efficiency for distinguishability of the hierarchy). Second, selecting a facet value or adding a search term removes all non-matching documents from the view (losing the context of the global document space). Finally, its authors explicitly state that FacetMap is specifically targeted at the 'personal data store', which they note does not have the "limited interface requirements imposed by the simple keyword/list box paradigm." In contrast, ResultMaps always show results in the context of the full information space and is explicitly designed for online digital libraries, which are subject to that paradigm.

Motivation

We have built the ResultMap system as a vehicle for exploring research questions that previous systems have raised or not addressed. Lin's SOM-based approach was not tested as an interactive system, which prevented inquiry into many interesting questions. Kules himself raised questions such as: what is the value of consistency in these visualizations (what Kules calls the "visual substrate")? Similarly, the FacetMap system did not improve targeted or browsing task performance, and Kules did not report objective performance data (Lin did, but again not using an interactive system). Since the ResultMaps system augments the existing query paradigm, users should perform at least as well; a raw quantification of that benefit (if any) would be valuable, for example.

RESULTMAPS

We use ResultMaps to augment the standard query/result set response search engine paradigm. Figure 2 shows the results of searching for 'information visualization' in our HCC Education Digital Library. The central text area ('Document Search Results') shows a relevance-ordered list of document results. The bottom right area ('Category Search Results') shows sections of the hierarchy that the engine also returns as query matches. At the top right we show a ResultMap (the *summary ResultMap*) of the hits on the current page.

Figure 1 shows a larger version of the summary ResultMap. Each document is a leaf node; our implementation does not currently weight node sizes, though a characteristic such as document popularity or rating could be used to do so. We use the term *search result set* to refer to documents on the *current search results page* rather than all the documents returned by the search engine. In Figure 2, the search result set is the first 10 items (out of 108 total hits) returned by the engine.

Node color depends on whether a document is in the current search result set: if it is, its color is according to its document (not file) type. The key below the ResultMap shows the color/type correspondence. Those documents outside the result set are varying grayscale shades. Documents lower in the hierarchy are slightly darker to emphasize the hierarchy structure, a technique used in previous treemap applications (e.g., [4]).

A small frame delineates each category from its parent. Categories that are returned as hits in the category search results are outlined in light blue (see Figure 1 top left, which highlights the Information Visualization category). Each category search result (see Figure 2 bottom right) has a thumbnail version of the ResultMap with only that category highlighted, showing the correspondence between the highlight and the category hit. We confine the ResultMaps to a 350 x 233 pixel area to preserve sufficient space for the textual result listings.

We use a number of brushing interactions to link different parts of the search interface. When the mouse enters the border of any highlighted node, three events are triggered. First, a tooltip showing the document title appears at the mouse location. Second, the appropriate document type in the key is highlighted in red text with over and underlines. Third, the corresponding entry in the search result listing is highlighted via a background color change to light gray. In Figure 2, the mouse pointer is hovering over the document highlighted in pink at the upper right portion of the summary ResultMap. The corresponding search result in the list is one for the "Prefuse Information Visualization Toolkit," and its section of the result list has a light gray background (see the 3rd result from the bottom of the document search result section in Figure 2).

The effect of clicking on a highlighted ResultMap node is shown in Figure 3 (in this case, clicking on the "Prefuse Information Visualization Toolkit" square from above). Clicking scrolls the browser view to the corresponding entry in the result list, outlines it, and expands it. The expansion shows the entry's untruncated description, the categories of which it is a member, and a thumbnail ResultMap. The thumbnail is similar to those in the category listing—it highlights only that document, allowing the user to see the connection between the highlighted element(s) in the summary ResultMap and that result entry.

The ResultMap system has several nice properties. It works well with normal search result set sizes (i.e., 10-100 per page). It provides an overview of the entire document space, and a glance-able representation of the search results within that context. The ResultMaps make cluster and outlier detection straightforward. Outlier detection especially is of interest, since such items are usually very interesting or very uninteresting. Repeated exposure to the repository with every search query also encourages a familiarity with the repository beyond a user's immediate interest, which could encourage repeat and increased usage.

We should also note that the ResultMap system inherits the layout instability of the general treemap technique: the addition of new nodes to or repositioning of existing nodes within a hierarchy may result in drastic changes to node locations. The implication is that changes to the repository contents may result in significantly different visual layouts. Different treemap layout algorithms have different stability properties; we use a squarified layout, which is designed to create low aspect ratio rectangles but has a medium level of stability. Though this means it is possible for our ResultMaps to change significantly with additions to the repository, these changes occur between search sessions. Since the utility of our tool is the pattern of search hits only within the context of our ResultMap, we do not expect changes between sessions to have a major impact on users.

System Implementation

Our search engine implementation uses indices and relevance scores generated by the Lucene open-source search engine, and the Apache/Tomcat JSP engine to manipulate and display the result to the end-user. Lucene indexes all metadata fields, as well as the full text of any text-based document stored within the repository.

We use the prefuse infovis toolkit [6] to generate the ResultMap images, which are JPEG image files at present. We use an imagemap HTML element and JavaScript event handling to achieve the interactive brushing behavior within the summary ResultMap. We pre-generate and cache a number of page elements to improve performance: the appropriate imagemap coordinates for all nodes in the summary ResultMap and all ResultMap thumbnail images. This means we only have to dynamically create a single image for each search request (the summary ResultMap).

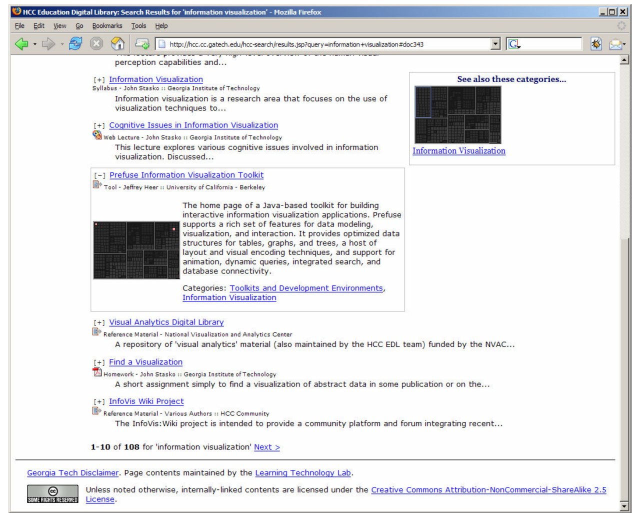

Figure 3 – The results of clicking on a highlighted node in a ResultMap: the browser scrolls to the corresponding entry in the search result list and expands the listing.

This implementation allows efficient interaction from the summary ResultMap *out to* the rest of the page (e.g., highlighting the search result text on a mouseover), but inhibits interaction from the rest of the result page *into* the ResultMap (e.g., highlighting a search result hit in the ResultMap image when hovering over an item in the textual result listing). This is a significant weakness, as Kules notes that tight coupling between representation and text is an important design principle [9]. Fortunately, this technical limitation is correctable: prototyping indicates DHTML elements are a feasible replacement for our image-based approach.

FUTURE RESEARCH FRAMEWORK

Our research framework involves several questions:

- Are there performance advantages for ResultMap users vs. traditional search users?
- Does using the ResultMap system increase users' broader understanding of a repository's contents?
- What representation techniques are feasible when the repository size exceeds the available pixel space?
- Does adding ResultMaps to a faceted browsing system increase its effectiveness?

The evaluation portion of this framework is not trivial. An infovis goal is to facilitate serendipitous knowledge discovery, which is difficult to evaluate—especially in traditional lab-based study settings. Since performance is still of interest, and other types of studies (e.g., longitudinal field studies) are often expensive and challenging to design and execute, lab studies supplemented with questionnaires and case studies are still popular approaches.

Evaluation

Our evaluation approach has two major components: controlled lab experiments and the analysis of field data. Concurrent with all of our studies are also qualitative, informal focus groups we commonly conduct to receive feedback on and inform future design decisions. Our formal studies seek to establish if the repeated contextual exposure in the ResultMap system results in better knowledge of that context (i.e., the repository as a whole). They also investigate if there are any performance differences between the ResultMap and the traditional systems.

System Enhancements

We are investigating three primary areas for enhancing the ResultMap system: scalability, technical implementation, and integration with a faceted browsing system. We know that treemaps scale relatively well [4], but our space constraints are more severe than many infovis applications. Our goal of leaving ample space for text result listings implies ResultMaps are limited to smaller sizes than many infovis systems. Our current summary ResultMap, for example, is 350x233 pixels (81,550 total). Since any document can appear in a search result list, a ResultMap must also highlight any item in a meaningful way—that is, so users can both detect that a node is highlighted and easily interact with it. There are a variety of methods for addressing this issue, including over-sampling pixels, node enlargement within and without treemap layout constraints, or interactive zooming. We seek a comparison of these methods within our specific usage context.

Faceted ResultMaps

Faceted browsing systems allow users to move easily between searching to browsing tasks. However, such systems focus on the items *remaining* in the search context—they do little to place the remaining items within the overall context of the entire repository. This problem is very similar to what motivated the ResultMap system—indeed, using ResultMaps as we have described is nothing more than using ResultMap with a single hierarchical facet. Accordingly, we can also apply ResultMaps to generalized faceted browsing, using a ResultMap for each facet—i.e., *faceted ResultMaps*.

The benefits of such an application are similar to those described previously: ResultMaps provide a glance-able, space efficient summary of the results for each facet. In addition, the ResultMaps can work together. Mouse-over on an item in one ResultMap can highlight the corresponding item in other ResultMaps, providing an easy way of linking the attributes of specific items (for example, whether an outlier in one facet is an outlier in other facets).

Likewise, the use of multiple summary ResultMaps provide information about how different facet selections affect others. For example, one facet may be highly related to another (e.g., the subject area of a document and the author who created it). Faceted ResultMaps could reveal such a connection by visual reasoning. If a user's selection of a category in one facet causes all of the results to disappear from a section of another facet, that change can be visually apparent from one ResultMap to the next. We can employ a device like color saturation difference to indicate whether a node is still in the result set or whether it was one of the items removed by the most recent constraint (or whether it was removed longer ago). Note that careful design of such indicators is critical, as is evaluation of how employing faceted ResultMaps affects the utility of the baseline system.

REFERENCES

1. Bruls, M., Huizing, K. and van Wijk, J. Squarified treemaps. In *Proc. of Joint Eurographics and IEEE TCVG Symposium on Visualization '00*, pp. 33-42.
2. Clarkson, E., Day, J. and Foley, J. An Educational Digital Library for Human-Centered Computing. In *CHI '06 Extended Abstracts*, pp. 646-651.
3. Clarkson, E. and Foley, J. Browsing Affordance Designs for the Human-Centered Computing Education Digital Library. In *Proc. of JCDL '06*, p. 361.
4. Fekete, J., Plaisant, C. Interactive Information Visualization of a Million Items. In *Proc. of InfoVis '02*, pp. 117-124.
5. Hearst, M. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In *Proceedings of CHI '95*, pp. 59-66.
6. Heer, J., Card, S. and Landay, J. prefuse: a toolkit for interactive information visualization. In *Proc. of CHI '05*, pp. 421-430.
7. Johnson, B. and Shneiderman, B. Tree-maps: A space filling approach to the visualization of hierarchical information structures. In *Proc. of IEEE Visualization '91*, pp. 284-291.
8. Klerkx, J., Duval, E. and Meire, M. Using information visualization for accessing learning object repositories. In *Proc. of IEEE International Conference on Information Visualization '04*, pp. 465 - 470.
9. Kules, W. *Supporting Exploratory Web Search with Meaningful and Stable Categorized Overviews*. Doctoral Dissertation, University of Maryland, 2006.
10. Lin, Xia. Map Displays for Information Retrieval. *Journal of the Amer. Soc. for Info. Science*. 48(1): pp. 40-54.
11. Shen, R., Vemuri, N., Fan, W., da S. Torres, R. and Fox, E. A. Exploring digital libraries: integrating browsing, searching, and visualization. In *Proc. of JCDL '06*, pp. 1-10.
12. Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G. and Tan, D. FacetMap: A Scalable Search and Browse Visualization. In *Proc. of InfoVis '06*, pp. 797-804.
13. Yee, P., Swearingen, K., Li, K. and Hearst, M. Faceted Metadata for Image Search and Browsing. In *Proc. of CHI '03*, pp. 401-408.

AVATAR: Beyond Keywords – Collaborative Information Extraction and Search

Eser Kandogan, Rajasekar Krishnamurty, Sriram Raghavan, Shivakumar Vaithyanathan, Huaiyu Zhu
IBM Almaden Research Center
650 Harry Rd.
San Jose, CA 95120, U.S.A
{eser,rajase,rsriram,vaithyan,huaiyu}@us.ibm.com

ABSTRACT
Existing search tools provide little support for information interaction activities beyond simple keyword-based source retrieval. In order to effectively support the full lifecycle of information interaction activities, tools need to support users in extracting, retrieving, manipulating, presenting, collaborating, and using information in their day-to-day activities. In this paper, we present our efforts in expanding the simple search interaction paradigm beyond source retrieval to collaborative information extraction and argue that we need to move towards an *information interaction language*, one that is simple and informal, yet supports users in the full lifecycle of information activities.

Author Keywords
Information Interaction, Search, Collaboration, Information Extraction.

INTRODUCTION
Searching is the most basic way of accessing information on the web. Nielson study points out that 88% of the web interactions start with a search [4]. Advances in search technologies transformed the way we think about information seeking. Advanced search algorithms, large scale systems, and simple search interaction paradigm played significant roles to bring quality, performance, and usability into the overall information interaction experience.

Over the years researchers developed various models of user-information interaction models that go beyond source-retrieval tasks [3,1,6,5]. For example, Marchionini's model also covers information extraction, analysis, and use [3]. In this paper we consider information interaction activities that begins with source retrieval and spans various activities including information extraction, manipulation, presentation, and integration/use. Source retrieval refers to finding relevant source documents in response to a user query. Information extraction, in our context, refers to examining and extracting valuable and relevant pieces of information from these documents, perhaps building a knowledge base of extracted information. Once extracted, information can be further manipulated to satisfy more complex queries that relate information in the knowledge base, perhaps aggregating information extracted from multiple source documents/queries. Bates' model also suggests that information needs are likely to be fulfilled not by one query/document but by a set of queries and documents [1]. Optionally, users may want to present information visually (or in other forms) to develop further insight and/or share with others. Next, information is utilized in some form in work activities, perhaps integrated to business processes (Figure 1.) Finally, search is typically an iterative process where the results of a query lead to further queries. As Shneiderman et al. suggest queries are likely to be refined and reformulated to satisfy information needs [6].

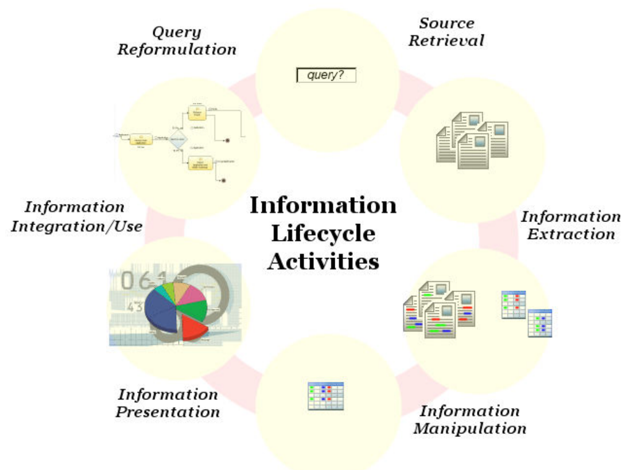

Figure 1. Information Lifecycle Activities.

Without a question, search tools have been very successful in answering user information needs for keyword-based source retrieval. The simple keyword textbox and search button interface significantly lowered barriers to entry allowing millions of users search and retrieve documents of interest. Without the complexity of formal query languages, logic expressions, users simply enter search terms into a textbox and find relevant information. We believe these search terms are in fact a simple and informal *information interaction language*. It is simple in that users do not have to specify any predicates indicating any kind of relationship among terms; it is informal as it does not necessarily conform to formal language syntax.

While in current tools information interaction language is limited to source-retrieval tasks we strongly believe it has to expand to cover the full lifecycle of information interaction lifecycle, capable of allowing users to express how information should be extracted, manipulated, presented, and integrated with processes. A 2002 study by Spink et al. points out that over the years the query formulation skills of the users have not improved significantly beyond typical queries containing just a few terms [7]. This is further evidence that cautions us that whatever information interaction languages the community may develop have to be simple and informal, yet able to satisfy needs in the full lifecycle of information interaction.

Below, we describe AVATAR, which demonstrates our efforts in expanding the simple search interaction language beyond source retrieval to collaborative information extraction and search.

AVATAR: COLLABORATIVE INFORMATION EXTRACTION AND SEARCH

AVATAR is an information extraction and search engine that exploits annotators to automatically tag information in documents. Annotations are performed by using high-precision information extraction techniques to extract facts (e.g. date, time, phone number), concepts (e.g. person, organization), and relationships (e.g. person's phone number) from text [2]. These facts, concepts, and relationships are represented and indexed in a structured data store such that queries can run efficiently.

In AVATAR, search queries are interpreted in the context of extracted information and converted into one or more precise queries over the structured store. For example, a query "steve phone" can utilize results from several annotators such as Person, PhoneNumber, PersonPhone, and PersonHomePage and yield multiple interpretations, including web pages that contain Steve's phone number, pages that mention Steve and a phone number, and pages authored by Steve where he mentions a phone number (Figure 2.) These interpretations in effect begin a dialogue between the system and the user to satisfy user's information needs.

AVATAR supports end-user development of annotators to facilitate collaborative information extraction (Figure 3). Our approach enables end-users to define annotator patterns for information extraction, just like they would do search, leveraging the same simple interaction paradigm. This is achieved by allowing users to simply type in patterns to detect facts, concepts, and relationships into a textbox (Figure 4.)

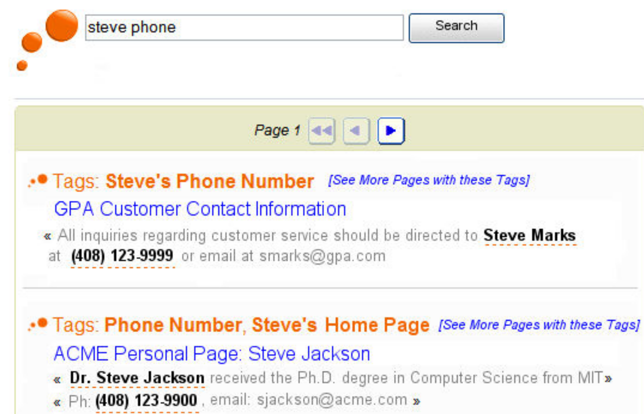

Figure 2. User query terms (e.g. "steve phone") are interpreted in the context of annotators (e.g. Person, PhoneNumber, PersonPhone, PersonHomePage), returning precise query results with various possible interpretations.

It is important to note that supporting collaborative end-user development of annotators is a must in information extraction for several reasons. First, not all users might know all the patterns to detect a piece of information. Only through collaboratively building on each other's patterns it is possible to achieve high recall rates. Second, a lot of the concepts are context and location sensitive. For example, phone number patterns vary significantly across different parts of the world. It is impractical to expect any system to have complete coverage. Finally, sufficiently large numbers of annotators are necessary to satisfy any realistic information needs. A work of this scale is only achievable through collaboration within communities.

In AVATAR, to define an new annotator, users can either start from scratch or build on existing annotators. For example, to define an annotator to tag cell phone numbers, one can use the built-in phone number annotator and contextual text around it, by the use of simple patterns such as, *cell:* PhoneNumber, *cell at* PhoneNumber, PhoneNumber *(cell)* , PhoneNumber *(mobile)* (Figure 4.)

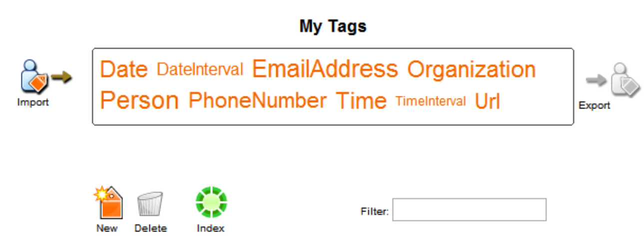

Figure 3. AVATAR annotators can automatically tag content using high-precision information extraction techniques.

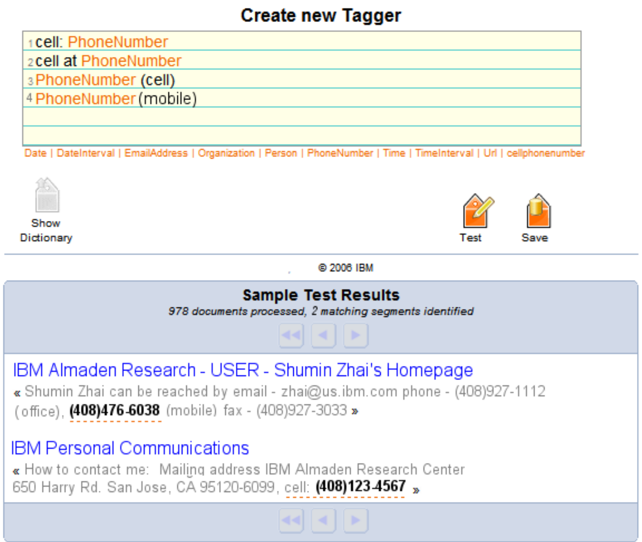

Figure 4. Users can collaboratively develop new annotators (e.g. CellPhoneNumber) building on existing annotators (e.g. PhoneNumber) through simple search like interface by defining patterns.

Annotator patterns can be of varying levels of complexity. Expert users can use regular expressions to detect concepts purely based on the format of the information. Novice users on the other hand can utilize existing annotators and AVATAR's natural language relaxation heuristics to define precise annotators. Typically, these patterns can build on a single annotator along with contextual text to specialize a concept, such as phone number vs. cell phone number. Other patterns can build on two annotators to define a relationship between two concepts, such as a person's phone number using Person and PhoneNumber annotators. In this case, the user would define the relationship often by specifying text that would occur between the concepts, such as Person *can be reached at* PhoneNumber. Users can also use dictionaries to define concepts. In this case, the dictionary would simply contain a collection of words associated with the defined concept. For example, to define an annotator to tag U.S. States, the user would build a dictionary of state names, such as "AZ", "CA", "NY", etc. and upload the dictionary to the system.

In order to support collaborative development, AVATAR also provides a facility to import and export annotator definitions. This way, users can share their annotators with their friends and colleagues, who can further build on these.

We expect three groups of AVATAR users: 1) Expert users who are technically savvy to build complex annotators from scratch, 2) Knowledgeable users, who are less technically savvy but can build on existing annotators, and 3) Novice users who never develop their annotators but can utilize shared annotators to improve the quality of their search. We believe by leveraging the simple search-like interaction model for annotator development we support users of variety of skills to perform effective information extraction.

A Use Case: Finding Product Prices

Let's go through a use case scenario to demonstrate the collaborative information extraction with AVATAR. In this use case several users each with different skills and backgrounds will build annotators leveraging each other's annotators.

Consider Aaron who works in the finance department, developing custom applications. As a programmer he is comfortable with regular expressions and often uses them in his applications. Aaron does his searches using AVATAR, and has developed a regular expression based annotator to detect numbers, called Number, with the regular expression \d+(\.\d{0,2})?. Using the Number annotator Aaron keeps track of various figures in the budget, which are typically distributed over many spreadsheets, documents, and internal web pages.

Michael is a product manager and uses AVATAR from time to time to search the web to check various blog sites to see what consumers are saying about his products. To do this more effectively he has developed a simple dictionary based annotator, Product, which detects product names such as Garmin Streetpilot, Nuvi, Forerunner, etc.

Mark on the other hand works in the human resource department and is interested in purchasing a GPS for his niece. While he doesn't have any software development skills he is a savvy web user. Using AVATAR, he built two simple annotators to help him find a good price for the product. One annotator is simply built on Aaron's Number annotator to identify prices, called Price using the pattern \$ Number to detect dollar signs preceding numbers. Another one, called ProductPrice, uses the Product and Price annotators to detect product prices using the following patterns:

1. Product *costs* Price
2. Product *is available at* Price
3. Product*:* Price
4. Product *Price:* Price

Using the ProductPrice annotator Mark simply does a web search to find retailers and their prices for various products he is interested in, all in the search results without even going to the retailer web site.

CONCLUSION

In order to effectively support the full lifecycle of information interaction activities, tools need to support users in extracting, retrieving, manipulating, presenting, collaborating, and using information in their day-to-day activities. Towards this end, we present AVATAR, an information extraction and search engine that exploits the familiar search interaction paradigm/language to also perform information extraction using annotators to tag content. AVATAR aims to leverage communities of knowledge users, who collaboratively define annotator patterns to extraction valuable information by building on each others work.

We believe that research on information interaction languages is a fruitful area, one that the HCI research community can contribute significantly. In our experience with AVATAR, the design of such an information interaction language necessitates going beyond usability and also addressing expressibility and scalability concerns. From this perspective, we found collaboration among HCI and Database researchers particularly rewarding.

REFERENCES

1. Bates, M. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review. Vol. 13, No. 5, pp. 407–424.
2. Kandogan, E., Krishnamurthy, R., Raghavan, S., Vaithyanathan, S., and Zhu, H. 2006. Avatar semantic search: a database approach to information retrieval. In *Proc. ACM SIGMOD International Conference on Management of Data* (Chicago, IL, USA, June 27 - 29, 2006). SIGMOD '06. pp. 790-792.
3. Marchionini, Gary N. (1995). *Information seeking in electronic environments.* Cambridge, Eng.: Cambridge University Press.
4. Nielsen, J. (2004) When Search Engines Become Answer Engines. www.useit.com/alertbox/20040816.html
5. Pirolli, P. & Card, S.K. (1999). Information Foraging. *Psychological Review*. APA, Vol. 106, No. 4, pp. 643–675.
6. Shneiderman, B., Byrd, D., & Croft, B. (1998). Sorting Out Search – A User-Interface Framework for Text Searches. Communication of the ACM. ACM Press, Vol. 41, No. 4, pp. 95–98.
7. Spink, A., Jansen, B., Wolfram, D., and Saracevic, T. (2002). From E-Sex to E-Commerce: Web Search Changes. IEEE Computer, Vol. 35, No. 3. IEEE Computer Society, 107—109.
8. White, R. W., Kules, B., Durcker, S., Schraefel, M. (2006). Communications of the ACM, Vol. 49. No 4.

combinFormation: Exploring Multiple Searches Together through the Mixed-Initiative Composition Space

Eunyee Koh
Interface Ecology Lab
Computer Science Department
Texas A&M University, 77843-3112, USA
eunyee@cs.tamu.edu

Andruid Kerne
Interface Ecology Lab
Computer Science Department
Texas A&M University, 77843-3112, USA
andruid@cs.tamu.edu

ABSTRACT

combinFormation enables people to search, browse, collect, and compose information on the Web, supporting information exploration and discovery. Through mixed-initiatives, a visual representation of news feeds, searches, and document crawls is assembled. A generative agent collects information, forms visual surrogates and continuously develops a composition that represents them as a connected collection. The composition space enables direct manipulation of the results that the agent is generating.

We present a usage scenario, in which a participant engages in mixed-initiative composition in order to explore topics in the day's news that she is particularly interested in. A composition develops iteratively, over time, using information from news feeds, and popular web searches and sites. The participant manipulates the information sources, and the visual design of the composition. In addition to the scenario, the mixed-initiative composition methodology is validated with data from a field study.

Author Keywords

Exploratory search, search interface, information discovery, compositional hypermedia, surrogates, visual interfaces, semantic web

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Sometimes, people lack the knowledge or contextual awareness to formulate queries that meet their information needs. As affordances, the text field and results list be inadequate for successful task completion. Users may require browsing, exploring, and learning, rather than just looking up search results. In such exploratory search tasks, the human formulates a tentative query and moves forward, by browsing retrieved information, seeking, and obtaining cues about where the next steps lie, and iterating [9].

In order to support exploratory search, Marchionini has suggested more active user involvement in the search interface, by blending querying and browsing strategies [7]. Cutrell and Dumais developed the Phlat interface for personal search, which combines keyword search and metadata browsing in a seamless manner [2]. This allows people to quickly and flexibly find information based on whatever they may remember about the information that they are looking for.

cOMBINFORMATION MIXED-INITIATIVES

Like the prior research, combinFormation integrates searching, browsing, and navigating. However, combinFormation is a mixed-initiative [3] exploratory search and creativity support tool that integrates processes of searching, browsing, collecting, mixing, organizing, and

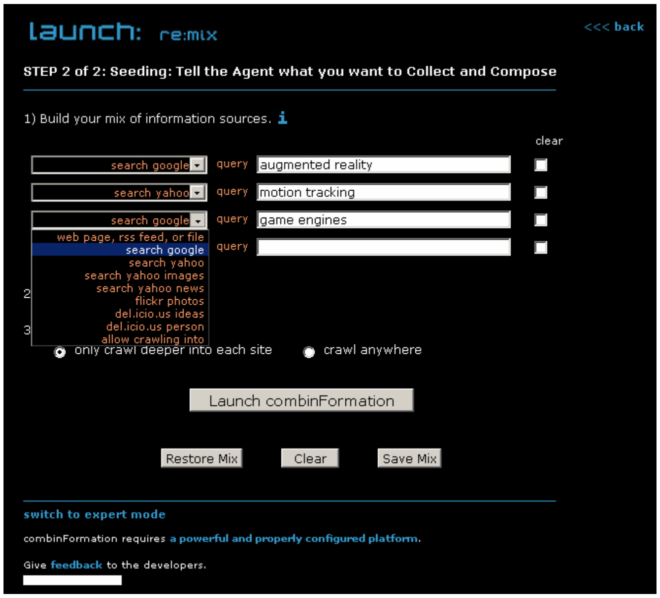

Figure 1. ***re:mix*** **interface: users can put together any number of searches (using multiple engines), sites, and feeds.**

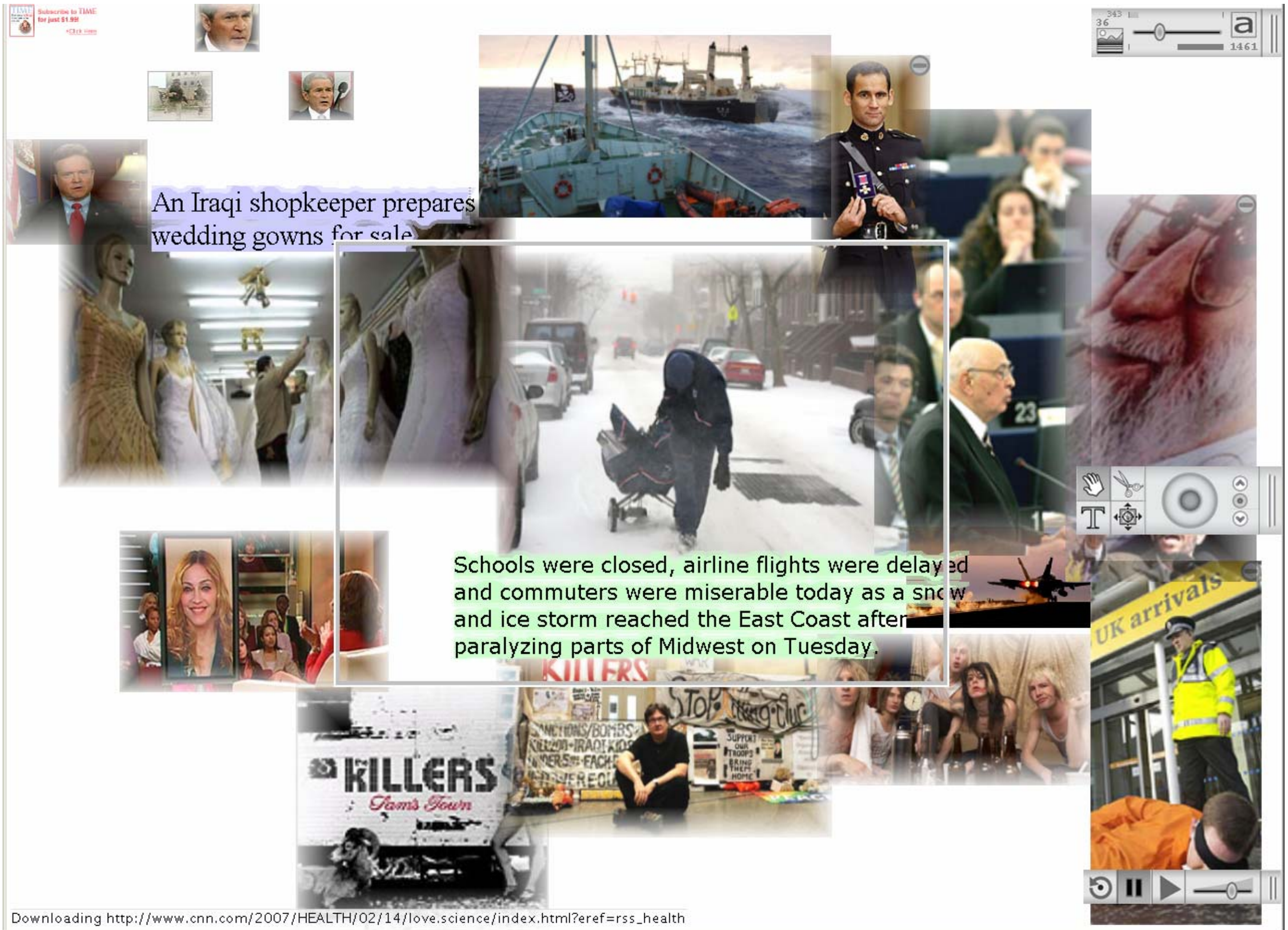

Figure 2. An early stage of the composition developed in the scenario. At this point, the seeds come from the *news re:collection*.

thinking about information [4, 5]. Digital information functions to support creative idea generation. By mixed-initiative, we mean that generative agents work in partnership with a human participant to collect relevant information, and represent the collection in the form a visual composition that changes over time. Images and text engage complementary cognitive subsystems. Each collection of information resources is represented as a connected whole. This promotes information discovery, the emergence of new ideas in the context of information.

combinFormation redesigns the searching and browsing interface to better support exploratory search. Use of combinFormation typically begins with launch, in which seeds are selected, to feed the generative agents. Each seed specifies an information source. The *re:mix* interface enables the participant to specify and assemble any number of seeds (see Figure 1). Each seed specifies either a document (HTML, PDF, or file system directory) location, an RSS feed, or a search query. For each search query, a particular search engine source may be selected, including Google, Yahoo, Yahoo Images, Yahoo News, del.icio.us, or Flickr. The *re:collections* interface enables the selection of previously curated seed collections. One such collection, *news*, assembles material from online news sources, including The New York Times and The BBC. Another, *popular buzzelicious*, combines the searches currently most popular on Yahoo with the items most tagged in del.icio.us.

combinFormation enables people to explore, compare, collect, and compose multiple information resources in a composition space. In the composition space, multiple seeds are mixed and visualized with image and text surrogates. Metadata details on demand are provided in-context [5]. Intermediate results can be directly manipulated by the user at any stage in the process.

Image and Text Surrogates

combinFormation represents search results with image and text clippings from documents. These clippings function as surrogates that the participant can use to navigate back to source web pages. Dual coding of results with image and text surrogates promotes cognition [8]. Each image surrogate is formed from a significant image from the document. Instead of using a summary, we build texts surrogate by extracting significant textual phrases. When a

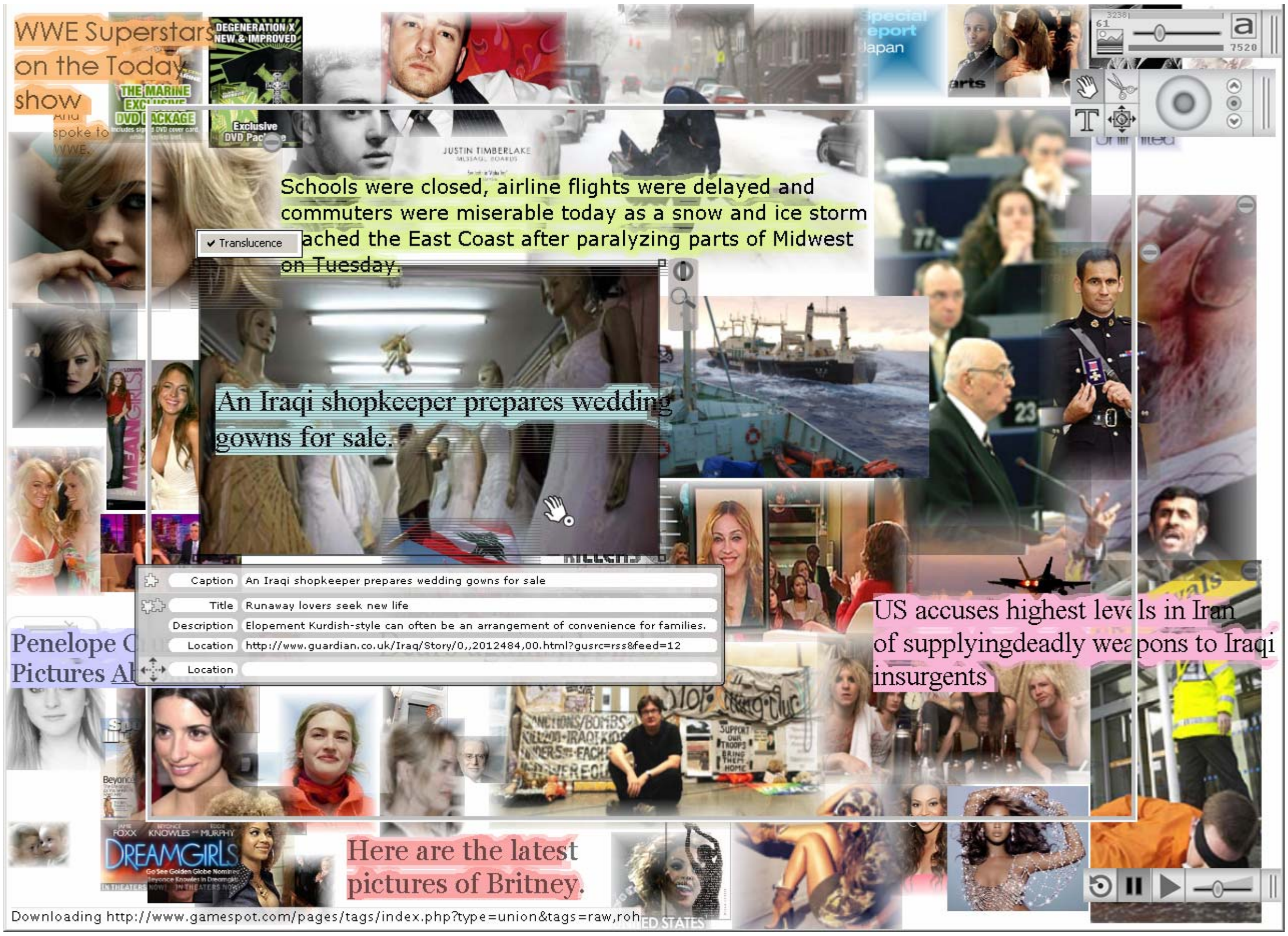

Figure 2. Later in the usage scenario, the participant has mixed the *popular buzzelicious* in with the *news*. In-context metadata details on demand are displayed below an element activated through brushing by simple mouse over.

surrogate is brushed with mouse over, the visual representations are augmented by metadata details on demand [1], such as the title of a document, and the caption of an image (see Figure 2). The details are provided in-context, to reduce cognitive load. The goal is to use surrogates to focus the representation of finer-grained ideas that lie within documents, as appropriate, in ways that reflect the intentions of document authors, and the needs of people collecting information. The surrogate serves to represent significant ideas from the document, in the context of the emerging collection and the associated task. The surrogate, in the composition space, also affords interest expression, which enables the participant to provide relevance feedback that affects the system's semantic model of the information and the participant's interests. This, in turn, affects the agent's on-going generation and exploration of the collection.

Composition Space

The composition space is a means for developing a representation of the collection as a connected whole. It is a mixed-initiative space, in which the agent and the human participant can work concurrently. Within the composition space, the human can directly position, resize, color, edit, and remove surrogates. S/he can also navigate from each surrogate in the composition space back to its hyperlink and source document, in a traditional web browser. A tape recorder transport enables the participant to pause the agents, taking more control of the composition. When the human finds information that s/he wishes to collect in a source document, s/he can click and drag to select it in the web browser, then drag it over to combinFormation, and drop it into the composition space. For the drag and drop operation to represent material that is selected in the source web page as surrogates in the composition, the program needs the context of the web address of the source container document. We developed a Firefox plug-in to pass this contextual metadata through drag and drop.

USAGE SCENARIO

A usage scenario was developed, with the simple exploratory search goal of exploring what is going on today in the world and on the internet, developing a focus on material of interest to a participant. The participant begins by selecting the *news* collection using the *re:collections* launch interface. As you can see in Figure 2, materials flowed into the composition on diverse topics, including the war in Iraq, a winter storm, and Madonna's appearance on the Oprah Winfrey Show. The participant decides that she is interested in simultaneously exploring the current most popular topics on the internet, so she returned to *re:collections*, while the composition was running, and selected *popular buzzelicious*. combinFormation obtained the 15 most popular Yahoo searches via XML, then issued the searches, and downloaded 15 documents for each search. Likewise, a feed specifying the most popular documents from del.icio.us was obtained, and these documents were downloaded. All the documents were processed to form image and text surrogates, which then became surrogate candidates for the composition. While this was happening, the participant used combinFormation's direct manipulation tools to reposition and resize surrogates of interest, and eliminate those that seemed uninteresting. A later stage of the exploration is presented in Figure 3.

DATA

Beyond this usage scenario, to validate the efficacy of the mixed-initiative composition methodology, we are conducting an on-going field study, in which combinFormation is being used on assignments in an undergraduate course on The Design Process [5]. The information discovery [6] assignments, which require the development of new inventions, require open-ended information exploration. Students used either combinFormation or Google and Word for information exploration. We investigated how students performed on their assignments, according to grades assigned by the teaching assistant. The ones using combinFormation performed better than those using Google and Word, and the results were statistically significant [5]. The TA also found that representations of collections assembled in the medium of composition of image and text surrogates are better than textual lists for understanding, developing ideas, and the communication of meaning.

Recently, student participants described the experience:

P1: When using this program, I learned that after I have all information gathered in one area, it is easier to play around and come up with something new. It takes in all the different ideas in our brains and just basically puts it on the table for us.

P199: The current and future application of ambient intelligence was very interesting especially as it applied to the medical field and to our everyday lives. It was something I did not know very much about, but there was a lot of info on. The really neat thing about the program was that instead of reading tons of long articles to find ideas or get concepts, it pulls quick quotes that allow you to connect lots of ideas.

CONCLUSION

combinFormation supports people in exploring, understanding and learning about information on the Web, by providing a mixed-initiative composition space to search, browse, compare, and compose information. The composition space promotes exploration and information discovery, by juxtaposing relevant surrogates. This provokes the human mind to consider alternative explanations about the relationships between connected elements. This paper identifies significant interface components to support exploratory search, which are search query mixing, image and text surrogates, the composition space, and in-context details on demand. combinFormation development is conducted through an on-going iterative design process, based on feedback from field and laboratory studies. We will continue to advance the interface, visualization, and semantic model to better support exploratory search and information discovery.

ACKNOWLEDGMENTS

Support is provided by NSF grant IIS-0633906.

REFERENCES

1. Ahlberg, C., Shneiderman, B., Visual Information Seeking: Tight Coupling of Dynamic Query Filters With Starfield Displays, *Proc of CHI 1994*, pp. 313-317
2. Cutrell, E., Dumais, S.T., Exploring Personal Information, Communications of the ACM, 49(4), 2006, 50-51.
3. Horvitz, E., Principles of Mixed-Initiative User Interfaces, *Proc CHI 1999*, pp. 159-166.
4. Interface Ecology Lab (2006), combinFormation, http://ecologylab.cs.tamu.edu/combinFormation/, last accessed 01/08/06.
5. Kerne, A., Koh, E., Dworaczyk, B., Mistrot, M.J., Choi, H., Smith, S.M., Graeber, R, Caruso, D., Webb, A., Hill, R., Albea, J., combinFormation: A Mixed-Initiative System for Representing Collections as Compositions of Image and Text Surrogates, *Proc JCDL 2006*, 11-20.
6. Kerne, A., Smith, S., The Information Discovery Framework, *Proc DIS 2004*, 357-360.
7. Marchionini, G., Exploratory Search: From Finding to Understanding, Communications of the ACM, 49(4), 2006, 41-46.
8. Pavio, A., Osapo, K., Concrete Image and Verbal Memory Codes, Journal of Experimental Psychology, 80(2), 279-285.
9. White, R.W., Kules, B., Drucker, S.M., Schraefel, M.C., Supporting Exploratory Search, Communications of the ACM, 49(4), 2006, 37-39.

Exploratory Tag-Based Search in Multiple Enterprise Domains with the Malibu Productivity Assistant*

Michael Muller, Werner Geyer, Casey Dugan, Beth Brownholtz, Eric Wilcox, David R Millen
IBM T.J. Watson Research
One Rogers Street, Cambridge, MA 02116
{ michael_muller, werner.geyer, cadugan, beth_brownholtz, eric_wilcox, david_r_millen}@us.ibm.com

ABSTRACT
The Malibu system provides peripheral search capability and awareness of activities, tasks, social-bookmark resources, and feeds to assist knowledge workers in their activity-centric work. We describe the experimental system, a usage scenario and some preliminary usage data.

INTRODUCTION
Knowledge workers integrate search activities with their other work. We describe an experimental enterprise search prototype that integrates with a user's on-going work at various levels:

- Plug-in architecture to run within an enterprise instant-messaging platform
- Peripheral specification and results-examination to maintain the user's current task context
- Tag-based search across multiple internal services that rely on collaborative tagging

New tools require new learning and new "destinations," in the sense of applications that must be installed, managed, and maintained. One aspect of search integration is to build our enhanced socially-based search facility as a plug-in within an existing collaborative framework, namely the IBM Sametime 7.5 platform. This approach provides a "gateway" from the collaborative application into a collaboration-based search space, and allows resource-sharing and metadata pooling across the applications and services accessed through that platform.

Knowledge workers frequently change tasks, either by choice or through interruptions [9]. With an increased number of tasks and task switches, it becomes more and more crucial to restore the context of their current task. Providing a peripheral display of both search specification and abbreviated search results should assist users to maintain their current context while accessing new and enhanced services.

Previous research on activity-centric collaboration (e.g. [1], [6], [8], [6], or [11]) supports knowledge workers with context switching and resource rediscovery by organizing and integrating resources, tools, and people around the computational concept of a work activity. Many of these approaches have in common that they provide some structure within which all records of an activity may be collectively located and (re)discovered. However, many relevant resources are not architected to be resident in such collections of objects and actions, and others must be discovered "in the moment," perhaps because they have only recently become available; discovery of relevant resource among dynamically-changing data stores is a major problem. Yet many of the resources are created and stored in known online places, and within known online communities. Increasingly, these resources may be related to one another through shared collaborative tags. Accessing shared tags among diverse resources and services is likely to be of value to knowledge workers who are trying to integrate dynamically-changing information across domains and constituencies.

We have developed the experimental Malibu system to assist knowledge workers in their activity-centric work, with the goal of improving or overcoming the limitations of the current activity management approaches. Malibu supports resource aggregation by providing fast access to information from different data sources; it finds relevant information across data sources using social tags (known from systems like flickr or delicious) and person information; it allows the user to rapidly switch and restore contexts; and it helps users manage tasks by unifying to-dos, activities, and reminders from various data sources.

MALIBU SYSTEM
Malibu runs as a desktop side bar ("Malibu Board") that either slides out when users hover with their mouse at the left or right side of the screen or remains sticky and always visible on the desktop. The Malibu Board contains a series of configurable views, each one displaying one data source (see Figure 1). Views are frequently updated. As such, Malibu provides peripheral access to and awareness of multiple data sources while the user focuses on her main work on the desktop. With this design, Malibu becomes an always-on companion that can display information that is contextually relevant to the current desktop activity of the user, draw attention to new and important events, and provide quick access to data sources that need attention.

---

* A different version of this position paper was submitted as a Work In Progress to CHI 2007.

The system currently supports the following views: tasks (A), activities (B) (from a collaborative task management system [6]), shared bookmarks (C) (from a social bookmarking system [10]), feeds (D) (RSS and ATOM) and people.[1] There is also no explicit people view since Malibu runs as an extension of a corporate instant messaging client. Note that similar to the Google side bar [7], Malibu can be extended with new views.

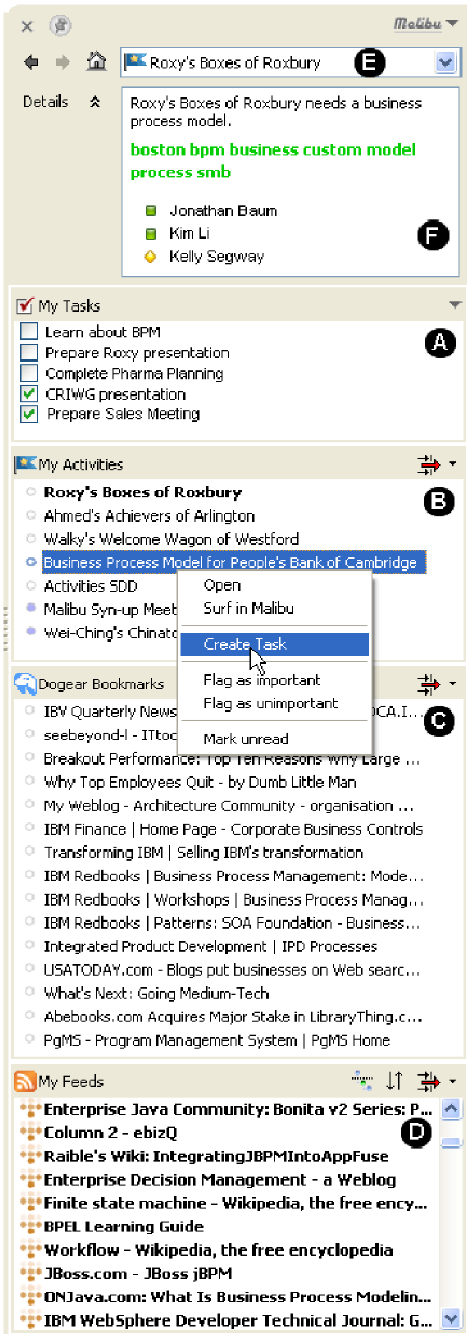

Figure 1. Malibu Board.

The key component in Malibu to support tag-based search is the Navigator located at the top of the Malibu Board. The "Navigator" (E) can be used to bring any item from any data source in Malibu into focus, i.e. a "focus" item is selected as pivot objects on which Malibu performs a tag-based search (called "Surf in Malibu" as shown in the context menu in Figure 1).[2] When a user surfs on an item, each Malibu view (window) receives a search event that contains all the tags of the surfed item. Each view uses this tag set to search for items with matching tags. The views reconfigure to display the search results related to the current surfed object; which becomes the current focus of the board. This kind of search allows users explore objects and their related content by navigating from one to another object instead of searching by keywords. However, each search can also be refined by selecting individual tags or manually adding key words to the query.

The details box (F) shows information about the current focus item, including the social tags from that focus item (which were used to perform the tag-based search in all of the windows). Similar to a browser, the navigation buttons and the history drop-down menu can be used to restore the search results of previous pivot objects. Details such as tags, description and author information can also be viewed in a slide-out window by clicking on items in any view.

The following usage scenario illustrates some core features and demonstrates how Malibu and its tag-based search can be integrated in a user's ongoing activity-centric work.

Kim, the lead member of a customer care team, opens the Malibu Board to start her day. The activities view (B) shows her activities filtered by importance. Kim sees that there is a new activity named ***Roxy's Boxes of Roxbury****. It is marked* ***unread*** *(bold) and shows up at the top of the activities list. Kim clicks on the activity and a details view slides out. Kim sees that the date of the activity was yesterday afternoon; and the creator was Kelly from sales. The description says "Roxy's Boxes of Roxbury needs a business process model". The details slide in and Kim creates two personal tasks in the task view (A)* ***Learn about BPM*** *and* ***Prepare Roxy presentation****.*

Kim right-clicks, chooses "Surf in Malibu" to pivot on the ***Roxy's Boxes of Roxbury*** *activity. Malibu reconfigures with the focus on that activity (E). The activities view shows some related activities as search results. The* ***Business Process Model for People's Bank of Cambridge*** *activity is similar to the new assignment and might contain useful*

---

[1] We decided to implement Malibu with those four data sources first, because of their popularity inside the company. Email and calendaring are important productivity tools too and may be added in a later version.

[2] In our data sources, tags are used to describe not only resources [12] but also activity records [6], selected feeds, and soon people [5]. Note that Malibu leverages user-generated tags from the data sources it aggregates, i.e. the tags are created in those data sources but not in Malibu.

modeling information. Kim right clicks on the ***People's Bank of Cambridge*** *activity and selects "Create Task". A new task linked to the activity shows up in the task view and Kim renames it to* ***Evaluate usability of BPM for Roxy's****.*

An incoming chat from Mary interrupts Kim. Mary is asking for the planning documents for the "Pharma" project. Kim right-clicks and surfs on Mary's name in the buddy list. Malibu reconfigures and shows all activities shared with Mary. The activity ***Pharma Customer Planning*** *shows up in the list. Kim had previously finished the planning document and from the desktop, Kim drags the document onto the* ***Pharma Customer Planning*** *activity to upload it. Kim replies to Mary in the chat, informing her that the document is complete and now available in the activity. Additionally, she checks off the* ***Complete Pharma planning*** *task. To recover from the interruption, Kim presses the back button. Malibu reconfigures into the pre-interruption state.*

Kim knows little about business process modeling in Roxy's business domain. She checks the Social-Bookmarks (C) and Feeds (D) views for related information. Kim notices a feed entry with a news article that could be interesting to the entire Roxy team. She drags and drops the entry onto the ***Roxy's Boxes of Roxbury*** *activity to share it with the team. Kim also sees an interesting shared bookmark referring to a BPM tutorial she would like to look at later. She drags and drops the bookmark onto the* ***Learn about BPM*** *task created previously. In the task view (A), Kim is reminded of the* ***Evaluate usability of BPM for Roxy's*** *task she created previously. She double-clicks on the task which opens up the activity the task is referring to. The Malibu Board slides in as Kim is focusing on the activity in the browser restoring the entire screen as her work area. After reviewing the activity, she decides to work on something else. She hovers with her mouse at the right side of the screen to trigger Malibu to slide out again and then presses the "home" button in Malibu to restore overall work awareness again, clearing the previous focus on Roxy's. The activity list shows all her activities filtered by importance.*

PRELIMINARY USAGE DATA

We have provided Malibu for internal download. As of December 14, 2006, 189 people had installed and run the Malibu client on a regular basis (defined by interaction with Malibu-mediated content on 50% of work days)[3]. In interviews with selected high volume users, we learned that people relied on Malibu to help them organize their daily work – e.g., "I use it, and keep it open and refer to it pretty consistently," and "[I] Live and breathe in an interrupt-driven environment. I use Malibu to return me to sanity," and "When it's not working, I get lost."

Users reported that they used Malibu to organize their tasks and activities. They particularly appreciated the ability to create their own, immediate priority-ordering of activities using Malibu's importance-flagging features – "The things [activities] I flag are the ones I tend to add content to… quicker access, I like [to see them] in the same place."

Users also reported problems. Some Malibu features were not fully-featured (e.g., the peripheral feed reader). Other features helped users, but engendered expectations that we did not fulfill (e.g., users could use drag-operations to create an object within an activity, and therefore expected to be able to create a new activity using the same action).

RELATED WORK

Researchers have studied the usefulness of side bars for peripheral awareness in Side Show [2]. Side Show evolved into a side bar for Microsoft Vista. Other similar products are DesktopSidebar [4], and Google Sidebar [7]. Watson [13] is a side-bar-like product that features automated search based on the active application on the desktop. Dashboard [3] is similar but shows all sorts of personal information related to the current desktop activity, e.g. emails, bookmarks, chats etc. Our work goes beyond these projects because it provides stronger connections among the different types of resources, through tagging. Specifically, we use tag-based search to find related resources, based on the use of tags in multiple internal social software services ([12]; see also [5]).

CONCLUSION

The Malibu Board provides peripheral access to and awareness of important data sources for knowledge workers. In contrast to other side bar prototypes and products, Malibu leverages social tag information to mash-up different data sources, and by doing so it can support the knowledge worker in finding relevant information to better support the current task.

Navigation around objects (on activities in particular) provides semantic search on a higher level than simple keyword search, and allows users to easily restore context as illustrated in our scenario, or explore relationships between otherwise disconnected data sources using multiple tags. We are currently developing a second version of Malibu that leverages the rich activity structure (using more than tag data) as the major component for finding information related to the current desktop activity of a user. To that end, we are also instrumenting Malibu with context sensors to detect what users are doing on their desktops. This information can be used by Malibu to identify the current task and proactively find related information.

---

[3] The 189 people were 14% of those who examined the experimental client when it was offered on an internal server. However, we note that only 46% of the people who examined the client were users of the Activity service (the primary intended user group). When we focused on Activities users, we found a slightly higher 17% of them used Malibu regularly.

ACKNOWLEDGMENTS
We thank all our anonymous users and the activities product team for their support.

REFERENCES
1. Bardram, J.E., Activity-based computing — Lessons learned and open issues. Position paper at ECSCW 2005 workshop, *Activity — From a theoretical to a computational construct*. Paris, September, 2005.
2. Cadiz; J.J., Venolia, G.D., Jancke; G., Gupta, A.; "Designing and deploying an information awareness interface," in: *Proc. ACM CSCW 2002*.
3. Dashboard, http://www.nat.org/dashboard/, retrieved 9/28/06.
4. DesktopSidebar, http://www.desktopsidebar.com/index.html, retrieved: 9/28/06.
5. Farrell, S.P., Lau, T., Muller, M.J., Ehrlich, K., & Wilcox, E., "Augmenting employee profiles via people-tagging," paper submitted to CHI 2007.
6. Geyer, W., Muller, M.J., Moore, M., Wilcox, E., Cheng, L., Brownholtz, B., Hill, C.R., Millen, D.R., "ActivityExplorer: Activity-Centric Collaboration from Research to Product," IBM Systems Journal – Special Issue on Business Collaboration, Vol. 45, No. 4 (2006), 713-738.
7. Google Desktop, http://desktop.google.com/, retrieved 9/28/06.
8. Kaptelinin, V., "UMEA: Translating Interaction Histories into Project Contexts," *Proc. ACM CHI 2003*
9. Mark, G., González, V., and Harris, J. No task left behind? Examining the nature of fragmented work. *Proc. CHI 2005*.
10. Millen, D.R., Feinberg, J., Kerr, B., "Dogear: Social Bookmarking in the Enterprise," in: *Proc CHI 2006*.
11. Muller, M.J., Geyer, W., Brownholtz, B., Wilcox, E., and Millen, D.R. One-hundred days in an activity-centric collaboration environment based on shared objects. *Proc. CHI 2004*.
12. Muller, M.J., "Patterns of tag usage across four diverse enterprise tagging services." Plenary paper to be presented at HCIC 2007.
13. Watson, http://www.intellext.com/, retrieved 09/28/06.
14. Whittaker, S. Supporting collaborative task management in email. *Human Computer Interaction 20*, 1&2 (2005), 49-88.

Challenges and Opportunities for Visual Exploratory Search in User-generated Media Content Collections

Jasminko Novak
University of Zurich, Institute of Informatics
Information Management Group
Collaborative Technologies & Social Media Lab
Binzmühlestr. 14, CH- 8050 Zurich
novak@ifi.unizh.ch
+41-44-635-7137

ABSTRACT

This workshop contribution presents experiences from our work on supporting exploratory search tasks by using visualization to augment the search process and information sensemaking in a specific domain of cross-community knowledge exchange. We then discuss how the lessons learned here are related and can be transferred to supporting exploratory access to user-generated media collections. The first part presents three main results of our recent work: 1) a concrete visual information interface for explorative, multi-perspective access to community information spaces, 2) an evaluation framework for measuring search performance in terms of user-centered knowledge construction effects and 3) the results of an empirical laboratory study applying this framework to compare the use of a visual information interface (Knowledge Explorer) and a traditional search system (Google). The second part discusses how these experiences are linked to the currently exploding phenomenon of personal and participatory media where huge amounts of information are increasingly produced and shared between users, in a number of different media (audio, photo, video) with no professional mediators and little or no structuring. We argue that user-generated content creation and sharing exhibits structural properties typical for explorative rather than goal-directed search needs. This requires new kinds of search tools integrating collaborative techniques such as community-based content aggregation and filtering with visual exploration interfaces.

Author Keywords
Exploratory search, visual information interfaces, user-generated content, participatory media.

VISUAL EXPLORATORY SEARCH FOR SENSEMAKING IN UNFAMILIAR INFORMATION SPACES

Exploratory search interfaces supported by visual techniques are frequently found to perform worse than traditional keyword search with list-based search result presentation and navigation [3],[4]. In contrast, the results of our work suggest that such seemingly general weakness of exploration and visual search may be due to two main reasons: 1) task-bias i.e. generic comparison vs. domain-specific evaluations (explorative strategies may be more suitable for specific classes of application or task domains than others) and 2) too narrow evaluation metrics largely focusing only on document precision and recall, not considering the broader context of the search process and its aims (such as the topical overview of the collection or the learning effect in understanding previously unfamiliar domains).

We have investigated the development and use of visual and exploratory search techniques for specific classes of applications which require intrinsically explorative access by their nature. A concrete example is the development of methods, evaluation frameworks and tools to support a combination of explorative and goal-directed search and sensemaking in unfamiliar community spaces as part of the process of cross-community knowledge exchange [7, 8]. Seeking information in unfamiliar communities is typically motivated by ill-structured problems which cannot be solved within the user's familiar knowledge context (e.g. in innovation processes, strategy making, interdisciplinary research or new product development). In such contexts, the information need is very ambiguous and difficult to resolve through goal-directed search. They contain tasks of inherently explorative nature and require learning about potentially relevant but unfamiliar topics and knowledge domains in order to structure the problem and identify gaps and knowledge needed for a solution, in the first place.

As a result, rather than as an information retrieval problem, in such situations the information seeking process is more correctly viewed as a sensemaking and knowledge construction process [9, 10]: an activity in which users look both for relevant information as well as for appropriate contextualization structures in order to interpret the meaning (make sense) of unfamiliar information and its relationship to the task at hand. We have addressed this problem by developing a method and an information interface for visual exploration and multi-perspective access to unfamiliar information spaces (the Knowledge Explorer) as well as an evaluation framework for capturing user-centered search performance in terms of knowledge construction effects rather than classical document retrieval measures (document precision & recall).

The Knowledge Explorer is a multi-view visual information interface enabling the elicitation and use of personal and shared concept structures for the structuring, navigation and exploration of community information spaces from different points of view. Originally, it has been developed to support cross-community knowledge exchange mediated through information access in unfamiliar community spaces. As this application domain is characterized by highly ill-defined information needs the information seeking process is highly explorative in nature.

To address the needs of such situations the Knowledge Explorer supports a combination of exploratory visual search with goal-directed search queries and concept-based navigation (Fig. 1). It employs a combination of 2D visual document maps providing overviews of inherent semantic structure of document collections and 1D concept maps (folder trees) reflecting conceptual structures of individuals and communities of users. The document maps are based on a combination of text or multimedia-feature analysis and a self-organized neural network which groups items into clusters of semantically related content while preserving global inter-item similarity relationships [5]. The personal and shared structures are elicited implicitly from users' bookmarking patterns and can be applied to dynamically restructure or filter a document collection, based on a specific personal or community point of view. Document Maps and Concept Maps of an unfamiliar community allow non-members to gain a quick overview of the community knowledge structure: the main topics, documents, concepts and relationships between them. Instead of displaying community maps, using personal maps of individual users presents only a specific portion of documents which both reflect personal knowledge of the map author and are relevant for a specific information need. Similarly, applying a personal map to the information space of an unfamiliar community will classify unfamiliar documents into thematic clusters defined by the map author.

The developed evaluation framework extends document precision and recall with measures for the quality of topical structuring of user search results and the users' learning effect of the previously unfamiliar collections as a result of the search process. An empirical laboratory experiment suggests that this framework is well-suited to measure effects of information consumption in the search process and can compensate for the bias of pure document retrieval measures [7]. The results of the experiment applying this framework to compare the use of a visual information interface (Knowledge Explorer) and a traditional search system (Google) suggest better performance of visual exploratory search in contexts characterized by ill-defined information needs and terminology problems and requiring a topical understanding of an unfamiliar information space [7, 8].

A significant difference of our solution to a number of other information and knowledge visualisation systems is that the latter focus on visualising data patterns, rather than on personal and shared knowledge structures of human users, and are conceived as tools for specialized analysis tasks (e.g. data mining, analysis of specialized document collections). In contrast, the Knowledge Explorer enables

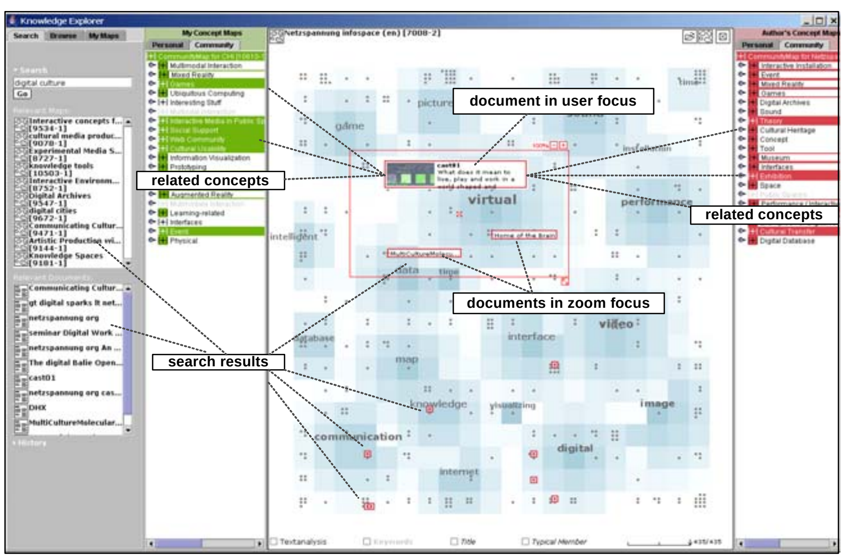

Fig. 1. The Knowledge Explorer interface

the elicitation and visualisation of implicit knowledge structures of human users and focuses on its application to information exploration in unfamiliar domains. Thus, it is intended as a powerful and yet simple to use visual information access tool for normal users rather than specialized analysts. This poses special requirements on its design, requiring real-time performance, ease of use and interactive visualisation adapted to the needs of typical information access tasks in unfamiliar domains (cf. information sensemaking). To achieve this, knowledge elicitation and knowledge discovery methods have been tightly integrated with multi-view visualization and navigation techniques in an intuitive user-centred interface. The development of an appropriate interface and interaction design has been informed both by insights from well-known information access task models (e.g. [2], [12]) and search process models (e.g. [1], [11]) as well as by studies on knowledge construction during information seeking in unfamiliar domains (e.g. [9], [10]). In this way the vast amount of experience on developing visual information interfaces could be productively applied, while considering the particular needs of the specific exploratory search requirements of the cross-community application domain.

THE CHALLENGE OF USER-GENERATED MEDIA CONTENT

The focus on community information spaces links this work to the current phenomenon of large scale user participation in creation and sharing of content. Huge amounts of information are increasingly produced and shared between users, in a number of different media (audio, photo, video) with no professional mediators and little or no structuring. The currently most dominant form of structuring in this context is collaborative tagging (folksonomies) which exhibits a great heterogeneity and idiosyncracy of semantic descriptors and requires large-scale user communities in order to produce usable terminological aggregations. At the same time, personal media use and sharing encompasses a number of different motivations and modes of use. Especially the "gratification mode" is increasingly superseding the more traditionally investigated information and communication modes. In addition, end-user content creation and use have been increasingly linked to shared social contexts such as computer-supported social networks and online communities (e.g. mySpace, YouTube, Wikitravel). We argue that this change in the modality of information access together with continuing growth of personal and shared media collections fundamentally changes user search patterns and requirements for new search interfaces. The gratification mode of (multimedia) information consumption implies that the search scope is much more explorative than goal-directed, as the information need becomes even more ill-defined and difficult to express than it is commonly observed (e.g. how to formulate search terms for "nice picture", "cool music"?). As a result, participatory media platforms tend to combine tag-based search with browsing and filtering through large collections based on aggregated patterns of user feedback (implicit or explicit) such as rating, tag frequency or number of user views (e.g. Flickr, YouTube). While this introduces some simple and easily usable structuring it also results in users being presented only with the tip of the iceberg. Since the largest part of the collection follows the long tail pattern (many small portions visited or rated by many small groups of users), the majority of items not appearing in the "top ten" remain practically invisible and very hard to find.

Furthermore, the gratification mode also changes the nature of relevance of retrieved information (e.g more relevant are items which are "more fun to see") which makes the whole notions of precision and recall difficult to define. This implies that applying existing metrics to evaluate the suitability of existing search tools for such collections and task-domains is also becoming increasingly difficult if not downright inappropriate.

Hence, personal and shared media collections together with the phenomena of user-generated content seem to exhibit structural properties which characterize more explorative than goal-directed search patterns: the information need is extremely vague and difficult to describe, the relevance is difficult to determine, the collection is unfamiliar to the user, highly amorphous and must be structured to allow access from many different, context-dependent and often very personal points of view (e.g. moods).

OPPORTUNITIES FOR EXPLORATORY SEARCH INTERFACES

Such structural characteristics rather closely resemble our previously described work in the application domain of cross-community information seeking. The results of this work suggest that in such situations a combination of visual explorative search with goal-directed search queries and a collaborative elicitation of personal and community-based information structures can provide valuable support to the users and may outperform traditional search interfaces, both on objective measures and subjective user satisfaction [7, 8]. Our results also suggest that accessing and sharing personal points of view of unfamiliar collections can be more rewarding and effective than exploring and navigating the entire collaborative structure [8]. And we also discovered how "personality matters": the possibility to personalize the visual presentation of one's own collections appeals to users and is quoted as an important incentive for sharing i.e. "presenting" the personal collection to others [7]. In summary, examining the nature of current phenomena of user-generated content and personal media sharing, we find it striking how direct the resemblance of structural properties and the relationship to our experiences with exploratory search in cross-community information seeking seems. This leads us to believe that there is great potential in investigating how the existing body of experience in exploratory search, visual support and community-based content aggregation and filtering can be brought together in order to create new kinds of explorative search interfaces, especially for user-generated multimedia content collections. More specifically, we propose to extend existing social bookmarking practices with richer means for expressing personal views and creating visual representations of multimedia object collections, where spatial arrangement carries semantic meaning reflecting a user's point of view (e.g. playlists vs. music maps).

As argued above, our experiences with facilitating information access through sharing personal visual document maps suggests that users are willing to invest effort in creating richer representations of their personal information collections in certain conditions - such as when the maps are aimed at being shared with others (self-presentation, reputation and hedonic factors) or when gratification is the primary motivation of information use.

As demonstrated in our previous work, user bookmarking patterns can be used to construct personalized views of information collections presented in form of visual overviews which map similar items close to each other, based on examples previously classified by the user. Similarity measures can also be defined based on document collocation patterns (e.g. the dice coefficient [6]) by communities of users (which have declared similar interests or opted for explicit community membership). This is a well-known technique which can avoid the problem of high idiosyncracy of user tagging patterns and actually be used to infer relationships between different classes of semantic descriptors for mutually related items. As a side-effect, tags signifying different but possibly related contexts of use (e.g. "easy listening" and "jazz" used for the same music) can be put in relation to each other.

Accordingly, user-induced "personalized media maps" can then be used as visual semantic templates for guiding exploration of otherwise unstructured collections, from different personal points of view. They can function as personalized templates for contextualizing keyword searches or conducting implicit search through "filtering by example" in situations where the information need cannot be easily or unambiguously expressed. And they can be aggregated across a group of related users (e.g. members of a community or a specific thematic social network) to provide visual means for browsing areas of the collection corresponding to shared interests of a group of users.

Finally, since they represent rich semantic contexts, such maps can be used to contextualize keyword-based search results: identifying the most relevant personal collection for a given search query provides a visual context within which the query results can be presented to the user (classified and spatially positioned into topical areas defined by the map author). In this way, related items can be discovered by visual exploration even when they greatly mismatch the original query: either by serendipitous browsing or by being spatially collocated with some of the retrieved documents but based on other attributes not targeted by the query.

Thus, by facilitating the creation and sharing of user-induced visual semantic templates we could overcome the limitation of current solutions for explorative access to user-generated media content collections which tend to provide rather restricted "tip of the iceberg" views. Instead, combining explorative visualization with personalized classification and community filtering seems a promising way for enabling rich visual exploration of user media collections based on different granularities of access (from personal to community level) and different facets of contextualization (e.g. different maps representing different selection criteria, based on varied needs and use contexts).

REFERENCES

1. Bates, M.J. The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5):407.424, 1989
2. Belkin, N.J., Stein, A., Thiel, U.: Cases, Scripts and Information Seeking Strategies. On the Design of Interactive Information Retrieval Systems. Expert Systems with Applications, Vol. 9(3): 379--395, 1995.
3. Chen, C., Yu, Y. Empirical studies of information visualisation: a meta-analysis, Int. Journal of Human-Computer Studies, Vol. 53,No.5, 851-866, 2000
4. Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces Sebrechts, Cugini, Laskowski, Vasilakis and Miller, Proceedings of SIGIR 99, Berkeley, CA, 1999.
5. Kohonen, T., Kaski S., et al.: Self Organization of a Massive Document Collection, IEEE Transactions on Neural Networks, Vol. 11, No. 3, May 2000.
6. McGill, M.et al.:An evaluation of factors affecting document ranking by IR Systems," Project Report. Syracuse University, School of Information Studies, Syracuse (NY), USA, 1979
7. Novak, J. Collaborative Knowledge Visualisation for Cross-Community Knowledge Exchange, PhD Thesis, University of Duisburg-Essen, Germany, 2006
8. Novak, J. Helping Knowledge Cross Boundaries: Using Knowledge Visualization to Support Cross-Community Sensemaking, Proc. HICSS-40, Hawaii, 2007
9. Qu, Y., Furnas, G.: Sources of Structure in Sensemaking. Proc. of ACM CHI 2005, Extended Abstracts and Applications, Portland, OR, USA, 2005
10. Russell, D. M., Stefik, M. J., Pirolli, P., Card, S. K.: The Cost Structure of Sensemaking. Proc. of ACM INTERCHI'93, pp. 269-276, 1993
11. Saracevic, T. The stratified model of information retrieval interaction: Extension and applications. In Proceedings of the American Society for Information Science 34 (1997), 313–327
12. Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualisation. Technical Report 96-66, Institute for Systems Research, University of Maryland, 1996

Encouraging Exploration with Elroy: A New User Experience for Web Search

Daniel E. Rose
A9.com, Inc.
130 Lytton Ave., Palo Alto, CA 94301 USA
danrose@acm.org

Swati Raju
Yahoo! Inc.
701 First Ave., Sunnyvale, CA 94089 USA
swati@yahoo-inc.com

ABSTRACT
Most search engines support a very limited interaction paradigm that reflects design constraints from the early days of the web. In this paper, we describe an experimental system called Elroy that is an attempt to explore an alternate user experience based on design attributes not found in conventional web search engines. In particular, Elroy has several features that encourage exploration by breaking down barriers between searching and browsing.

INTRODUCTION
When the first web search engines were introduced in 1995, they provided an extremely useful resource to users who wanted to find information among the of millions of pages available on the World Wide Web. But the search engines of 1995 also reflected the technical constraints of their time as well as the interests and experience of their designers. CPUs and modems were relatively slow, so web search engines passed the minimum amount of data between client and server and used simple text interfaces. The early web search engine creators were often computer scientists with little experience in information retrieval, library science, or human-computer interaction. Considerations of human information-seeking behavior or support for different search strategies were largely absent from these systems, which instead emphasized speed and corpus size. In particular, no attempt was made to encourage users to iterate and explore. Rather, search engines seemed to be designed to require users to know exactly what they wanted in advance and be able to formulate the perfect query to find it.

While these decisions may have been understandable in 1995, it is surprising to see that so little has changed more than a decade later. If an AltaVista user from 1995 could be transported to the present, he might marvel at many of the changes in technology and be overwhelmed by the rich-media, highly interactive web services of today. But sit him down in front of today's Google, and he'd be right at home.

Is it possible that the lack of change in a decade of search engine interfaces is due to the fact that the existing paradigm represents the pinnacle of evolution for the search user experience? If this were true, then it should be the case that search engines of today possess all the attributes we could imagine that make a good search user experience.

Results as relevant as possible
Short delay between query and results
Clean and uncluttered presentation
Gives user feeling of direct engagement
Allows seamless transition between search and browsing
Fun to use
Rewards user for giving more information
Interaction appropriate for type of task
Limit visual noise / optimize data-ink ratio
Minimizes scrolling

Table 1. A few possible answers to the question, "What makes a good search engine user experience?"

While there is no definitive set of such attributes, our research has identified a large number, some of which are shown in Table 1. It should be apparent at a glance that only a few of these attributes are found in a typical search engine. In particular, attributes such as "seamless transition between search and browsing" and "interaction appropriate for type of task" are absent, discouraging users from conducting more exploratory searches.

To investigate this discrepancy, we designed an experimental dedicated information-seeking environment called Elroy that attempts to explore *one* possible alternative for interacting with search.

ELROY DESIGN
We began by focusing on three specific desiderata that we felt were clearly absent from current web search systems:

- Search engines should *reduce or eliminate the barrier between search and browsing.* Information seeking includes more than explicit search; users often move towards their goal in an indirect or iterative fashion [3] and may not even have a clear idea of the destination when they start. Yet today's search engines are designed as though search occurs in isolation, as a one-time event. As a result, information seeking requires frequent context-switching between the search engine and navigation from various destination sites.
- Search engines should provide a feeling of *direct engagement* [5]. In a direct engagement interface, physical actions by the user are immediately reflected into actions by the system. Cars have direct engagement

interfaces; as the user turns the steering wheel, the car turns. In the virtual world, many video games provide direct engagement interface, often offering a viscerally immediate form of interaction.

- Search engines should be *fun* and provide *joy of use*. In the 1980s, many human-computer interaction researchers focused on ease of use. As a result of their efforts, most ordinary people can use computers today. But few designers have focused on joy of use, the idea that using the system could itself be an enjoyable experience.

Although these might not be the most important attributes for search, we felt they were worth exploring specifically because they receive short shrift today. Accordingly, we decided that the three desired attributes above would serve as our basic design principles for Elroy. Note that by making this choice, we were consciously deciding *not* to optimize for attributes already exhibited by current systems.

Our next task was to create a set of features that would embody or promote the design principles. We identified several new features, each of which is intended to support one or more of the design principles. Three of the features, described below, encouraged exploration by users.

Semantic Fisheye

A "fisheye" lens in photography is one that causes the center of the field of vision to be magnified, while the periphery is reduced to allow a wide angle of view to be included. The fisheye concept has been used in some interfaces such as Apple's Mac OS X dock [2] and explored in earlier research on information visualization [4] and in prototypes such as Xerox PARC's DocumentLens [8]. Fisheye interfaces are often described as providing a "focus + context" view of the content being displayed.

In Elroy, we introduce the notion of a "semantic fisheye" and apply it to navigating web search results. The semantic fisheye is an information visualization technique for browsing a list of documents such as web search results.[1] In the semantic fisheye view, one item in the list is the current focus, and this item is displayed with great detail. Items adjacent to the focus in the list are shown with moderate detail. Items farther away are shown in less detail, using only one line of text.

Unlike traditional fisheye views, semantic fisheyes do not require the graphic capability for visually magnifying and shrinking parts of the image. Instead, items grow and shrink by the addition or subtraction of content, which may be contained in the item itself or metadata about the item.

The effect is similar in some respects to Paek et al.'s WaveLens system [7], although the details are quite different. For example, in WaveLens, the concepts of focus and pointing are interleaved, with mouse hover and click actions used to change focus. In Elroy, focus and clicks are orthogonal. The focus is moved by an alternate control mechanism (for example, arrow keys), while the user may click on any item without requiring a focus shift. Also, the expansion in WaveLens takes the form of increasing the amount of page content shown, while in Elroy, expanded detail takes many forms including previously inaccessible metadata and suggestions for exploration.

History

In Elroy, the user's search/browse history is a first-class part of the user experience. Unlike a typical browser history menu or sidebar, Elroy not only shows where the user has been, it shows the order in which he or she visited. Each time step (defined by changing the search being conducted or the page being viewed) is represented visually as a frame in a kind of filmstrip. Our intent is to allow the user to "play back" the film to retrace the thought process that led to certain results.

Since search and browsing is not necessarily a sequential activity, and users backtrack and form new trails, some have suggested that a hierarchical tree structure would be a more appropriate representation of history. However, we believe that both the cognitive simplicity of a linear representation and the linear nature of time make a linear portrayal of history more appropriate for Elroy.

Dynamic Suggestions

In order to begin reducing the barrier between search and browsing, we designed a dynamic suggestion mechanism for Elroy that offers browse-type hyperlinks created in response to what we know about the documents from search-style indexing.

Some search engines have shown limited examples of query-specific suggestions in the past. For example, Yahoo! has an "Also Try" feature, Ask Jeeves has "narrow" and "expand" terms, and AltaVista used to show search suggestions related to the query [1]. Shopping sites such as Amazon.com have item-specific suggestions ("others who bought this item also bought...").

Elroy offers suggestions at three points in the information-seeking process: after issuing a query, when examining an individual result, and when browsing a specific web page. The query-specific suggestions include suggestions for other queries on related topics. Result-specific suggestions, shown on demand for each search result when that item becomes the focus of the semantic fisheye, may be additional pages from the same site or concepts describing what the site is about. The page-specific suggestions shown at browse time are an expanded version of result-specific suggestions; these appear in an alternate display pane while the user is browsing a specific web page. As the user goes from page to page, the suggestions dynamically update to reflect items relevant to the current page content.

[1] Note that our "semantic fisheye" is unrelated to similarly named "semantic fisheye views" of Janacek and Pu [6].

In all of these cases, the suggestion takes the form of a hyperlink that either generates a new search, navigates to another web page, or has some other action. Thus a user who is researching a topic on the web will not need to keep going back to the search engine query box and typing another query from scratch.

EXPLORATION WITH ELROY

We built an Elroy prototype using Flash, a platform designed for interactive animations. A screen shot of the interface is shown in Figure 1. The essential idea of the design is that at any given time during the search/browse process, the user will be given suggestions (in the form of clickable links or images) of different directions to try or different paths to follow. This is visible in the figure.

The large pane on the left, which we refer to as the display pane, shows a list of search results as rendered by the semantic fisheye. Each search result appears in one of three levels of detail, depending on the location of the focus. In the figure, item 5 is the current focus and therefore is shown in greatest detail, with the title, abstract, URL, a thumbnail of the page, a list of concepts describing it, and a list of other pages from the same website that also match the search query. So while the search result occupies a place in one dimension (the space of documents matching the query), the detail links invite the user to explore two others – the spaces of concepts and documents on the same site).

Items 4 and 6 are shown with intermediate levels of detail, with title, abstract, and URL (similar to a typical search engine result today). Other search results show only the title. Reducing detail in some items allows more results to appear on the screen, encouraging users to browse a larger set than is common in most search engines. (The result list extends beyond the end of the page, and is accessible by moving the focus past the last visible item.)

The pane on the upper right is the suggestion pane, here showing the query-specific suggestions. Each result set is thus portrayed as a jumping-off point for other possible queries. The pane across the bottom is the history, each frame representing a conceptual point in time when the display changed. By making it easy to return to a previous state of work, we believe users will be more likely to explore alternatives.

Elroy is a dedicated searching and browsing environment. This means that when a user clicks on a search result, the target web page is displayed in the same (display) pane where the search results were, while the other interface elements (the suggestion and history panes) remain in place. This is shown in Figure 2; the user has clicked on the focus result from Figure 1 ("washingtonpost.com: AIDS in Africa") and that page is shown in the display pane at left. Note that the suggestions have been altered to reflect awareness of the page the user is now viewing. Similarly, the history pane has been updated with a thumbnail of the latest page at the right.

CONCLUSIONS

Current search engines do a poor job of supporting different search strategies and styles; in particular, they offer little support for exploration. Elroy is an attempt to investigate how a web search engine might support a different set of attributes, including breaking down barriers between search and browsing. This is achieved by providing better tools for navigating result sets, better support for going back to previous states of work, and a constant dynamically updating set of visible options suggesting new directions for users to explore.

However, Elroy is just one point in a very large design space. User research is needed to find out whether the design goals have been achieved. We hope to learn whether there are particular information-seeking tasks for which a system like Elroy is especially well-suited. At the same time, we would like to know whether the benefits resulting from Elroy's underlying design principles outweigh the drawbacks of not supporting some attributes of existing search interfaces.

Although it is too early to tell whether this particular design will be successful, we believe the knowledge we gain about how different features affect the search experience for different users will inform our designs for all search interfaces in the future.

REFERENCES

1. Anick, P.J., "Using Terminological Feedback for Web Search: A Log-Based Study," *Proc. SIGIR Conf.,* 2003.
2. Apple Computer, Inc., "Mac 101 Lesson 4: The Dock", http://www.apple.com/support/mac101/tour/4/, accessed 30 October 2005.
3. Bates, M.J. (1991). The Berrypicking Search: User Interface Design. In M. Dillon (Ed.), *Interfaces for Information Retrieval.* Westport: Greenwood Press.
4. Furnas, G.W., "Generalized Fisheye Views," *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 1986.
5. Hutchins, E.L., Hollan, J.D., and Norman, D.A., "Direct Manipulation Interfaces," in *User Center System Design,* D.A. Norman and S.W. Draper, eds., Lawrence Erlbaum Associates, 1986.
6. Janacek, P.J., and Pu, P. "Opportunistic Search with Semantic Fisheye Views." EPFL Technical Report: IC/2004/42, School of Computer and Communication Sciences Swiss Federal Institute of Technology, Lausanne (EPFL), May 2004.
7. Paek, T., Dumais, S., and Logan, R. "WaveLens: A New View Onto Internet Search Results," *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2004.
8. Robertson, G.G., and Mackinlay, J.D., "The Document Lens," *Proc. 6th Annual Symposium on User Interface Software and Technology*, 1993.

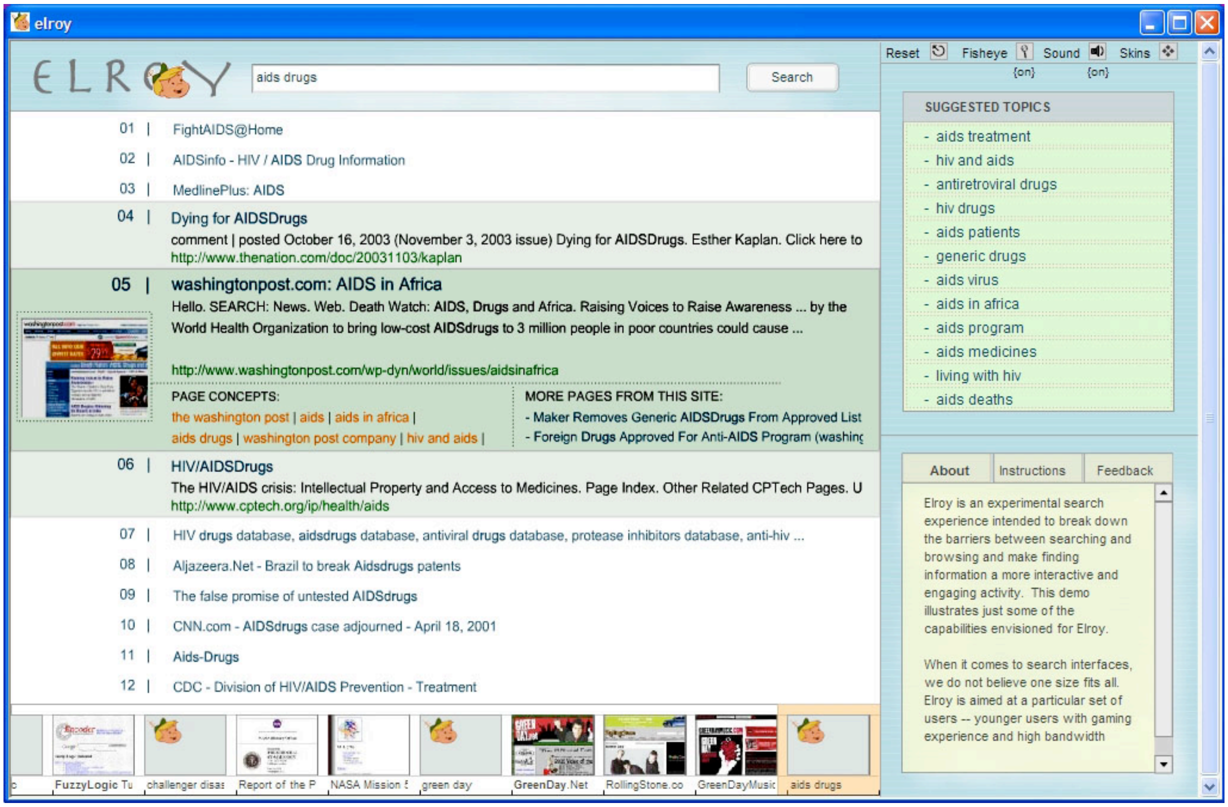

Figure 1. The Elroy Prototype Interface.

Figure 2. The Elroy interface when a document is selected. Note that the suggested topics and history have changed.

Exploratory Search for Business Users

Omar Alonso, Frankie James, Matthias Kaiser
SAP Research Center Palo Alto
3410 Hillview Avenue
Palo Alto, CA 94304
{firstname.lastname}@sap.com

ABSTRACT

Exploratory search systems have the potential to improve the overall user search experience by presenting relationships and mechanisms for discovering new insights. From a business perspective, this is in line with what a typical user needs to make better decisions. In this paper, we present a focus + context interface that is adapted to business needs to show a company view while presenting relevant external information.

Author Keywords

Exploratory search, focus + context, business process

ACM Classification Keywords

H.5 [Information Systems]: Information interfaces and presentation

BACKGROUND

Having the right information at the right time is an essential precondition for successfully managing enterprises of any kind, anywhere. Advances in research areas such as information retrieval, natural language processing, intelligent interfaces and modern Web technology provide opportunities for significant advancements in the quality of information management.

The success of Internet search engines has made a clear impact on the enterprise. Users within companies want to have the same Web experience that they have at home. However, the domains are very different and the single view of top-10 hits seems too narrow for the typical business user.

Web search engines use a wide range of technologies and heuristics to provide the best search results on the first page. The outcome of the competition for the highest spot on the hit list determines the overall search quality. The business user, on the other hand, is usually more interested in finding the right information based on his or her context. The context, in a business environment, consists of the user's business goal(s), the role the user plays within the company, and the current status within a concrete business process. Finding information that is contextually relevant is important, since it helps to minimize the process of reverse engineering search keywords to get good hits. In this sense, the success of the hit list construction depends a lot on the business context.

Exploratory search systems have gained a lot of attention lately as they help users discover new relationships for solving problems. [2] [3] We argue that exploratory search activities fit well in business scenarios where users need more contexts to reach solutions or find new insights. Also, exploratory search should provide not only the company view, but also overviews of every information source that matters. The goal is to provide an interface that helps the user makes better business decisions while presenting information that is focused on the company within its external context.

SCENARIO

We present our ideas using the following business scenario: Company X produces houseware products, and coffee makers are the best sellers. For the coffee maker product line, there are a number of models that differ in price, features, and style. The company has software for tracking every detail of the manufacturing and sales process.

Since there are different roles in the company, the right search results depend on the context of the question. A query for a product description "stainless steel coffee maker with grinder" or product name "CM-123-07", would result in information about product details, known issues, a particular claim that the customer has entered, or license negotiations with a major reseller. All of them are valid answers, but the determination of the context (e.g., design, consumer support, or sales) is the key for providing the right answer.

Most enterprise systems offer a very company-centric view, since all product information is centralized within the systems used from design to sales. There are other public sources that we would like to use to add more contextual information. For example, Company X would like to use sentiment tracking on certain blogs about coffee, or check certain models on prominent retailers and auction sites.

The notion of exploratory search is to combine all information available within the company boundaries with information that is available outside of the company, on the Web. The end result is a search interface that allows a user to get information within the right context, augmented with external sources that are beneficial to making decisions within the business process.

Let's go back to our coffee maker scenario. If the enterprise system only shows internal corporate data, Company X will get one picture of the sales for the coffee makers. The sales data and invoices might suggest that the grinder feature is a big seller: twice as many units are selling with grinders than without. The company can go back to the manufacturing plant and have them step up production of those units.

By adding data sources that are external to Company X, we may get an entirely different picture. Three days after the product launch for the coffee makers, an exploratory search uncovers hits on eBay for near-new condition units of the coffee maker plus grinder, but only 3 hits for the non-grinder version. Epinions shows a lower satisfaction rating for the grinder model than the non-grinder, and there are also a *lot* of negative reviews about the grinder. Finally, CoffeeBlog has an article about how bad the grinder is, with 400 comments agreeing. Consumers may have bought the coffee maker with the grinder, but they are now trying to off-load them on eBay and are generating a lot of negative press. This data suggests another strategy for Company X: stop production on the coffee maker with the grinder until they can talk to the product team and find out what's wrong with that grinder! Once it's fixed, they can market a new version, and possibly also offer some kind of upgrade or rebate for the folks who bought the faulty first version.

Note that the exploratory search approach can be used to show all possible contexts. In many business situations and for many business roles, however, some contexts are more relevant to others. For example, while a sales person is interested in the number of coffee makers sold and external information about satisfaction about the product, a technician trying to fix a certain problem (e.g., the faulty coffee grinder) is only interested in technical features and specifications.

Another important consideration is the business process in which a search is launched. Different results and data from different sources are relevant for the fulfillment of different tasks within a business process. This knowledge can be used to limit the search significantly, while at the same time increasing the probability for expected results. If a sales analyst is preparing a sales report, for a certain region (e.g., the US), she is interested in the results of previous sales in the United States, not in the sales planned in the future or in another region.

EXPLORATORY INTERFACES: BUSINESS FOCUS + CONTEXT

The focus + context technique states that the user needs detailed information (focus), but also needs to retain a global context. [1] The information in the overview may be different from the one presented in detail; for example, computer games often show the user's current situation (the focus), as well as a map of where the user is in the virtual world (the context). The presentation of both focus and context within a single display is an important feature, since it allows item-level exploration while maintaining the user's sense of the whole space.

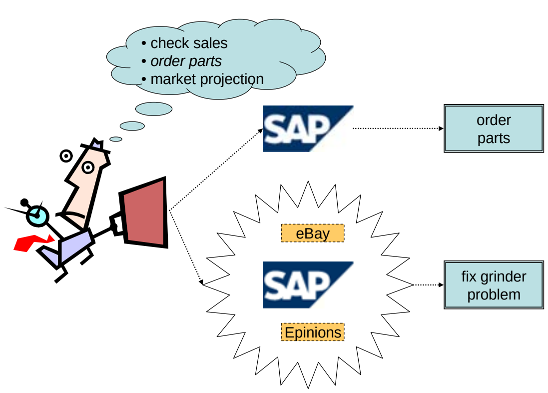

Figure 1: Current enterprise systems allow users to achieve relevant subgoals (e.g., check sales, order parts). The addition of external context can help to suggest new subgoals (e.g., fix grinder problem) that better address their real goals (e.g., maximize sales).

We extend the concept of focus + context to an enterprise situation where the focus is a company-centric view of a particular step in a task. The context is then is a set of overviews from external sources that are relevant to the focus. The focus at any given time is the current goal or subgoal in a (business) process which the user wants to achieve. We argue that the context in which a goal is in focus can actually change the priority of this goal to the extent that it may be discarded (usually temporarily) for the pursuit of another goal which, given based on information in the context, gains higher priority.

Returning to our coffee maker example, if the user's overall goal is to maximize profit on coffee makers and sales are high, this indicates that more products should be manufactured and sold. However, if the context indicates that, in fact, the customer satisfaction is low despite the high sales figures, another goal should probably be instantiated: namely, to find the cause for the dissatisfaction and fix it. After this has been done, the previous goal (sales maximization) can once again be pursued. Figure 1 (above) illustrates this concept.

In essence, a focus + context interface will allow the user to focus on their company within the context of a consumer- or industry-level view, to help the user achieve larger business goals, such as maximizing sales. This is a contrast from traditional enterprise systems that may cause the user to focus too narrowly on subgoals suggested by internal company data that are not supported by outside sources.

Figure 2 shows a user interface mock-up for our exploratory search prototype, based on the scenario used in this paper. The usual search box sits on top. The business focus is presented on the left (in this figure, the sales chart and list of recent invoices). Actions related to the business, such as a "people search" to connect the user to people in the enterprise relevant to this task (e.g., the product manager or sales representative), are presented in the lower left. The right side of the interface contains information from sources that are external to the company, but relevant to the coffee maker search. These include results from common consumer tools such as Epinions and eBay, blog hits, and news stories. The "Actions" area in the lower right allows actions related to the external data, such as external searches and personalization of the external tools.

CONCLUSION

In today's business, having a company-centric view of the world can lead to underestimation of one's consumers and competition. Using a focus + context approach can help bring things together. We believe that using this interface in an exploratory search system can improve the overall user experience, and can be applied to a range of business scenarios. For example, we can imagine that sales representatives might want to search on a customer prior to an upcoming meeting. The information that will be important is not only what is stored in the company database, but possibly also the company's recent press activity, consumer feedback, and other external data.

The ideas presented so far have been situated within a desktop application. We also see potential in a focus + context approach for mobile devices. For example, small amounts of "focus" material (such as the top hit from a search engine) could be presented using audio, while the context of the rest of the hits or other related information could be presented on the device's small screen. This multimodal "I'm feeling lucky" style of presentation may be of value in many mobile situations, such as while driving.

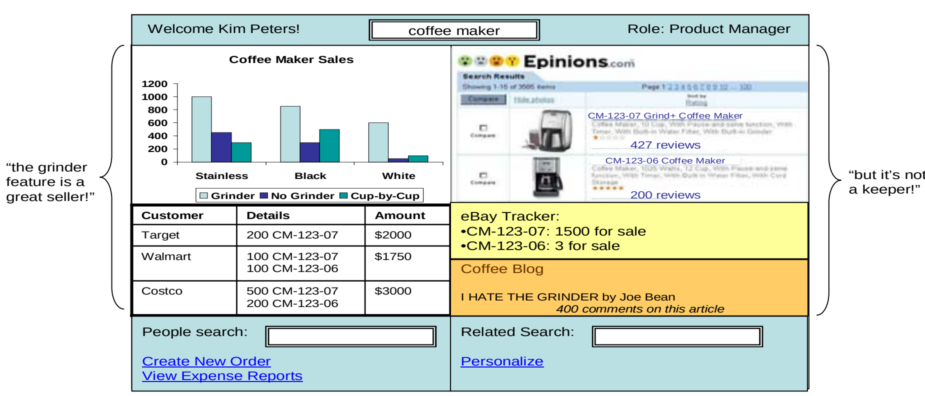

Figure 2: Exploratory Interface for Product Manager. Business focus is presented on the left. Relevant external content is presented on the right.

REFERENCES

1. G. G. Robertson and J. D. Mackinlay, "The Document Lens," *6th Annual ACM Symposium on User Interface Software and Technology*, Atlanta, GA, 1993.
2. R. White, B. Kules, S. Drucker, and M. Schrafael (Eds.) "Supporting Exploratory Search," *CACM*, Vol. 49, No. 4, 2006.
3. R. White, G. Muresan, and G. Marchionini (Eds.) "Evaluating Exploratory Search Systems Workshop," *ACM SIGIR*, Seattle, WA, 2006.

Interfaces to Support the Scholarly Exploration of Text Collections

Georg Apitz
Department of Computer Science & HCIL
University of Maryland
College Park, MD 20742, USA
geapi@cs.umd.edu

Jimmy Lin
College of Information Studies & HCIL
University of Maryland
College Park, MD 20742, USA
jimmylin@umd.edu

ABSTRACT

The analysis of text collections forms the basis of scholarship in many disciplines in the humanities and social sciences. Despite the growing availability of electronic texts, automated techniques have not been effectively exploited to support the activities of scholars in these fields. We present a prototype search interface for exploring text collections that places equal emphasis on content, what the document is *about*, and metadata, the context that situates a piece of text. As a start, we focus on a selection of briefs and opinions from the U.S. Supreme Court to support legal scholars.

Author Keywords

User interfaces, exploratory search, creativity support tools

INTRODUCTION

Scholarship in many areas of the humanities and social sciences involves large collections of textual documents. Consider, for example, a legal scholar who combs through hundreds of court documents to find support for a particular hypothesis, or a literary scholar who analyzes countless pages of writings to ascertain stylistic influences among contemporaries. Traditionally, such studies have been laboriously carried out by hand, and despite the growing availability of these texts in electronic format, automated systems are not usually exploited to support scholarly endeavors.

At a broad level, our work aims to develop innovative interfaces to help scholars in the humanities and social sciences explore large text collections. We seek to develop creativity support tools [12] for scholarship in these various fields. The goal is not to supplant, but to augment the scholar in the creation of knowledge—by utilizing the tool to visualize relationships between different texts, to drill down into a text, to draw connections otherwise not apparent, and most important of all, to form hypotheses that provide the basis for further exploration. We take the view that computers by themselves do not generate knowledge—its creation falls within the purview of humans who interpret system output. Thus, the system's primary role should be to facilitate the human-centered processes of hypothesis formulation, evidence gathering, etc. As a start, we focus on a collection of legal briefs and opinions from U.S. Supreme Court cases.

It is apparent that the tools we're describing lie at the intersection of human-computer interaction and information retrieval. Information retrieval and related text processing techniques (data mining, linguistic analysis, text classification and clustering, etc.) can unearth characteristics of texts that may be of interest to scholars. However, these findings need to be synthesized and pre-digested into a form suitable for consumption by individuals who may not be experts in computational techniques. This requires, for example, visualizations that render various relationships apparent, explanation tools that help the scholar understand the findings of automated algorithms, and controls to subsequently affect the behavior of these algorithms.

This paper focuses on the interface aspect of such creativity support tools. Using off-the-shelf information retrieval technology, we demonstrate a prototype interface for exploratory search that emphasizes both the presentation of content and associated metadata. The emphasis on both aspects provides scholars with a rich environment to engage a text collection.

METADATA VS. CONTENT

A *document* (used broadly to encompass any piece of text) is characterized by its content, on the one hand, and its associated metadata, on the other hand. Content is comprised of the words that make up the text and define what the document is *about*. Metadata define important characteristics such as authorship, document type, time of creation, association with other documents, etc. In a sense, metadata define the context of a particular document.

Traditionally, the exploration of text collections is performed primarily through the manipulation of either metadata or content. In the legal domain, one might use metadata to explore briefs and opinions by case, by author (e.g., judge), by issue area, or chronologically. This is often facilitated by inserting metadata into a database, which allows users to issue arbitrary relational queries. For specific tasks, this can be a very useful and efficient method for accessing documents.

Free-text search on document content represents an alternative to exploration by metadata. Given the prevalence of Web search engines such as Google, this mode of information access is most familiar to users today. In response to a keyword query, retrieval engines return a list of "hits" that are likely to be relevant. However, metadata are not leveraged in the search algorithm for the most part, and when they are, results often conflate multiple factors (for example, Google rankings combine topical relevance, authority, popularity, etc.). In a general Web environment, one might argue that these issues are less important to the typical end user. However, this is not the case for scholarly collections, where the context of a particular document, as defined by metadata, is often the subject of exploration itself.

We believe that a tool for exploring document collections must allow users to manipulate both content and metadata. Although some search engines provide mechanisms to search on metadata (through special query operators or in different fields), these capabilities do not go far enough and are not supported by corresponding interfaces that render the relevant relationships transparent. In our opinion, the traditional output of most search engines—a list of hits sorted by relevance—is not sufficient to support exploration. To address this, we focus on search interfaces as a starting point.

PREVIOUS WORK

We divide our discussion of previous work into two parts, one surveying the literature from information visualization, and the other reviewing work in information retrieval. Although visualization interfaces are clearly relevant to this work, most existing systems focus solely on metadata. For example, BiblioViz [11] employs a combination of different visualization techniques and clustering approaches to present bibliographical data (based on contributions to the InfoVis 2004 contest). Chen [2] also presents 3-D and 2-D visualizations of citation and co-citation networks as a summary of the InfoVis 2004 contest entries. The Hierarchical Clustering Explorer (HCE) [9] supports the user in exploring a dataset based on the metadata—the idea is that the user first explores the data in 1-D, then in 2-D, and then uses a rank-by-feature mechanism to expose more interesting patterns.

In the information retrieval literature, there is a substantial body of work on search interfaces. Some of these focus on query specification [14] or graphically convey the distribution of query terms in retrieved document sets [3, 15]. Others attempt to visualize the relationship between documents in the result set [1, 6]. Cat-a-Cone [4] represents an interesting attempt to combine search and browsing, but was primarily designed for hierarchically-categorized documents. The work of Nowell et al. [7] share the most similarities with ours, although their Envision system was designed as a tool for accessing digital libraries.

THE DIGITAL DOCKET

This work is situated in the context of the Digital Docket, a recently-funded NSF project that aims to apply automated content analysis to support legal scholarship. Previous research of judicial systems has faced a trade-off between large scale quantitative inquiries focused on readily-counted behaviors and smaller studies that allow closer examination of legal texts. This project, a multi-disciplinary collaboration between several units on the University of Maryland campus, aims to apply computational techniques to the study of the U.S. Supreme Court.

By viewing the legal system as an intricate and complex web of communication, discourse, and debate, the project aims to better understand the role and influences of various actors through analysis of written records. Those records include, for example, briefs written by litigants and other stakeholders, and opinions written by judges and justices. The application of computational techniques to model the U.S. judicial system represents an opportunity to overcome many of the bottlenecks associated with traditional labor-intensive methods in political science, and also provides a new environment for the advancement of text algorithms.

Propagation of "Legal Memes"

As a first goal, we are developing a tool that assists legal scholars in exploring the propagation of "legal memes". Consider the issue of affirmative action, a topic that has recently received substantial coverage in the popular press. In the landmark Supreme Court case *Regents v. Bakke (1978)*, Justice Powell wrote that an admissions program must be "flexible enough to consider all pertinent elements of diversity". Such a statement introduces a coherent concept into the "ecosystem" of the legal system—subsequent litigants read the opinion, future briefs may quote or cite this idea, law review articles may debate its interpretation, etc. A legal scholar may be interested in the following:

- Where did this "meme" come from? Did it originate from Powell, or from an even earlier piece of writing?
- How influential is this idea? Obviously, direct quotes and citations are one way to quantify this, but ideas are often paraphrased without explicit reference.
- Do references to this meme change over time? Is it currently "out of fashion" to discuss affirmative action issues in these terms?
- What are the subsequent interpretations of this idea? Do other justices have incompatible views?
- How does this idea affect the perception of the individual? Does one gain prominence as a result of widely-adopted ideas, or vice versa?

A tool for exploring court documents can be valuable to legal scholars in formulating hypotheses and in gathering evidence. Content is clearly important in this application, but so is metadata, since it provides the basis for answering many of the above questions.

A PROTOTYPE

We have developed a prototype exploratory search tool, the DigiDock Explorer, that allows legal scholars to explore documents associated with cases heard by the U.S. Supreme Court—it begins to support some of the activities discussed

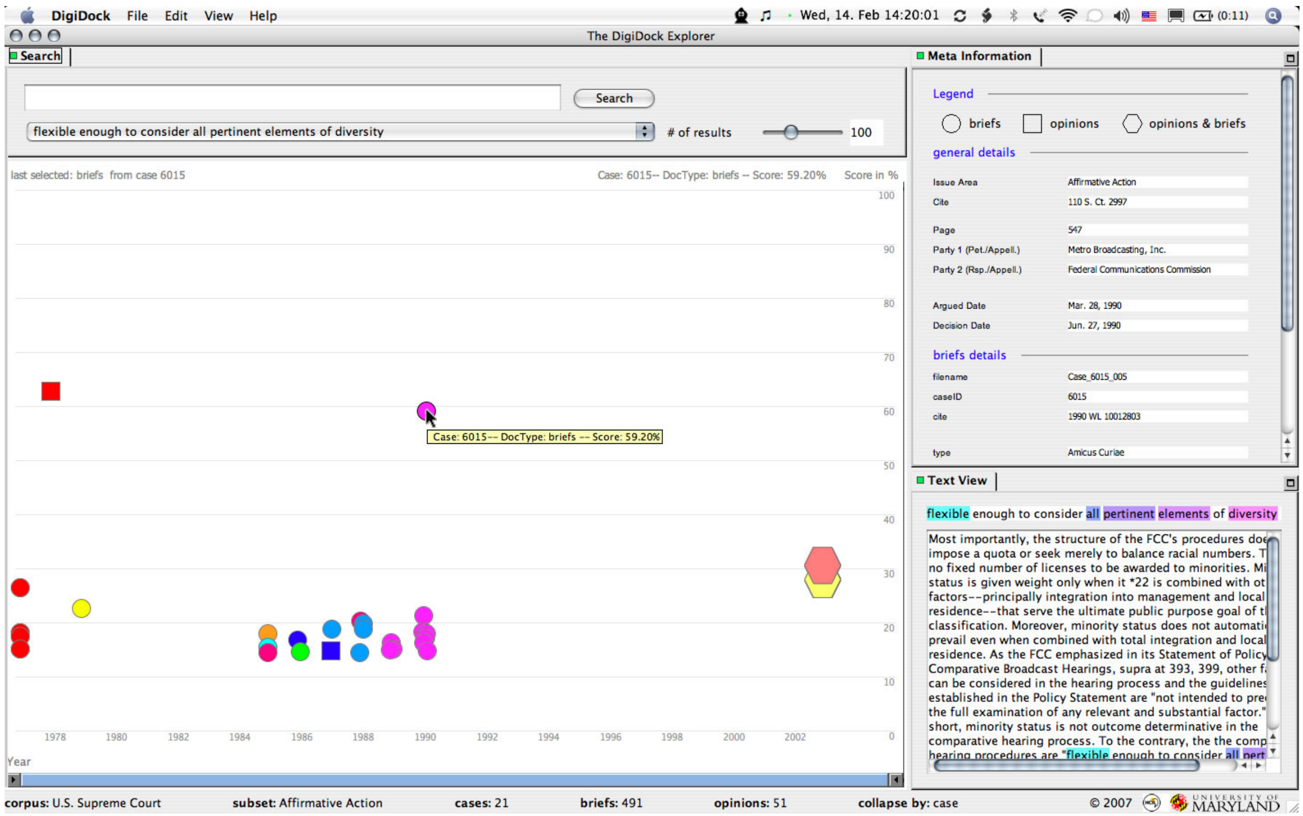

Figure 1. The DigiDock Explorer screen. The main portion of the screen shows the search results for the current search phrase displayed as time (x-axis) vs. relevance (y-axis). Several results have been aggregated by case (hexagons). The right side shows details about the selected item, these are: general and document type specific information as well as a text view. At the bottom a status bar shows details about the corpus and current settings.

above. The application is implemented in Java using the Prefuse [5] Information Visualization toolkit. A screenshot is shown in Figure 1.

Document Collection and Preparation

Twenty U.S. Supreme Court cases about affirmative action between 1978 and 2003 were selected for inclusion in our sample collection. For each case, we obtained all briefs from the litigants, all *amici* (third-party) briefs, and all opinions of the court. This totaled approximately five hundred documents, about half a million words. Although this represents a small collection, it is useful for legal scholars since its coverage of one specific issue area at the Supreme Court level is relatively complete.

Each document in our collection is associated with metadata mined from various sources. They include properties such as the document type, the author, date, the case association, etc. Different types of documents have additional specialized metadata; for example, opinions can be majority, dissenting, etc. In addition, we have metadata about the cases in general, e.g., what the outcome was, the vote, etc.

The entire collection was indexed with Lucene, an open source search engine that uses modified *tf.idf* weighting. Since the documents are very long on average, we segmented them into paragraphs and indexed each separately.

Interface

Like most retrieval systems, the starting point of the DigiDock Explorer is a search box at the top of the interface, where users can input a query string or select from predefined queries (that may be populated, for example, by text mining algorithms). Retrieved results (paragraph segments) are displayed in a scatterplot in which the x-axis shows the creation date of the document and the y-axis shows the similarity score returned by Lucene.

We have chosen this display format since feedback from legal scholars has indicated that time is the most important dimension to consider when examining cases. Since the legal system is causal in that previous cases shape the argumentation and outcome of subsequent cases, an explicit visualization of this property provides important cues. Different icons represent different document types (circles for briefs, squares for opinions). Users can group documents to reduce clutter—for example, a large hexagon is an aggregate representation of all segments belonging to a single case. With this display, scholars can get a broad overview of the information landscape and directly answer questions such as:

How many potential instances of this concept exist in the collection? How are they distributed temporally?

The right hand side of the application displays metadata associated with the retrieved segment and its contents. Keyword highlighting is employed to facilitate browsing of text. Arbitrary segments can be selected and used as "queries", thus allowing drill-down. Users can also use the metadata panel to control the display, for example, to show only certain types of documents, text by certain authors, etc.

NEXT STEPS

With our early prototype already in use by scholars we are using their feedback to extend the functionality of the DigiDock Explorer. We are planning to allow the user to have different visualizations for the search results, as well as very fine-grained control over how the search is conducted.

Furthermore, we are investigating ways of visualizing the propagation of legal memes. An initial 'river like' layout seems like a good starting point but we are envisioning representations that show the flow, as well as the kind of documents that are influences and the people who are influenced. This aims at the representation of several features in the propagation and should potentially support the user in ranking these features by their importance, similar to Seo and Shneiderman's Rank by Feature approach [10].

In addition to "free association" exploration, we envision that scholars use our tool in *sessions* where they aim to accomplish a specific goal by performing several iterations of drill-down, refinement, and restriction of intermediate results with the tool. These *sessions* themselves encode information about a search strategy, the steps involved, and the criteria applied. Thus, we want to provide users with a recording functionality that allows them to store, replay, and share such sessions.

The legal domain represents one instance where scholars rely on large collections of text. We imagine a general purpose tool that can be applied to other disciplines in the humanities and social sciences. Comparative literature represents an interesting application—see, for example, [8]. Tools for exploring large text collections can be extended even further into domains such as biology and medicine.

The evaluation of interfaces for domain experts is a very specific challenge we must address. Traditional user studies, where subjects work 30 minutes or so with an application are not well suited for our purposes. Thus, we are planning on applying a technique called Multi-Dimensional, In-depth, Long-term, Case studies (MILCs) [13]. With this evaluation methodology, it is possible to observe expert users in their traditional environment for longer periods of time and gather information that only occurs in natural settings. Logging capabilities can deliver fine-grained information about user interactions that are difficult to observe. In addition, interviews, self reports, and think-aloud methods can provide more subjective insights from the scholars.

CONCLUSION

The exploration of text collections is a common task among scholars, yet so far, there is no solution that leverages information about the documents (metadata) as well as information on what the document is about (content). We aim to develop systems that combine the benefits from both with the domain expertise of scholars.

Our prototype demonstrates a new way for legal scholars to explore, research, and examine judicial systems. This application enables them to look at document collections in ways never before possible and to gain insights that traditionally would require much manual labor. We believe that this example illustrates a new way of interacting with text collections that enables scholars to more fully exploit the benefits of technology in combination with their knowledge.

ACKNOWLEDGMENTS

This project is funded by NSF under award BSC-0624067. Any opinions, findings, and conclusions expressed in this paper are those of the authors. We would like to thank Ben Shneiderman for valuable discussions. The authors would like to thank Anja, Esther, and Kiri for their kind support.

REFERENCES

1. M. Chalmers and P. Chitson. Bead: Explorations in Information Visualization. In *SIGIR 1992*.
2. C. Chen. Information Visualization Research: Citation and Co-Citation Highlights. In *INFOVIS 2004*.
3. M. Hearst. TileBars: A Visualization of Term Distribution Information in Full Text Information Access. In *CHI 1995*.
4. M. Hearst and C. Karadi. Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy. In *SIGIR 1997*.
5. J. Heer. Prefuse: Information Visualization Toolkit. *www.prefuse.org*, 2004.
6. R. Korfhage. To See, or Not to See—Is That the Query? In *SIGIR 1991*.
7. L. Nowell, R. France, D. Hix, L. Heath, and E. Fox. Visualizing Search Results: Some Alternatives to Query-Document Similarity. In *SIGIR 1996*.
8. C. Plaisant, J. Rose, B. Yu, L. Auvil, M. Kirschenbaum an M. Smith, T. Clement, and G. Lord. Exploring Erotics in Emily Dickinson's Correspondence with Text Mining and Visual Interfaces. In *JCDL 2006*.
9. J. Seo and B. Shneiderman. Interactively Exploring Hierarchical Clustering Results. *Computer*, 35(7):80–86, 2002.
10. J. Seo and B. Shneiderman. A Rank-by-Feature Framework for Interactive Exploration of Multidimensional Data. *Information Visualization*, 4(2):96–113, 2005.
11. Z. Shen, M. Ogawa, S. Teoh, and K. Ma. BiblioViz: A System for Visualizing Bibliography Information. In *Proceedings of Asia-Pacific Symposium on Information Visualization*, 2006.
12. B. Shneiderman. User Interfaces for Creativity Support Tools. In *Proceedings of the 3rd Conference on Creativity & Cognition*, 1999.
13. B. Shneiderman and C. Plaisant. Strategies for Evaluating Information Visualization Tools: Multi-Dimensional In-Depth Long-Term Case Studies. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors*, 2006.
14. A. Spoerri. InfoCrystal: A Visual Tool for Information Retrieval & Management. In *CIKM 1993*.
15. A. Veerasamy and N. Belkin. Evaluation of a Tool for Visualization of Information Retrieval Results. In *SIGIR 1996*.

Synergistic interaction: visualizing and exploring together

Russell Beale
School of Computer Science, University of Birmingham
Edgbaston, Birmingham, B15 2TT, UK
R.Beale@cs.bham.ac.uk

ABSTRACT
This paper presents an overview of an exploratory system developed to allow user to work synergistically with the system in order to explore and understand data. The system comprises an interactive visualization component for multidimensional data, and a descriptive component that provides rules to explain discoveries, and which can feed back into the visualization. The system has been used for supporting exploratory searching in a digital library as well as for knowledge discovery in large datasets.

INTRODUCTION
Data mining, or knowledge discovery as it is sometimes called, is the application of artificial intelligence and statistical analysis techniques to data in order to uncover information[5]. Given a number of large datasets, we are fundamentally interested in finding and identifying interesting relationships between different items of data. Whatever the domain of the data, we are engaged in a search for knowledge, and are looking for interesting patterns in the data.

But what is "interesting"? One day, it may be that the data falls into a general trend - the next it may be the few outliers are the fascinating ones. Interest, like beauty, is in the eye of the beholder. For this reason, we cannot leave the search for knowledge to computers alone. We have to be able to guide them as to what it is we are looking for, which areas to focus their phenomenal computing power on. In order for a data mining to be generically useful to us, it must therefore have some way in which we can indicate what is interesting and what is not, and for that to be dynamic and changeable[4].

There are further issues with current approaches to data mining, in that the answers are often almost as incomprehensible as the raw data. It may be that rules can be found to classify data correctly into different categories, but if the rules to do so are pages long, then only the machine can do the classification and we haven't gained knowledge: we may know how to do the classification but have no insight into why it may be like that. We believe that we should be able to understand the answers that the system gives us. In order to achieve this, it may be that we need broader, less accurate generalisations that are comprehensible to the human mind, but then feel confident in the main principles to allow the machine to do classification based on much more complex rules that are refinements of these basic principles. For example: "if it's red and squishy, it's a strawberry" is easy to understand. Even if that's true only 80% of the time, it's a useful rule, and easier to grasp than, say:

```
red, deforms 4mm under 2N pressure, >3cm diameter
= strawberry &
red, deforms 1mm under 2N pressure, <6cm diameter
= cherry &
red, deforms 3 mm under 4N pressure, >5cm diameter
= plum
else rasberry
```

which may be 96% correct but is hardly memorable. For many data mining systems, the rules developed are far more complex than this, each having numerous terms, with no overall picture able to emerge. For statistical-based systems, the parameter sets are even harder to interpret.

Since "interesting" is essentially a human construct, we have argued that we need a human in the data mining loop, and if we are to develop an effective system we need to allow them to understand and interact with the system effectively[2]. We should also take advantage of the capabilities of the user, many of which we have tried to emulate with AI systems for many years and are still a long way from reproducing effectively. A key example is the human visual system, which is very effective at picking out trends within a mist of data points, capable of dealing with occlusion, missing values and noise without conscious effort. On the other hand, processing vast numbers of points and deriving complex statistics is something much better suited to computers.

This leads us to conclude that a knowledge discovery system should be interactive, should utilise the best in artificial intelligence, evolutionary and statistical techniques in deriving results, but should be able to trade accuracy for understanding, and needs to provide a way of allowing the user to indicate what is interesting and to understand the suggestions that the computer makes. An ideal system should be symbiotic, each benefiting from the intrinsic abilities of the other, and holistic, producing results that are much more powerful than each could achieve on their own.

KNOWLEDGE DISCOVERY WITH HAIKU
The Haiku system was developed with these principles in mind, and offers a symbiotic system that couples interactive 3-d dynamic visualisation technology with a novel genetic

algorithm. The system creates a visualisation of the data which the user can then interact with, defining which areas are of interest and which can be ignored. The system then takes this input and processes the data using a variety of techniques, presenting the results as explanations to the user. These are in both textual and visual form, allowing the user to gain a broader perspective on what has been achieved. Using this information, they can refine what the system should look at, and slowly focus in on developing knowledge about whatever it is they are interested in. As well as using conventional rule generation techniques, Haiku also has a specifically designed genetic algorithmic approach to producing explanations of data. Each of these components is described in more detail below.

Visualisation

The visualisation engine used in the Haiku system provides an abstract 3-d perspective of multi-dimensional data. The visualisation consists of nodes and links, whose properties are given by the parameters of the data. Data elements affect parameters such as node size, mass, link strength and elasticity, and so on. Multiple elements can affect one parameter, or a subset of parameters can be chosen. Many forms of data can be visualisated in Haiku. Typical data for data mining consists of a number of individual "items" (representing, for example, customers) each with the same number of numerical and/or nominal attributes. What is required for Haiku visualisation is that a distance can be calculated between any two items. The distance calculation should match an intuitive view of the differences between two items. In most cases, a simple and standard Hamming distance measure performs well. Clearly, many variations of this exist - a weighted sum can be used, and so on. One of the characteristics of the system is that the user can choose which parameters are used to create the distance metric, and which ones affect the other characteristics of the visualisation.

In the visualisation, a node is created that represents an item. These nodes may be all equivalent, or may have characteristics inherited from the data (e.g. number of children may be used not in the standard distance measure, but in the mass of the node). Links are created between all the nodes, which act as springs and try to move the nodes about in the space. To create the visualisation, nodes are initially scattered randomly into the 3d space, with their associated links. This 3d space has obeys a set of physical-type laws, which then affect this initial arrangement. Links tend to want to assume a particular length (directly related to the distance measure between the nodes), and tend to pull inwards until they reach that length, or push outwards if they are compressed, just as a spring does in the real world. Nodes tend to repel each other, based on their mass. This whole approach can be seen as a force directed graph visualisation. This initial state is allowed to evolve, and the links and nodes shuffle themselves around until they reach a local minimum, low energy steady state. The reasoning behind these choices of effects are that we want related things to be near to each other, and unrelated things to be far away. Therefore, by creating links that are attractive between data points with similar characteristics, we achieve this clumping effect. The data points themselves, the nodes in the visualisation, are made repulsive so that the system does not collapse to a point, but instead are individually distinguishable entities, slightly separated from their similar neighbours.

This approach achieves a number of things. It allows us to visualise high-dimensional data in a comprehensible and compact way. It produces results that are similar to those achieved using approaches such as multidimensional scaling, but is somewhat more comprehensible because it tries to cluster 'similar' things with other 'similar' ones. It is certainly true that the choice of distance metric, and particularly which items to include and which to map to node characteristics, can affect the resulting visualisation, but we are searching for insight and meaning, not trying to come up with a single right solution. At different times, different features can be examined, and different results achieved - this is an inherent characteristic of searching for information, rather than an intrinsic problem with the approach. In any move from a high-dimensional space to a lower one, information will have to be lost - this approach at least preserves some of the main similarity characteristics of the original datasets.

The physics of the space are adjustable, but are chosen so that a steady state solution can be reached that is static - this is unlike the real world, in which a steady state exists that involves motion, with one body orbiting another. In the real physical world (a Newtonian space) in the steady state each node may potentially have zero or a constant velocity. In other words, the steady state solution has dynamic properties, with bodies moving in orbit, for example. In our space, we redefine the force equations and it can be shown that the final arrangement is static. To evolve the structure, each node is checked for links to other nodes, and the forces of those links is added vectorially, and the node is then moved according to that force. Computationally, the process scales exponentially with the number of links, which is usually proportional to the number of data points, and so the evolution to the stable structure moves from being a real-time process that you can watch towards one that has to be allowed to run for a long period of time as the dataset increases in size. In general, this is not a problem, since the initial arrangement of data is random and the evolutionary process is not in itself informative (although it is interesting to observe). However, when the visualisation is used as a component in the data mining tool, this is designed to be an interactive process, and so we have taken a number of approaches to speeding up the relaxation to steady state. The most effective approach is to use predominantly local relaxation, however: instead of considering all the forces to act over infinite distance, we can limit nodal interactions to be very local, so that nodes a long way away do not exert any forces on the ones in question (much like assuming that the gravitational effects of all the stars except the sun are negligible). Once the system has undergone some initial relaxation, which provides some level of organisation, we can also focus on the local neighbourhood much more, and occasionally recompute the longer-range interactions. This is akin to organising a tight cluster properly, but then treating that as one structure for longer-range effects. This and other approaches allow us to produce an effective steady state representation even with large datasets, in interactive time.

This representation can then be explored at will by rotating it, zooming in and flying through and around it. It is a completely abstract representation of the data: different data to attribute mappings will clearly give different structures, but the system can at least produce a view of more than 3 dimensions of the raw data at once. A typical structure is shown in Figure 1.

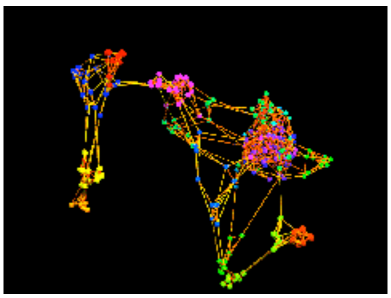

Figure 1. Nodes and links self-organised into stable structure.

The interface provides full 3D control of the structure, from zooming in and out, moving smoothly through the system (flyby), rotating it in 3D, and jumping to specific points, all controlled with the mouse. The visualisation in itself provides a lot of information about the dataset. We have used the visualisation in isolation for a number of tasks, including the visualisation of users internet browsing behaviour, information retrieval tasks within a digital library[1], and exploration of classic soyabean and Boston housing data sets[3].

Interaction with the Data Visualisation

When features of interest are seen in the visual respresentation of the data they can be selected using the mouse. This opens up a number of possibilities: data identification; revisualisation; explanation. The simplest of these (data identification) is to view the identifiers or details or items in the feature, or export this information to a file for later use. Another option is re-visualise the dataset without the selected data or indeed to focus in and only visualise the selected data. This can be used to exclude distorting outliers, or to concentrate on the interactions within an area of interest. Of course, we can data mine the whole dataset without doing this, the approach taken by many other systems. One of the features of the Haiku system is this interactive indication of the things that we are currently interested in, and the subsequent focussing of the knowledge discovery process on best describing that data only.

A key feature of the system is that this user selection process takes full advantage of the abilities of our visual system: humans are exceptionally good at picking up gross features of visual representations. Our abilities have evolved to work well in the presence of noise, of missing or obscured data, and we are able to pick out both simple lines and curves as well as more complex features such as spirals and undulating waves or planes. By allowing user input into the knowledge discovery process, we can effectively use a highly efficient system very quickly as well as reducing the work that the computational system has to do.

The most striking feature of the system is its ability to "explain" why features of interest exist. Typical questions when looking at a visual representation of data are: "Why are these items out on their own?", "What are the characteristics of this cluster?", "How do these two groups of items differ?". Answers to these types of question are generated by applying a machine learning component.

The interaction works as follows: First, a group or number of groups is selected. Then the option to explain the groups is selected. The user answers a small number of questions about their preferences for the explanation (short/long) (Highly accurate / characteristic) etc. The system returns a set of rules describing the features selected. As an alternative, the classic machine learning system C4.5 may be used to generate classification rules.

GENETIC ALGORITHMS FOR DATA MINING

We use a genetic algorithm (GA) approach for a number of reasons[6]. The first is that a GA is able to effectively explore a large search space, and modern computing power means we can take advantage of this within a reasonable timeframe. We use a special type of GA that evolves rules; these produce terms to describe the underlying data of the form:

IF `term` OP `value|range` (AND …) THEN `term` OP `value|range` (AND …)

where `term` is a class from the dataset, OP is one of the standard comparison operators (<, >, =, ≤, ≥), `value` is a numeric or symbolic value, and `range` is a numeric range. A typical rule would therefore be:

IF colour = red AND consistency = soft THEN fruit = strawberry

A set of these rules can, in principle, describe any arbitrary situation. There are two situations that are of interest to us; classification, when the left hand side of the equation tries to predict a single class (usually known) on the right hand side, and association, or clustering, when the system tries to find rules that characterise portions of the dataset. The algorithm follows fairly typical genetic algorithmic approaches in its implementation, but with specialised mutation and crossover operators, in order to explore the space effectively. We start with a number of random rules, and evolve the population through subsequent generations based on how well they perform. The genetic algorithm aims to optimise an objective function, and manipulation of this function allows us to explore different areas of the search space. For example, we can strongly penalise rules that give false positive results, and achieve a different type of description than rules that may be more general and have greater coverage, but make a few more mistakes. Each rule is analysed in terms of the objective function and given a score, its fitness. The fittest rules are then taken as the basis for the next population, and new rules created. Crossover points are chosen to be in syntactically similar positions, in order to ensure that we are working with semantically meaningful chunks. Mutation is specialised: for ranges of values it can expand or contract that range, for numbers it can increase or decrease them, for operators it can substitute them with others.

Statistically principled comparisons showed that this technique is at least as good as conventional machine learning at classification, but has advantages over the more conventional approaches in that it can perform clustering operations too. One of the key design features is to produce a system that has humanly-comprehensible results. Rules in conjunctive normal form are inherently much more understandable than decision trees or probabilistic or statistical descriptions. It is also true that short rules are going to be easier to comprehend than longer ones. Since the GA is trying to minimise an objective function, we can manipulate this function to achieve different results. If we insist that the rules produced must be short (and hence easier to understand) then the system will trade off accuracy and/or coverage but will give us short rules, because they are 'fitter', which provide a general overview that is appropriate for much of the data. Because the Haiku system is interactive and iterative, when we have this higher level of comprehension, we can go beck into the system and allow the rules to become longer and hence more specific, and accuracy will then increase.

Feedback

Rules generated using C4.5 or the genetic algorithm based method can be visualised within the system to give extra insight into their relationships with the data. Rules are usually represented by massive nodes that don't move far in space, and are regularly spaced. Links show which rules apply to which data, and hence unclassified data and multiply classified data are shown well, as in Figure 2.

In this figure, rules are the large purple, fuschia and green spheres, with the data being the smaller spheres. Links are formed between the rules and the data that is covered by the rule, and the visualisation has reorganised itself to show this clearly. We have additionally coloured the data according to its correct classification. A number of things are immediately apparent from this visualisation, much more easily than would the case from a textual description. On the very left of the figure, one rule, the fuschia sphere, covers exactly the same data as the other fushia sphere, except it also misclassifies one green data point. But the rightmost fushia rule, whilst correctly classifying all the fushia data also misclassifies much of the other data as well. On the right hand side, the purple rule clearly does very well; it covers all its data and doesn't misclassify anything. The green rule at the top has mixed results. The system is fully interactive, in that the user can now identify different characteristics and instruct the GA to describe them, and so the process continues.

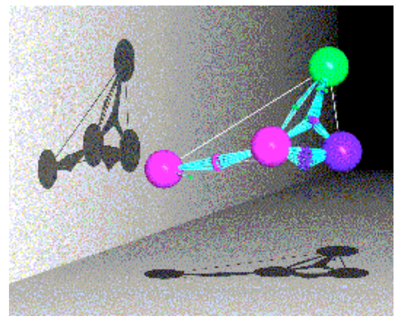

Figure 2: Rules and classified data

This synergy of abilities between the rapid, parallel exploration of the structure space by the computer and the user's innate pattern recognition abilities and interest in different aspects of the data produces a very powerful and flexible system.

REFERENCES

[1] Beale, R. Supporting serendipity in data mining and information foraging. *International Journal of Human Computer Studies*, *65* (5).(2006), 421-433.

[2] Beale, R., Pryke, A. and Hendley, R.J., Evolutionary approaches to visualisation and knowledge discovery. in *Computer Human Interaction*, (Rotorua, New Zealand, 2004), Springer-Verlag, 30-39.

[3] Blake, C.L. and Merz, C.J. UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html.

[4] Ceglar, A., Roddick, J.F. and Calder, P. Guiding Knowledge Discovery through Interactive Data Mining., Technical Report KDM-01-002. KDM Laboratory, Flinders University, Adelaide, South Australia., 2001.

[5] Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. From data mining to knowledge discovery: an overview. in *Advances in knowledge discovery and data mining*, American Association for Artificial Intelligence, 1996, 1-34.

[6] Freitas, A.A. A survey of evolutionary algorithms for data mining and knowledge discovery. in *Advances in evolutionary computing: theory and applications*, Springer-Verlag New York, Inc., 2003, 819-845.

Seven Habits of Highly Exploratory Interaction?

Duane Degler
IPGems
Columbia, MD, USA
ddegler@ipgems.com

ABSTRACT

This position paper presents observations and perspectives on user habits for exploratory search and information synthesis, in order to consider possible interaction components that could enhance non-targeted information seeking. The observations and perspectives come from many years of personal research and professional practice designing and developing information systems for large organizations, utilizing user-centered methods and semantic/metadata-based models and tools to improve both user experience and organizational outcomes.

THE IMPORTANCE OF USER INTERACTION

> "In our expanding information universe, computer scientists must ensure that access enhances and enriches everyone's meaningful experience with information, rather than dehumanizes it by possibly omitting its context."
>
> *Argamon & Olsen, "Toward Meaningful Computing," 2006, p 35 (1)*

Users' goal achievement and satisfaction is driven by the quality of their interaction with both information and the mechanisms for reaching that information. It is actually very difficult to achieve quality interaction, and the design pursuit is often pushed to a lower priority when faced with the technical complexity of delivering rich exploratory and synthesizing environments, or simply meeting project deadlines. In other disciplines, the lack of focus on quality user interaction has marginalized the knowledge management community, slowed the progress of the semantic web, and possibly hindered the advancement of anything beyond "basic search."

A recent study of search behavior among consumers reported that only about 38% of people move beyond the first page of search results (2), reinforcing other study findings (3). This presents a risk that user needs are not adequately addressed by search except in very targeted situations, particularly if those needs are not well understood by the user. This risk is well known to practitioners who have observed users in the field as part of analysis and design activities. I have seen for many years that the initial interaction of search – selection of criteria or entering terms – is secondary to the interaction of *refinement*, where success relies on experience. "Satisficing" (as coined by Herbert Simon[1]) is a common user behavior – if the initial results seem sufficient, there is no need to apply further effort. Contextual observations and user interviews confirm that people prefer to do as little as possible initially to begin seeking information. Yet, when faced with results that don't meet their expectations, their motivation increases to use additional interaction devices and semantic aids to reach their goal.

THE CRITICAL ROLE OF CONTENT

> "Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing… The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain…"
>
> *Vannevar Bush, "As We May Think," 1945 (4)*

It is convenient, particularly in "open world" scenarios, to think of the challenge of exploratory search as a problem focused on metadata and interfaces. However, the success of exploration arises in the seamless flow between data and metadata, where a user searches, navigates, discovers, samples, interprets, qualifies and decides on the substance – the content – through the guiding framework of metadata interaction. It is an interesting and challenging question when designing information repositories and retrieval applications to identify where the metadata should *stop* and direct experience with the content should *start*, and then how the user's interpretation of that content can be captured to improve and inform any context/experience model used to support further exploration.

Certainly, simple examples of that seamless interaction exist already. The mSpace (5) preview feature is available at every level of exploration, allowing people to make value judgments without leaving the navigation context. It would be interesting to explore over time how the quality and "representativeness" of sound snippets influence the content decisions users then make. Another example appears in

---

[1] For more about Herbert Simon, see http://en.wikipedia.org/wiki/Herbert_Simon.

some Internet search engines[2], where users can expand/preview content directly in the results window, rather than leaving the results and having to "pogo stick" back and forth to explore a set of documents. While preview is clearly a style of interaction that supports exploratory behavior, and is convenient for users making decisions, it would be useful to evaluate the risks associated with people formulating judgments based on "sampling" or "nibbling" content rather than actually digesting it.

For example, a recent study in the Journal of American Medical Informatics Association (6) described the low incidence of successful confirmation of doctors' diagnosis resulting from searching for information on the Internet. Among the alarming findings was the fact that the percentage of preliminary diagnoses (based on experience) that *improved* as a result of searching for information was almost completely offset by the percentage of diagnoses that *went from correct to incorrect* as a result of searching and reading Internet-based information. While the focus of the study was not on the nature of the search (targeted, synthesis, learning, etc.), it potentially points to areas for further benefit from exploratory search.

SEVEN EXPLORATORY HABITS

While Steven Covey's original "Seven Habits" (7) focused on what people *should* do for effectiveness, the following habits are observed. They focus not on construction of the criteria for search, rather on how results are interpreted and collections managed over time. At their core are some behaviors that appear to need support for exploratory search to be meaningful. However, on their surface, the ways that users are currently using technology to achieve their goals are far from ideal.

There have been studies and design projects that have looked at aspects of this subject in the past (e.g. 3, 8, 9, 10), looking specifically at experienced user search, information foraging, following scents, etc. One question that may be valuable to explore further is "why do these insights tend to be slow to adopt in existing products and sites?" One reason may be commercial viability – there may not be enough financial/market incentive to radically change the existing interaction models. Another reason is that some of these behaviors are deeply rooted in habit, so it may be difficult to gain acceptance of novel interaction models.

The following are observations of typical scenarios illustrating users' exploratory search behavior, using existing technology, without necessarily having explicit exploratory support.

[2] See for example www.clusty.com (hourglass after link) or www.ask.com (binoculars before link).

1. Get the lay of the land

When authors prepare papers or when analysts begin to gather background information for a new project, they often start with very broad-based searches. The user goals are to survey the domain for new or unfamiliar information, and to begin to gather possible resources that could contribute to their main task: composition or synthesis.

One technique is to create a blank document that remains open during the process of searching. As interesting sites are found, the URL is copied from the address bar into the document. In most cases, a particular snippet of relevant text is also copied into the document just before or after the link (capturing context). Sometimes, but not always, an additional note is written by the user to describe why this site/snippet is interesting, or what subject it will relate to within their work. This research compilation can be saved in the same area as the draft being written, so it keeps the research process associated with the product.

Interesting maintenance behaviors begin to arise during and following the initial subject survey activity, such as managing references that at first looked interesting but later proved less useful – sometimes they are deleted, sometimes they are copied to the end of the document (for example, with a heading "unused"), and sometimes they are copied into a different document for later reference based on subject but outside the context of the current project.

2. Collect and compare useful resources

A comparison behavior is appearing as a result of browsers' tabbed interfaces. The user develops a habit of opening links in new tabs, rather than in the same or new windows. This harvests a set of resources from within one page/result set before moving to additional pages. The goal is to keep particular queries – or subjects – together when exploring and organizing possible resources for later use. After the harvesting activity comes the skimming process, where each harvested link is checked to see if it might be useful – at this point, the user may remove non-useful information from the collection by closing individual tabs. Comparison is very simple and tends to occur fairly often, switching between tabs to look back-and-forth between related documents. Digressions (at least among the more disciplined) may be handled by opening new windows, after which the process of collecting open tabs starts again on the new subject.

At this level, the "meta-model" of the subject and the information is reinforced by maintaining the group of items. However, there may be more that can be done to make that group identity explicit for future use, and to tie that session-level grouping into more persistent subject-level grouping. Another interaction that has been described as useful is easy annotation of judgments about each tab's reference, as part of the user's decision-making throughout the process.

3. Save and update information sets

Once a user has collected interesting information, it may be saved for later reference and then updated over time. For example, I have found in my own use of saved tabs that I may reopen the set of sites for further review and exploration (that is why I saved the set, after all). Within each tab I may then navigate to different pages, and even open new tabs for further exploration. At the end of a working session – then what do I do? Do I abandon the changed tabs? Do I save the tab set using the same name as I had used before, hoping that this new set will replace the old one… and do I still want all the open tabs for the future? Do I save them under a different name, and hope at some stage to find the time to clean out unused tabs and move them to be organized more effectively by subject?

Managing bookmarks and collections is known to be a growing problem. At the individual level, people like Tom Passin (11) have done interesting work harvesting bookmarks, then converting them into semantic data sets (RDF or Topic Maps), and even clustering them using standard clustering techniques to identify hidden patterns and relationships in the references. Tom began this experiment when he realized that it had become faster for him to search for a link than to look it up in his bookmarks.

Such bookmark management techniques have not yet been adopted by sites like del.icio.us, but it seems only a matter of time given the risk of decreasing usefulness in their flat tagging schemes. For example, in one recent session, I discovered a user with 44 stored links, applying 163 tags to describe those 44 links. Tags included both "workshop" and "workshops" as well as "semanticweb" and "semantic_web" – a range of semantic variation with no clear relationships. Looking then at the link lists for "semantic_web" I chose to quit after scanning lists of over 1,500 loosely organized links[3].

4. Identify answers in problem-solving sequences

Call center technicians and case workers in insurance/social services exhibit interesting synthesis habits. For example, over the course of resolving an issue by exploring information, ten documents may be reviewed. Document 3 might derail an initial hypothesis, sending the user down a completely tangential path to the original seeking trajectory. The important "aha!" moment might arise at documents 4 or 5. Document 4 might provide the first hint of the resulting solution, but only when combined with information from documents 5 and 8. Documents 9 and 10 might be viewed for negation – simply to rule out any secondary condition or situation, to reinforce the conclusion, and disprove alternatives.

[3] For examples, see http://del.icio.us/tag/semantic_web?page=16 and http://del.icio.us/peachris.

The immediate goal is to resolve the problem. However, the long-term goal is to recognize the pattern of information synthesis, in order to recall aspects of the solution in similar situations and to share knowledge with others. Support is needed within applications to expose and name patterns, as well as to contextually share those patterns with other users when they encounter similar situations.

5. Motivation informs strategy

The user's external situation that prompts exploration also has an impact on their evaluation of results and information. For example, a user who is face-to-face with a client needing the information being sought is likely to be motivated to find a short, targeted answer extremely quickly. In those cases, we have found that proactive delivery of information they may not know they need, based on the context of the situation, is extremely valuable to limit mistakes and improve quality of service.

In contrast, researchers or in-depth hobbyists have a different motivation focused on completeness. Their habits are more like casting nets and picking the best of each catch. However, they also retain a memory of what they have thrown back. When they stop seeing new or useful information, and begin seeing things that they have previously rejected, it is a sign that they may be nearing the end of the exploration.

In both cases, exploratory environments that can recognize the context – and thus to some extent the motivation of the user – might support users by adapting the tools, displays, and weighting of results to reflect each user's needs.

6. Revisit serendipity… a few days later

People sometimes come across something interesting that is out of context with their original exploration. They may allow the distraction to carry them down a small side path, before deciding they must return to their main task and complete it. However, some time later they may want to revisit that digression path that they discovered. At that point, they tend to struggle to remember key aspects of their original exploration, so they can use that thread of information to re-create the serendipitous discovery. Revisiting is an interesting secondary task to be considered when supporting exploratory search.

7. Take things out of context

The more granular the web becomes, the easier it is to find a snippet of information that appears interesting, and yet the user's interpretation may be misaligned with the intent of the author, due to a lack of context. One example of this arises in navigation of search results lists or long documents. Users scroll very quickly up and down, catching clues from bold text, highlighted phrases, and other markers. When they see something that seems to relate to their interest area, they focus on it. If it satisfies their need, the convenience of a link to click or selecting the back button means they often leave the page before seeing in detail what came before or after the snippet they selected.

In this case, aligning the goals of the author – comprehension of the message of the content, in context – with the goals of the user – easily and directly gaining the answer to a question – requires different kinds of signposting and support than is currently provided in the publishing-driven web. Perhaps when greater contextual details about the information can be compared with greater knowledge of the contextual environment of the user, mismatches can be highlighted in a way that alerts the user to the potential risk of misinterpretation.

CONCLUSION

> "Ask individuals to evaluate the relevance of search results, and their responses will vary according to what they already know and what they want to know. Even the same individual may evaluate the same results differently as her knowledge and interest changes over time… we should proceed with an understanding that relevance is subjective, situational, and dynamic."
>
> *Peter Morville, Ambient Findability, 2005, p 54 (12)*

The observations in this paper are written as a contribution to discussion about the types of situations that could benefit from enhanced support for user goals and exploratory tasks. Research, design, and development in the area of exploratory search must be grounded in a user-centered approach, in order to have lasting value as a guide to the next generation of interaction. Approaches for manipulating and holding information, paths, value judgments, and context will help improve the exploratory experience, and hopefully the targeted search experience, as well.

REFERENCES

1. Argamon, S. and Olsen, M. 2006. Toward Meaningful Computing. *Communications of the ACM, April 2006, 49(4),* 33-35.
2. iProspect and Jupiter Research. 2006. *iProspect Search Engine User Behavior Study*. April 2006. Online: http://www.iprospect.com/about/whitepaper_seuserbehavior_apr06.htm.
3. 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.
4. Bush, V. 1945. As We May Think. *The Atlantic Monthly, July 1945, 176(1),* 101-108.
5. mSpace. Online: http://mspace.fm/.
6. McKibbon, K.A. and Fridsma, D.B. 2006. Effectiveness of Clinician-selected Electronic Information Resources for Answering Primary Care Physicians' Information Needs. *Journal of the American Medical Informatics Association 13(4),* 653-659.
7. Covey, S. 1989. *The Seven Habits of Highly Effective People*. Free Press.
8. White, R., Kules, B., Drucker, S., schraefel, m.c. (Eds.). 2006. Special Issue: Supporting exploratory search. *Communications of the ACM,* 49(4), April 2006.
9. Wexelblat, A., Maes, P. 1999. Footprints: history-rich tools for information foraging. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 270-277.
10. Pirolli, P., Card, S., et al. There are a number of useful studies on information-seeking behavior using different types of visual displays, available in the ACM Digital Library between the years 1996-2006.
11. Passin, T. 2003/2004. Browser bookmark management with Topic Maps. In *Proceedings of Extreme Markup Languages 2003;* and On-the-fly Clustering as a Novel RDF Query Mechanism. In *Proceedings of Extreme Markup Languages 2004*.
12. Morville, P. 2005. *Ambient Findability*. O'Reilly.

Learning as a Paradigm for Understanding Exploratory Search

Bernard J. Jansen, Brian Smith, and Danielle Booth
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA, 16801, USA
jjansen@acm.org, bsmith@ist.psu.edu, dlb5000@psu.edu

Abstract
In order to investigate whether exploratory searching can be viewed as a learning process, we studied the searching characteristics of 41 participants engaged in 246 searching tasks. The searching tasks were classified according to Anderson and Krathwohl's Taxonomy of the cognitive learning domain, which is an updated version of Bloom's taxonomy. Results show that *Applying* takes the most searching effort in terms of queries per session, topics searched per sessions, and total time searching. Interestingly, *Remembering* and *Understanding* exhibit searching characteristics similar to the higher order learning of *Evaluating* and *Creating*. It appears that searchers rely primarily on their internal knowledge for *Evaluating* and *Creating*, using searching as fact checking and verification. Implications are the commonly held notion that Web searchers have simple information needs may not be correct. These simple searching expressions may be supportive of higher level information needs. We discuss the implications for exploratory searching interfaces and system design, including designing interfaces to support learning and exploration.

Authors Keywords
Information searching, Bloom's Taxonomy, Anderson and Krathwohl's Taxonomy

ACM Classification Keywords
H.3.3 [1] Information Search and Retrieval – *Search process.*

INTRODUCTION

How can we tell if someone is engaged in exploratory searching? A common paradigm for analyzing information searching problem solving or decision making. Donohew and Tipton [6, p. 251] state that information seeking research is intertwined with decision making. Other researchers have investigated aspects of information searching from a decision making or problem solving perspective [c.f., 3].

There has been, however, little empirical study of the relationship between information searching and problem solving. One exception is [7]. In this study, the researchers administered the Problem Solving Inventory (PSI) survey instrument to 40 participants of an information searching study. The PSI consists of a 35-item self-report measured in a 6-point Likert style format (strongly agree to strongly disagree). The PSI assesses an individual's perceptions of his or her problem-solving capabilities (i.e., a person's level of efficacy as a problem solver). Self-efficacy in a given area is correlated to actual performance [2]. The researchers report that there was no correlation between perceptions of problem solving ability and searching performance or perceptions of problem solving ability and exhibited searching characteristics. Given the prior work on information searching as a problem-solving activity [3], one would expect some relationship between problem-solving self-efficacy and searching performance [2].

We therefore explored other possible paradigms in which to view the information searching process – and thereby possibly exploratory searching -- namely as a learning activity. We define information searching, as the 'micro-level' of behavior employed by the searcher in interacting with information systems of all kinds [10, p. 49]. In the following, we present a brief literature review, research questions, preliminary results, and implications for online searching systems that support exploratory searching. We conclude with our future research aims and directions.

REVIEW OF LITERATURE

Information science and related literature contain references to an on-going learning process while a person is engaged in information searching. Zhang, Jansen, and Spink [11] showed that searchers exhibit linguistic characteristics that parallel learning of language formulation. Kuhlthau's Uncertainty Principle [9] includes a significant learning component.

Similar to the problem solving paradigm, there has been little research into how or even if learning explicitly manifests itself in the information searching process. Are there specific searching behaviors that one can map to a particular learning paradigm? It is this question which motivates our research.

RESEARCH QUESTIONS

Our research question is: *Is a learning paradigm effective for analyzing information searching?* In order to analyze this question, we used Anderson and Krathwohl's re-design of Bloom's taxonomy of learning in the cognitive domain to develop searching tasks for each of the six categories. We then analyzed the exhibited searching characteristics along the line of established information searching parameters to detect if there were differences in exhibited searching behavior.

METHODS

Bloom's Taxonomy is a primary classification of learning in the cognitive domain [4]. An updated version, Anderson and Krathwohl's Taxonomy, redefined Bloom's original classification [1]. Anderson and Krathwohl's Taxonomy is a six-tiered model for classifying learning according to cognitive levels of complexity. We devised searching scenarios for each of the six levels of Anderson and Krathwohl's Taxonomy, with each scenario correlated to one of the classifications. The scenarios were pilot tested twice before being used in a laboratory study. The six classifications with definitions [1, p. 67-68] and searching scenarios are shown in Table 1.

Over the course of a week, 41 participants engaged in a laboratory study. After they completed administrative and demographic paperwork, each participant was presented with the six searching scenarios and instructed to answer the questions and verify their answers. Each participant had access to an individual computer with Internet access. All user interactions with the computer were logged using a non-inclusive logging software package developed for use in information searching studies. At the completion of the study, we analyzed participant interactions in accordance with standard characteristics of information searching using transaction log analysis as the methodological approach. We analyzed the data using SPSS version 13.0.

RESULTS

Our research question is: *Is a learning paradigm effective for analyzing information searching?* Our three hypotheses addressing this research questions are:

Hypothesis 1. *There will be a significant difference in the number of queries per session among the classifications in Anderson and Krathwohl's Taxonomy.*

Hypothesis 2. *There will be a significant difference in the number of topics per session among the classifications in Anderson and Krathwohl's Taxonomy.*

Hypothesis 3. *There will be a significant difference in the duration of sessions among the classifications in Anderson and Krathwohl's Taxonomy*.

We define our terminology similar to that used in other Web transaction log studies [8]. *Query:* string of terms submitted by a searcher in a given instance. *Session:* series of queries submitted by a user during one interaction with the Web search engine. A topic is the information focus of one or more queries. A session may have several topics. The duration is the temporal length of session. We report our results in the following sections and examine if there is any difference among the six classifications of knowledge. We intend to examine other searching characteristics that are outlined in the future research section.

Queries Per Session

Our first hypothesis is: *There will be a significant difference in the number of queries per session among the classifications in Anderson and Krathwohl's Taxonomy*.

We used a one-way ANOVA statistical analysis to compare means and variance among the classifications. Our results indicate that there is a significant difference among the groups ($F(5) = 4.175$, $p < 0.01$; the critical value of $F = 2.29$). We ran a Tukey's HSD (honestly significantly different) test comparing group means.

Classification	Definition	Scenario
Remembering	Retrieving, recognizing, and recalling relevant knowledge from long-term memory	List 5 movies directed by Steven Spielberg.
Understanding	Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining	Give a brief plot summary of the TV show, Veronica Mars.
Applying	Carrying out or using a procedure through executing, or implementing	What are some possible characteristics of a person who would enjoy trip-hop music?
Analyzing	Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing	A certain television show contains intense violence and coarse language. Which rating should it receive?
Evaluating	Making judgments based on criteria and standards through checking and critiquing	Create a list of pros and cons for the new iPod Shuffle. Based off of this, would you purchase it (assuming you had the money)? Why or why not?
Creating	Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing	Which do you think will have better overall sales -- the XBox 360, the Nintendo Wii, or the Playstation 3? Why?

Table 1: Anderson and Krathwohl's Taxonomy with Searching Scenarios.

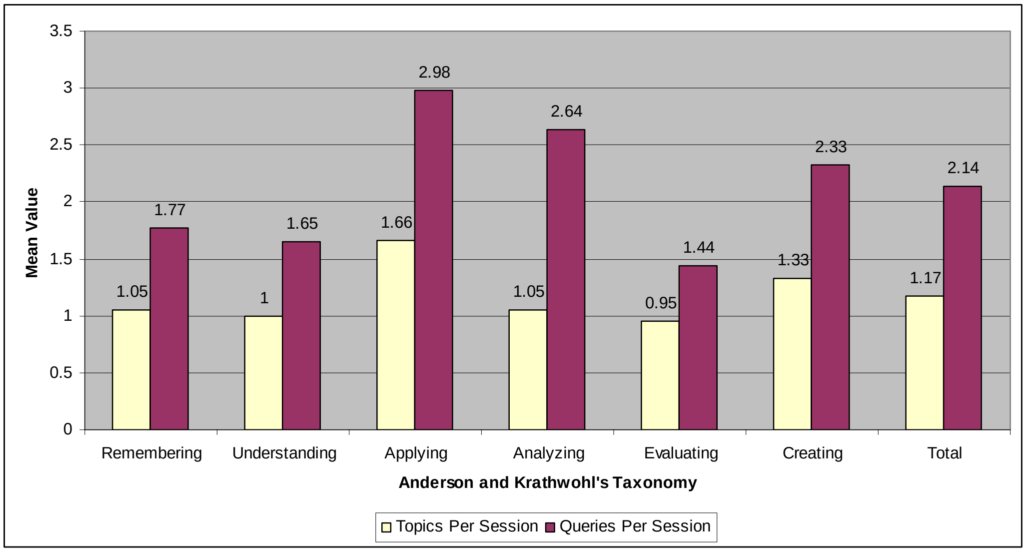

Figure 1. Queries per Session and Topics Per Session.

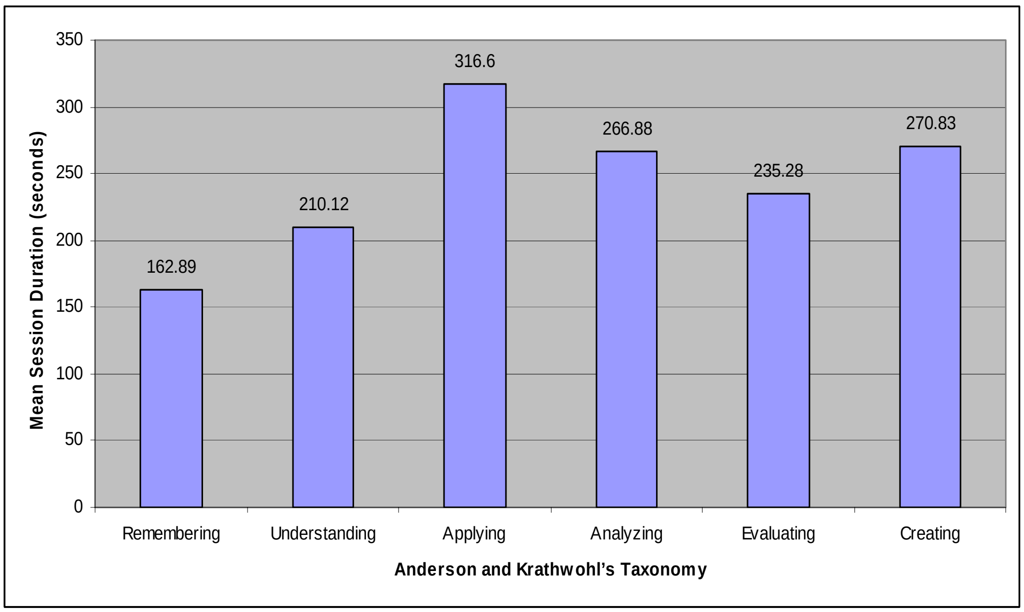

Figure 2. Session Duration in Seconds.

HSD results indicate that the classification *Applying* was significantly different then the classifications of *Remembering*, *Understanding*, *Analyzing*, and *Evaluating* ($p < 0.05$). *Applying* was not significantly different in number of queries per session than *Creating*. So, Hypothesis 1 is partially supported. Figure 1 show the mean queries per sessions of the six classifications.

Topics Per Session

Our second hypothesis is: *There will be a significant difference in the number of topics per session among the classifications in Anderson and Krathwohl's Taxonomy.*

Using a one-way ANOVAour results indicate that there is a significant difference among the groups ($F(5) = 6.869$, $p < 0.01$; the critical value of $F = 2.29$). Tukey's HSD results again indicate that the classification *Applying* was significantly different then the classifications of *Remembering*, *Understanding*, *Analyzing*, and *Evaluating* ($p < 0.05$). *Applying* was not significantly different in number of topics per session than *Creating*. So, Hypothesis 2 is partially supported. Figure 1 shows the mean topics per sessions of the six classifications.

Duration of Session

Our third hypothesis is: *There will be a significant difference in the duration of sessions among the classifications in Anderson and Krathwohl's Taxonomy*.

Again using a one-way ANOVA, our results indicate that there is a significant difference among the groups ($F(5) = 2.296$, $p < 0.05$; the critical value of $F = 2.29$). Tukey's HSD results indicate that the classification *Applying* was significantly different then the classification of *Remembering*. Hypothesis 3, therefore, is partially supported. Figure 2 shows the mean durations of sessions for each of six classifications.

DISCUSSION

We conducted a laboratory experiment to investigate whether or not one can view searching as a learning process. We utilized Anderson and Krathwohl's Taxonomy of the cognitive domain as the overall paradigm, designing searching scenarios consistent with each level of the taxonomy. We measured the differences in the searching characteristics of queries per session, topics per session, session duration among each classification. Results indicate that learning appears to be an appropriate paradigm in which to view searching. *Apply* appears to exhibit the most challenging searching context in all measured categories.

The classifications of *Applying* and *Analyzing* required more queries per session and more topics per session and took longer. These searching tasks exhibit searching characteristics that imply they are the most challenging for searchers. These levels may be the paradigm for identifying exploratory searching. Interestingly, the higher order learning of *Evaluating* and *Creating* exhibits characteristics generally similar to the lower order levels of *Remembering* and *Understanding*. While the higher order tasks are more difficult, they appear to depend more on the users' creativity and viewpoint. The additional knowledge searchers may need to complete the task are fact finding. Obviously, in these cases, users are missing out on serendipitous findings and alternative viewpoints. This would be a case for interfaces to facilitate exploratory.

The implications of this study are that commonly held notions of Web searchers have simple information needs [8] may not be correct. These simple searching expressions may be supportive of higher level information needs at the *Evaluating* and *Creating* levels of learning. Drawing on Kuhlthau's Uncertainty Principle [9], one can view uncertainty as an indicator of the presence of a learning opportunity. With this view, the goal of systems design shifts to one of not closing this uncertainty too quickly but instead accommodating exploring and formulating.

Conclusion and Future Research

People seek information for a variety of reasons, and our research has tried to categorize some of these reasons in terms of learning needs. We will use the results of this study to design search interfaces that facilitate different information searching modes, including exploratory searching. For instance, providing tools to collect and annotate findings might enhance *Applying* tasks as searchers could develop their arguments during the search process. Similarly, we could present people with multiple perspectives on an argument, along the lines of [5], in order to enhance their viewpoints when engaged in *Evaluating*. We envision search interfaces assuming various modes that correspond to searcher learning goals or intentions.

References

1. Anderson, L. W. and D. A. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman, 2001.

2. Bandura, A., "Self-efficacy," in *Encyclopedia of human behavior*, 4, V. S. Ramachaudran, Ed. New York: Academic Press, 1994, pp. 71-81.

3. Belkin, N., "On the nature and function of explanation in intelligent information retrieval," in *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*, Grenoble, France, 1988. pp. 135 - 145.

4. Bloom, B. S. and D. R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a Committee of College and University Examiners. Handbook 1: Cognitive Domain*. New York: Longmans, 1956.

5. Budzik, J., K. J. Hammond, L. Birnbaum, and M. Krema, " Beyond similarity," in *Working Notes of the AAAI 2000 Workshop on Artificial Intelligence for Web Search*. Austin, TX. USA, 2000.

6. Donohew, L. and L. Tipton, "A Conceptual Model of Information Seek, Avoiding, and Processing," in *New Models for Mass Communication Research*, P. Clarke, Ed. Beverly Hills, CA: Sage, 1973, pp. 243-269.

7. Jansen, B. J. and M. D. McNeese, "Evaluating the Effectiveness of and Patterns of Interactions with Automated Searching Assistance," *Journal of the American Society for Information Science and Technology*, 56, pp. 1480-1503, 2005.

8. Jansen, B. J. and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing & Management*, 42, pp. 248-263, 2005.

9. Kuhlthau, C., "A Principle of Uncertainty for Information Seeking," *Journal of Documentation*, 49, pp. 339-355, 1993.

10. Wilson, T. D., "Human Information Behavior," *Informing Science*, 3, pp. 49 - 55, 2000.

11. Zhang, M., B. J. Jansen, and A. Spink, "Information Searching Tactics of Web Searchers," in *the Annual Meeting of Society for Information Science and Technology*. Austin, Texas, 2006.

Methodology for Capturing Exploratory Search Processes

Makiko Miwa
National Institute of Multimedia Education
2-12 Wakaba, Mihama-ku, Chiba-Shi, Japan
miwamaki@nime.ac.jp

Noriko Kando
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
Kando@nii.ac.jp

ABSTRACT

This paper reports on methodological propositions for capturing exploratory search processes on the Web in the domain of history and geography based on a pilot study of a series of experiments. We propose three new methods for improving the capture of exploratory search processes from the searcher's perspective: (1) using "view" as a unit of data collection and analysis to capture physical behavior as well as the internal states of the searcher; (2) inviting and encouraging pairs of participants to talk aloud during searching; and (3) using an eye-tracker to capture search processes and show participants their recorded search processes to support their recall. Implications for several aspects of a Web user interface design intended for exploratory searches are also discussed.

Author Keywords

Exploratory search process, naïve ontology, taxonomy of knowledge modification, methodology, eye-tracking.

ACM Classification Keywords

H52 User Interfaces: Evaluation/methodology.

INTRODUCTION

Descriptive studies for capturing and analyzing exploratory Web search processes call for optimal methodologies that enable us to capture observable behavior as well as internal (cognitive and affective) processes of searchers [1]. However, this is not an easy task since most of the Web search processes are embedded in everyday life activities and searchers often have difficulty recalling and articulating what is happening or happened in their mind while they look for information and build knowledge.

To date, we have been using with some limited success a combination of observing (video recorded) search processes, the elicitation of a think-aloud protocol, and post hoc interviewing to reconstruct in as much detail as possible the micro level internal processes.

In this paper, we report a pilot study in which we conducted a series of experiments to seek an optimal method for identifying characteristics of exploratory search processes using the concepts of time and space. We chose the topics of history and geography as domain instances and invited fifteen students (undergraduates and graduates) as novice domain searchers.

CONCEPTUAL FRAMEWORK

In the following we define two terms, "a view" and "naïve ontology," to describe our intention to improve our methodology in capturing the micro level processes of exploratory search processes as well as in expanding our understanding of knowledge-building behavior of searchers.

A View as a Unit of Analysis

We used the concept of "a view" to develop the methodology for this study. Kwasnik [2] defined the notion of "a view" as "what a person articulates as seeing at one time [in his/her physical as well as cognitive world], that is, a span of attention" and proposed to use it as a unit of analysis in browsing. By expanding physical eye movements into the internal states of cognitive and affective worlds, a view could represent a synthesized chunk of external and internal behavior at each point of move in exploratory search processes. We thus introduce the notion of "view" as the unit of analysis for holistically capturing and analyzing exploratory search processes.

Naïve Ontology for Designing Interface for Domain Novices

The increasing availability of and accessibility to Internet resources by the general public allows domain novices to interact with unfamiliar document collections. In response to such trends, the main role of ontology seems to be shifting from a search aid to a navigation tool. "Naïve ontology" in this study is a conceptual framework for the type of information access interface that allows domain novices to refine their knowledge interactively by acquiring chunk by chunk information as they encounter in digital environments. Specifically, this is a framework for designing a kind of navigation tool that supports seamless switching between searching and browsing to acquire and build basic domain knowledge that enables domain novices to explore domain collections. We introduce the concept of naïve ontology to advocate the needs for developing a new type of information access interface intended to support domain novices' exploratory searches.

METHODOLOGY

We invited a total of fifteen novice searchers (seven undergraduate and eight graduate; ten female and five male students) to participate in a series of three experiments. We used an eye tracker to record searchers' eye movements and a HyperCam to capture screen shifts and talk aloud protocol. We showed the recorded eye movements superimposed on the screen shifts during the post search interviews to elicit searchers' views on each move during the exploratory search processes. We used a slightly different procedure in each of three experiments, as depicted in Figure 1.

Pre-search Activities

Three female college students participated in the first experiment. Each participant completed a short questionnaire that asked the frequency of their Internet use, favorite browsers and search engines, topics chosen at the college entrance examination, whether planning to get a teacher's certificate, and topics of interest in history and geography, in addition to demographic data. Each participant was then given two vignette scenarios prepared by the author. The first scenario asked the participant to conduct a Web search to prepare for teaching a middle school history class on a particular topic; the second scenario asked the participant to conduct a Web search to prepare for visiting a world heritage site.

In the second experiment, we invited four female participants in two pairs and conducted the same questionnaire. Then we asked each pair to discuss their topics of interest in the domains of history and geography for about ten minutes to let them choose two topics for a Web search. Each participant had either a history or a geography topic. This modification in the procedure was introduced since each participant in the first experiment was tense and nervous because s/he was required to wear unfamiliar devices, sit alone in a laboratory environment, and be observed by two researchers and a research assistant. This was also because one participant was unable to make progress in her exploratory search because of the confusion caused by encountering incomprehensible information.

In the third experiment, we invited three female and five male students. We conducted the same questionnaire with eight participants in four pairs. We then asked them to discuss their topics of interest in the domain of history and to choose two topics for a Web search, one for each participant in each pair. In addition, we gave a vignette scenario that asked the participants to conduct a search to prepare for teaching a middle school history class on any topic. The modification was introduced to capture self generated and imposed exploratory search processes.

In all three experiments, participants were asked what and how much they already knew about the topic of the search.

Recording of Exploratory Search Processes

In the first and second experiments, screen shifts and mouse movements were captured using a HyperCam. Participants' eye movements were captured using the EMR NL8 eye tracker which required fixing searchers jaw on the stand so they could not move their head.

In the third experiment, we captured participants' responses to a talk-aloud protocol that incorporated screen shifts and mouse movements using a HyperCam. This modification was possible because we used a new eye tracker VOXER which allowed searchers to move their head to enable them to talk freely. Each session was conducted over 15-20 minutes.

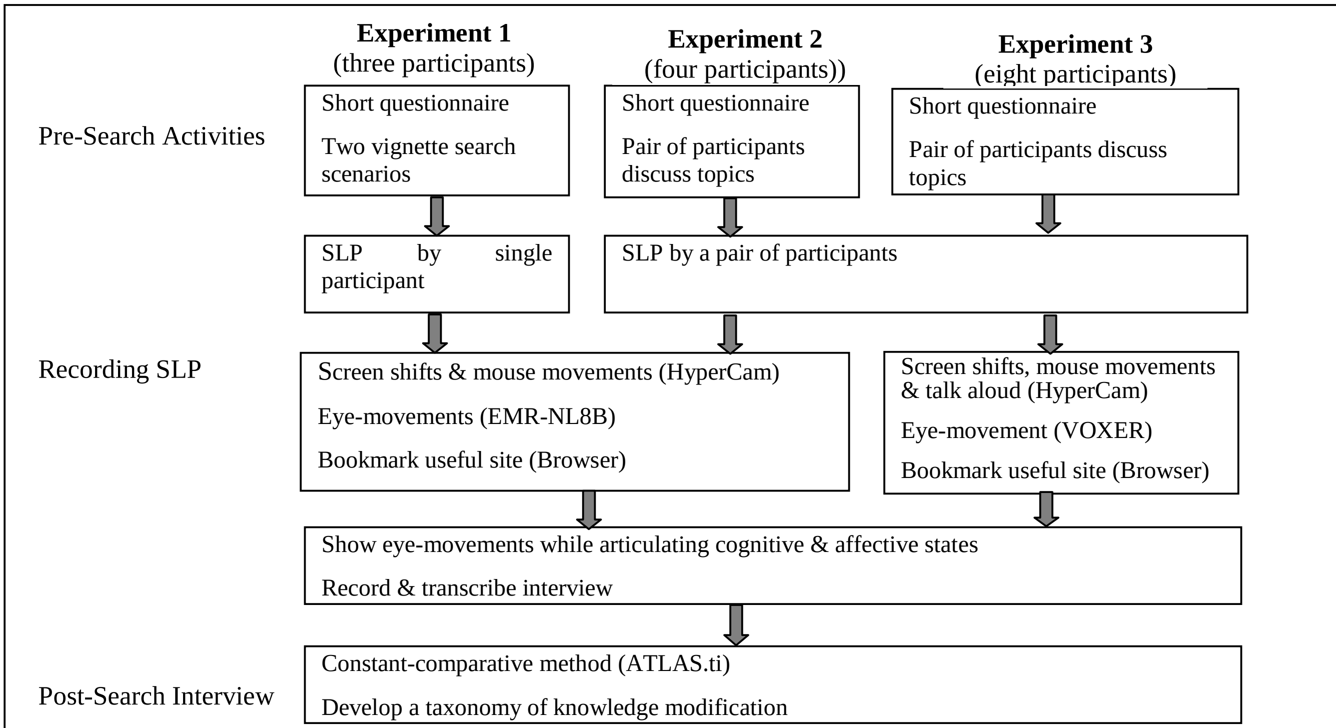

Figure 1: Procedure.

Post-search Interview

After each exploratory search session, participants were interviewed. During the interview, we showed the video recorded eye movements superimposed on the screen to elicit the searcher's view at each move. Specifically, we stopped the video at each move (typing, clicking, browsing, copying, pasting, bookmarking, and stopping), and asked the searcher reasons for the move as well as accompanying thoughts and feelings at that moment.

Data Analysis

Both pre search activities and post-search interviews were voice recorded and transcribed in detail. For the third experiment, talk aloud protocols during the exploratory search processes were also transcribed and compared to the post search interviews to increase reliability.

The transcribed interview data were coded for knowledge acquisition using ATLAS.ti , software for qualitative data analysis, following a bottom up strategy with the constant comparative technique.

RESULTS

In the post search interview, participants articulated what they thought, felt, expected, and intended to do at each point of move, either voluntarily or in response to the interviewer's inquiries. Analysis of the interview data lead us to develop a taxonomy of knowledge modification during the participants' exploratory search processes (Table 1). A detailed description of the taxonomy is reported elsewhere [3,4]. This taxonomy helps to increase our understanding on exploratory search processes in terms of how searches knowledge is modified while searching for information, which should eventually lead us to develop a naïve ontology.

IMPLICATIONS

The reported pilot study provides us with some methodological implications for capturing exploratory search processes. The study also provided us with useful information for developing naïve ontology for designing an information access interface intended for domain novices.

Methodological Implications

We perceived that three new methods implemented in this study were helpful in increasing the reliability of a descriptive study of exploratory search processes. They are: use of a "view" in data collection; showing eye movements in post search interview; and pairing of participants.

The use of an eye tracker in data collection and a "view" as the unit of analysis for exploratory search processes provided us with a coherent picture of searchers thoughts, feelings, and physical movements. First of all, we were convinced that the use of a "view" as a unit of analysis helped us capture the dynamically changing knowledge structure of domain novice searchers together with the ir externally observable behavior during their exploratory searches. In the post search interviews, participants reports of what they thought, felt, and had done related what they saw on the screen with their eye movements. They said the eye movements shown on the screen pointed to the exact places that they were looking at and that watching their own eye movements helped them recall their search processes. Thus, the method of using an eye tracker to collect data on exploratory Web search processes and showing the

Type	Definition	Example of Expressions*
Adding	Acquire novel information to increase knowledge	I didn't know that the origin of Bento is in Edo culture. I have never thought about it.
Correcting	Clean up a mis-understanding	I did not expect Johannes Brams to be classified here. I thought Brams is in a bit earlier in chronology, early romantic music…
Limiting	Narrow down the scope of the concept	When I examined the text, it was about the Renaissance era which is quite different from the industrial revolution I was looking for… I thought industrial arts may have influenced over civil life.
Relating	Relate a concept with another concept	It said that the admission ticket for Akropolis Museum and Partenon Temple are unified. So I thought the days for both of them is closed are the same. ..
Specifyin g	Increase the depth of meaning of the concept by increasing specificity	It was written in the Christian calendar for the beginning and the end of Taisho era
Transform ing	Understand a concept in a different framework	I did not know the experiential learning programs are offered by the [Edo Tokyo Museum]… I thought it's a pure museum

Table 1. Taxonomy of Knowledge Modification (*translated from Japanese).

recorded eye movements during the post search interview may be recommended for future studies. They will not only identify what searchers are looking at rather than what data are shown on the screen, but will also increase the reliability of data on internal processes of searchers, which eventually increases the reliability of research findings.

The method of pairing participants is helpful to elicit natural conversation during exploratory search processes as well as during post search interviews. This procedure may be particularly useful in Japan where people are shy about expressing their thoughts and feelings in front of strangers. In a series of three experiments, we used slightly different data collection procedures. In the first experiment, we invited three participants individually and asked each of them to conduct searches. During the post search interviews, participants explained their behavior, but seemed hesitant to articulate their thoughts and feelings. In the second and third experiments, we invited participants in pairs by asking each of them to bring an intimate friend as a partner to facilitate verbal communication. During exploratory search processes, one of the pair conducted the search and the other gave suggestions. At the post search interview, the participant who conducted the search articulated what s/he did, thought, and felt, while the other participant of the pair shared her/his own perspectives. We perceived that the paring of participants helped them articulate their thoughts and feelings more freely not only during the post search interviews but also during exploratory search processes, and provided more detailed data on their search processes compared to the first experiment .The presence of an intimate friend seems to relax participants even in the unfamiliar lab environment and leads them to a more naturalistic articulation than during one-to-one interviewing with the researcher. Naturally generated free flowing conversation between the pair during the search processes for the second experiment led us to introduce a voice recording of think aloud protocol at the search processes in the third experiment and triangulate it with post search interview data to improve the reliability of the interview data.

Implications for Interface Design

Elicitation and analysis of novice searchers' exploratory search processes lead us to make design suggestions for an information access interface intended for domain novices as synthesized in Table 2. Some of the functionalities proposed in Table 2 have been reflected in recently developed navigation tools, such as the Flamenco system [5] and Junii+ [6] that seamlessly combine browsing and searching. As demonstrated in these tools, an information access interface to support exploratory search processes is expected to be naturalistic for novice searchers to enable them to follow their own "view" of the world developed through everyday life experiences without being disturbed by the mandatory use of unfamiliar ontology intended for domain experts.

Aspect	Implications
Content	provide high-quality pictures and visual images
	present correlations among people and incidents using graphics
	provide dictionary of historical figures and events
	present the content in a simple manner
	use colors to categorize texts
	provide glossary of technical terms
Browsing	enable to browse time and space simultaneously or interchangeably
	enable to browse time and space from broad to specific
	enable to trace and overview historical transitions of people, culture, religions, and political rulers of each geographic area or country
	enable to choose pictures from among thumbnails
Navigation	provide metadata on genre of contents
	allow switching between different calendar systems
	provide links to contemporaneous figures and incidents
	provide historical and geographic frameworks people and incidents
	provide links to free online translation services or enable to switch languages

Table 2. Implications for Interface Design.

CONCLUSION

We conducted a series of three experiments in a case study in order to seek for optimal method for identifying characteristics of exploratory search processes, using the topic of history and geography as domain instances and invited fifteen students as novice domain searchers. We used a slightly different procedure for each experiment to seek an optimal method.

The results we obtained have useful implications for design of naïve ontology as the bases for an information access interface intended for novice domain searchers. The study also provided us with implications for three new methods for capturing exploratory search processes: the use of a "view" as a unit of analysis that may enables us to capture both physical and cognitive processes together in a reliable manner; the use of an eye tracker in capturing the exploratory search processes and showing eye movements superimposed on screen shifts to participants in post search interviews which seemed to help participants recall their internal processes; and pairing of participants, which seemed to help them freely articulate their thoughts and feelings, leading to elicitation of a rich and thick description of the processes.

FUTURE RESEARCH

We consider this paper as describing an initial step in developing naïve ontology. We will continue this research project by recruiting more participants. We hope to refine naïve ontology by refining the taxonomy of knowledge modification as well as exploring how such knowledge is used during exploratory search processes. Only after obtaining sufficient empirical evidence concerning relationships between acquisition and usage of knowledge may hypotheses be developed to test the ability to generalize these relationships as a base for developing naïve ontology.

Simultaneously, we need to verify the effectiveness of each of the three new methods by comparing the effects of their presence and absence on exploratory search processes.

ACKNOWLEDGMENTS

This research was funded by the National Institute of Informatics joint research grant, and we wish to thank Dr. Barbara Kwasnik at Syracuse University for her helpful suggestions and the notion of "view" she proposed as the unit of analysis for browsing, which lead to the research design of this study.

REFERENCES

1. Marchionini, G. Exploratory searching: From finding to understanding. *Communication of the ACM, 49,* 4 (2006), 41-46.
2. Kwasnik, B. A descriptive study of the functional components of browsing. Working Conference on Engineering for Human-Computer Interaction (August 10-12, 1992), In J. Learson & C. Unger (Eds.), IFIP Transactions, Proceedings of 5th IFIP Working Conference on User Interfaces. A-18, North-Holland, (1992), 191-203.
3. Miwa, M., & Kando, N. Role of naïve ontology in search and learn processes for domain novices. In S. Sugimoto., J. Hunter., A. Rauber., & A. Morishima (Ed.), *Digital Libraries: Achievements, Challenges and Opportunities, 9th International Conference on Asian Digital Libraries, ICSADL 2006 Kyoto, Japan, November 2006, Proceedings* (pp. 380-389). Berlin, Germany: Springer.
4. Miwa, M., & Kando, N. Naïve Ontology for Concepts of Time and Space for Searching and Learning. *Information Research 12,* 2. (2007). http://informationr.net/ir/12-2/paper296.html
5. Hearst, M., English, J., Shinha, R., Swearingen, K. & Yee, P. Finding the flow in web site search. *Communications of the ACM, 45,* (2002), 42–49.
6. Kando, N., Kanazawa, T. & Miyazawa, A. Retrieval of web resources using a fusion of ontology-based and content-based retrieval with the RS vector space model on a portal for Japanese universities and academic institutes. *Proceedings of the 39th Hawaii International Conference on System Science*, Kauai, HI, 2006. [CD-ROM].

CONCLUSION

We conducted a series of three experiments in a case study in order to seek for optimal method for identifying characteristics of exploratory search processes, using the topic of history and geography as domain instances and invited fifteen students as novice domain searchers. We used a slightly different procedure for each experiment to seek an optimal method.

The results we obtained have useful implications for design of naïve ontology as the bases for an information access interface intended for novice domain searchers. The study also provided us with implications for three new methods for capturing exploratory search processes: the use of a "view" as a unit of analysis that may enables us to capture both physical and cognitive processes together in a reliable manner; the use of an eye tracker in capturing the exploratory search processes and showing eye movements superimposed on screen shifts to participants in post search interviews which seemed to help participants recall their internal processes; and pairing of participants, which seemed to help them freely articulate their thoughts and feelings, leading to elicitation of a rich and thick description of the processes.

FUTURE RESEARCH

We consider this paper as describing an initial step in developing naïve ontology. We will continue this research project by recruiting more participants. We hope to refine naïve ontology by refining the taxonomy of knowledge modification as well as exploring how such knowledge is used during exploratory search processes. Only after obtaining sufficient empirical evidence concerning relationships between acquisition and usage of knowledge may hypotheses be developed to test the ability to generalize these relationships as a base for developing naïve ontology.

Simultaneously, we need to verify the effectiveness of each of the three new methods by comparing the effects of their presence and absence on exploratory search processes.

ACKNOWLEDGMENTS

This research was funded by the National Institute of Informatics joint research grant, and we wish to thank Dr. Barbara Kwasnik at Syracuse University for her helpful suggestions and the notion of "view" she proposed as the unit of analysis for browsing, which lead to the research design of this study.

REFERENCES

1. Marchionini, G. Exploratory searching: From finding to understanding. *Communication of the ACM, 49,* 4 (2006), 41-46.
2. Kwasnik, B. A descriptive study of the functional components of browsing. Working Conference on Engineering for Human-Computer Interaction (August 10-12, 1992), In J. Learson & C. Unger (Eds.), IFIP Transactions, Proceedings of 5th IFIP Working Conference on User Interfaces. A-18, North-Holland, (1992), 191-203.
3. Miwa, M., & Kando, N. Role of naïve ontology in search and learn processes for domain novices. In S. Sugimoto., J. Hunter., A. Rauber., & A. Morishima (Ed.), *Digital Libraries: Achievements, Challenges and Opportunities, 9th International Conference on Asian Digital Libraries, ICSADL 2006 Kyoto, Japan, November 2006, Proceedings* (pp. 380-389). Berlin, Germany: Springer.
4. Miwa, M., & Kando, N. Naïve Ontology for Concepts of Time and Space for Searching and Learning. *Information Research 12,* 2. (2007). http://informationr.net/ir/12-2/paper296.html
5. Hearst, M., English, J., Shinha, R., Swearingen, K. & Yee, P. Finding the flow in web site search. *Communications of the ACM, 45,* (2002), 42–49.
6. Kando, N., Kanazawa, T. & Miyazawa, A. Retrieval of web resources using a fusion of ontology-based and content-based retrieval with the RS vector space model on a portal for Japanese universities and academic institutes. *Proceedings of the 39th Hawaii International Conference on System Science*, Kauai, HI, 2006. [CD-ROM].

Evaluating an automatically constructed hypertext for improved searching

Michael Huggett **Joel Lanir**
Department of Computer Science
University of British Columbia
2366 Main Mall, Vancouver BC, Canada
{mike, yoel}@cs.ubc.ca

ABSTRACT

Finding related items easily should be possible even in large data repositories. Automatically-constructed hypertext (ACH) allows large numbers of documents to be linked faster and more consistently than manually possible. With a single link type based on semantic similarity, the result is a *similarity-ACH*. We believe that similarity-ACH can be used in exploratory search systems to improve searching in large heterogeneous document collections. In a user study, we supplement standard keyword search with navigation in a similarity-ACH. Using novel task metrics, we find that the addition of an ACH significantly improves user search behaviour and information finding, particularly for search tasks that are general or vague.

INTRODUCTION

Keyword search by itself is not enough. While keyword search is fast when the target of a search is known, it fails when users do not have and clear idea of what they seek. Users may be unfamiliar with a topic's vocabulary, or may be unskilled at composing accurate queries. Vague searches may return too many results, while over-specific queries may return too few. Keyword search also places a burden of decision on users to reformulate queries until they find what they want. Even with a good result in hand, with current search tools it is often difficult to find more related items. In other words, users seldom have the opportunity to say to their systems, "give me more like *this* one".

Humans are good at exploratory wayfinding. Such local sign-following is based on nearby cues in the environment, and occurs quickly since these cues are usually present [4]. When wayfinding through hypertext, users are likely to pick a next step that appears to bring them closer to their intended target, based on desirable attributes (such as keywords) visible in currently-retrieved objects [11]. Such iterative searching approaches items of interest in a semantic gradient descent that quickly narrows in on relevant items.

Ideally, information systems will suggest next steps to users, reducing the need for reformulation of direct queries. Keyword search is difficult, and refining queries can require significant user competence [11]. Keyword search also has semantic limits: users typically formulate queries with fewer than 3 keywords [5]; the inherent ambiguity of short queries limits the precision of the results, and precision also falls as the size of a corpus grows.

However, keyword searches can be a good jumping-off point for browsing. With a similarity hypertext, queries place users in the vicinity of targets, and once one target is found, browsing will lead to adjacent good targets. The similarity hypertext "suggests" next steps based on the similarity scores of adjacent documents. If the suggested documents have high similarity scores, users may be confident that items linked to them in a similarity-based hypertext may lead to other good results. Evidence shows that users react positively to link suggestions and navigate in a more structured way, reducing disorientation and task execution time [6].

In real-world conditions, users typically precede browsing with a keyword search. Although some studies [e.g. 3] seek to examine "pure" navigation with browsing only, this clearly does not simulate real-world conditions. Neither browse nor search is sufficient by itself to fulfill complex information tasks [7, 11], but each has its strengths. Keyword search is fast and "good enough", but ill-articulated information needs cause users to browse [7]. Hypertext works well for vague or difficult searches, as navigation through the semantic structure permits formulation of a query by identifying a semantic path.

Automatically-constructed hypertext (ACH) provides a navigable structure for exploratory search, with links that users can follow between nodes. ACH has been generated from document corpora in many ways. Most studies assume that building an ACH allows users with to browse primarily through documents based on semantic similarity. ACH are most commonly constructed using statistical measures of word similarity between documents. Documents are represented as vectors of weighted terms: if the vectors match well enough, then the documents are linked. There are many methods to determine inter-document similarity. Our position is that any method that produces a monotonic ranking of document similarities based on keyword matching will be sufficient for evaluating similarity-ACH.

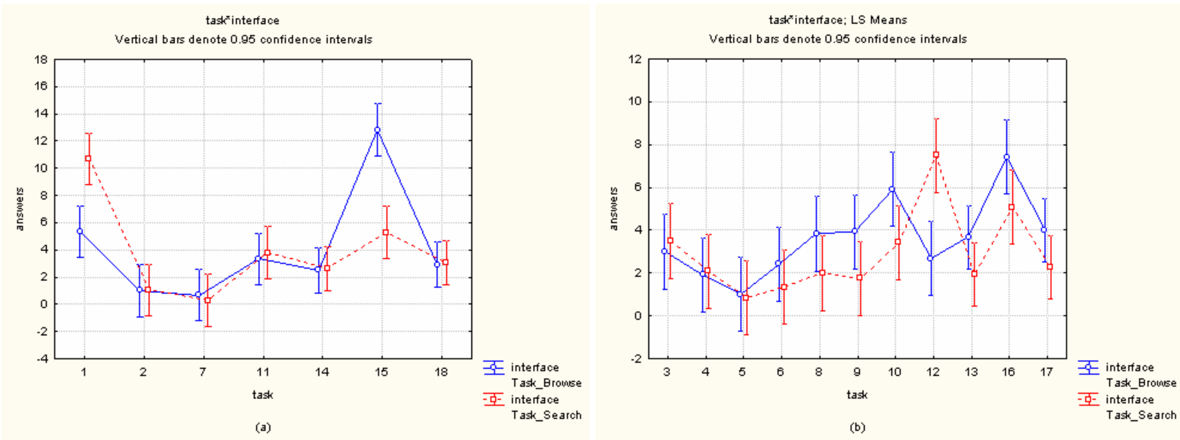

Figure 2 – Number of correct selections by subjects using the Browse and Search interfaces, for direct (a) and indirect (b)

For example, the direct task "Find all articles that discuss the jailing of Judith Miller" requires little more than a search on the terms "Judith" and "Miller" to find all related articles. On the other hand, the indirect task "Find articles that talk about financial predictions of the future" is not easy to execute using keyword search alone.

We believe that our approach, a user study based on a variety of realistic task types, is generally applicable to evaluation of exploratory search systems.

RESULTS

We ran a two-way ANOVA with repeated measures in order to examine the effect of interface type on user performance, with task and interface as the within-subject variables. The dependent variable reported in all the analyses was each subject's score result of number of correct answers on each task.

The results show that across all tasks subjects performed better in the Browse interface (mean: 3.75) than in the Search interface (mean: 3.2) {$F(1,22)=6.01$, $p<0.05$}. The results also show an interaction effect between task and interface {$F(1,22)=6.005$, $p<0.01$} indicating that scores in each interface depend on the task given.

To further compare subject performance in both interfaces, and to investigate the interaction effect, we performed a task by task analysis on our measures of *directness* and *extent*. A 2-way ANOVA was run on the direct and indirect tasks in order to examine the effect of directness on task performance in both interfaces. Results indicate that indirect-task performance was better in the Browse interface {$F(1,22)=4.50$, $p<0.05$} while for the direct tasks no significant difference was found. Thus although the Browse interface showed better performance on all tasks, it was more beneficial to the indirect tasks. Figure 2 shows a task-by-task comparison of both direct (2a) and indirect (2b) tasks.

We further compared tasks according to the number of targets for each task. Performance in the high-extent tasks was not significantly better for either interface (mean 6.19 for the Browse interface, and mean 5.92 for the Search interface), while in the low-extent tasks the Browse interface showed better performance {$F(1,22)=16.93$, $p<0.01$}. These results suggest that the Browse interface was especially beneficial for low-extent tasks, i.e. tasks with few target answers.

We compared the Search and Browse interfaces to estimate the information gain in each. Subjects performed an average of 4.03 queries per task in the Search interface. In the Browse interface subjects performed an average of 4.72 combined retrieval operations (i.e. both queries and navigations).

66.8% (320/479) of navigations were through documents already selected as answers. This surprisingly high number indicates that users most often used navigation to retrieve results similar to a known strong correct result. Open-ended answers in the post-questionnaire suggested another navigational behaviour. Some users indicated that they used the navigation button when they didn't get enough strong results in their initial search. Instead of refining their search, they used the navigation button to browse from the most likely-looking result to "home in" on good targets.

DISCUSSION

Over-all, there was a clear advantage in using the Browse interface, compared to the Search interface. The navigation option in the Browse interface is most helpful for indirect tasks, and for tasks of low extent. There are two main reasons why the Browse interface would be better for indirect tasks.

First, the browse condition provides results of greater relevance. Whereas the search condition only shows users what they have asked for directly with their queries, in the browse condition users have the advantage of additional relevance information. A known good document is likely to be linked to other relevant documents. Also, subjects using the Search interface may feel that they have exhausted their search terms and stop seeking more information, while with the Browse interface it is easier for subjects to continue with their search.

Second, the browse condition is less effortful. It gives users direct suggestions on possible avenues for exploration, befitting human wayfinding [4]. In the Search interface subjects must perform their own iterative query refinement. By contrast, in the Browse interface subjects are less burdened with query refinement and instead can follow their intuition to explore the information space along promising trails. Navigation exposes them faster to a greater amount of related information, which reduces the time required to find accurate results. Users take advantage of the implicit keywords inherent in the links of the similarity-ACH, as the system makes ‘suggestions’ for documents that seem reasonable based on the semantic content of the corpus. Subjects can then examine the descriptions and content of the suggested documents to discover other relevant terms that can be used in further keyword searches.

To understand better the advantages and disadvantages of the similarity network, tasks which showed worse performance for the Browse interface offer some useful clues. Only two tasks showed significantly better performance for the Search interface in the task-by-task analysis. The description for Task 1 was: “find articles which mention the jailing of reporter Judith Miller”. This is an excellent example of a direct task. Almost all other direct tasks showed no differences between Search and Browse interfaces, but here the search terms Judith and Miller are highly appropriate and specific, and any use of navigating the similarity network merely delays and distracts subjects from the better tactic of reformulating queries around those two terms.

Task 12 was described as “find articles which mention one company referring to another company”. We judged it as an indirect task because it doesn’t directly suggest any obvious search terms (unlike Task 1). Given that the relationship between two companies could be described in any number of ways – by mergers, lawsuits, subcontracting, competition or cooperation, etc. – there was less likelihood that target articles would share keywords, and thus less likelihood that any two target documents would be linked. This would render the similarity-based navigation option less effective, as finding one target would not easily lead to more. Thus, similarity browsing will fail when few target articles share common terms within the scope of a task.

CONCLUSION

The advantages of browsing compared to keyword search for hypertext are already well-known, and have been demonstrated primarily with user studies using manually-built networks (e.g. the Web). We have shown that an automatically-constructed hypertext based on similarity (similarity-ACH), constructed from a heterogeneous document corpus using simple tools, produces similar results in simulated real-world exploratory search tasks. Our user study confirmed that browsing in similarity hypertext is effective for tasks that have few targets in the corpus (low *extent*), and for tasks that are vaguely described (low *directness*). Results also indicated that there was more activity (i.e. more results were retrieved) in the browsing condition, and that users were exposed to more information.

We suggest that exploratory search tools that use automatically constructed similarity hypertext as a basis for browsing in large document collections can substantially enhance user experience and improve the overall quality of results.

REFERENCES

[1] Bhavnani, S.K., Drabenstott, K., and Radev, D. (2001) Towards a Unified Framework of IR Tasks and Strategies. In *Proc. ASIST'2001*, 340-354.

[2] Carmel, E., Crawford, S., and Chen, H. (1992) Browsing in hypertext: A cognitive study. *IEEE Transactions on Systems, Man, and Cybernetics*. 22 (5), 865-83.

[3] Green, S.J. Building Hypertext Links by Computing Semantic Similarity. *IEEE Transactions on Knowledge and Data Engineering*, 11 (5), 1999.

[4] Hunt, E. and Waller, D. (1999) Orientation and wayfinding: A review. Technical Report N00014-96-0380, Arlington, VA. Office of Naval Research.

[5] Jansen, B.J. and Spink, A. (2006) How are we searching the world wide web? *Information Processing & Management*, 42 (1), 248-263.

[6] Juvina, I., & Herder, E. (2005) The Impact of Link Suggestions on User Navigation and User Perception. In *Proc.UM2005*, 483-492.

[7] Marchionini, G. *Information seeking in electronic environments*. Cambridge University Press, 1995.

[8] Marchionini, G., Dwiggins, S., Katz, A., and Lin X. (1993) Information Seeking in Full-Text End-User-Oriented Search Systems, *Library and Information Science Research*, 15 (1), 35-69.

[9] Reuters-21578 Test Collection. http://www.daviddlewis.com/resources/testcollections/reuters21578/.

[10] Salton, G. *Automatic Information Organization and Retrieval*. McGraw Hill Text, 1968.

[11] Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R. (2004) The Perfect Search Engine is Not enough. In *Proc. CHI 2004*, 415-422.

Exploring the interaction effects between search tasks and interfaces

Muh-Chyun Tang
Department of Library and Information Science
National Taiwan University
mctang@ntu.edu.tw

Kuo-Ting Huang
Graduate Institute of Networking and Multimedia,
National Taiwan University
b90010@csie.ntu.edu.tw

ABSTRACT
In this paper we propose a naturalistic methodology tailored to test two term suggesting tools designed to support exploratory search with PubMed. Terms co-occur with user's query terms in the initial returned set are extracted and present to the user in two ways: one makes use of the MeSH tress structure; the other rank terms by their mutual information with the query. The user study adopts a quasi-experimental design, comparing the performance of the two interfaces with the baseline PubMed interface on subject searches. Of particular interest is the usefulness of the proposed interfaces for difference types of search tasks. One novel aspect of our methodology is the use of real user and real search tasks in naturalistic work setting, instead of using assigned tasks in a controlled experimental setting. The performance criteria will be users' satisfaction with the search results and search experiences.

Author Keywords
Information Retrieval Evaluation, User study, exploratory search interface

1. INTRODUCTION
Researchers in information seeking behaviors have long recognized search situations where information needs are not static, but are emergent phenomena out of users' interaction with the information objects [1,5]. Recently there have been growing interests in interfaces designed to support exploratory search and methodologies to evaluate their effectiveness [6].

Exploratory search interfaces pose a new set of challenges to system evaluation. Thus far proper metrics that reflect desired results are still lacking. Furthermore, exploratory search interfaces inevitably entail users' active participation therefore the need to counterbalance the variance introduced by users in experiment design. Simulating exploratory search tasks in an experimental setting presents another great challenge as it is difficult to design tasks that are semantically open to allow evolving. And users' exploring and learning with a novel interface also pose as a threat to study validity.

In traditional IR evaluation, every effort is made to make sure that everything is equal other than the system components being compared. User, tasks and their interaction effects are regularly treated as random variance the experimenters strive to systematically control and minimize.

Such a controlled-experiment approach is very effective if the systems or system features compared are conceptualized as tools of general purposes. In traditional IR evaluation, the assigned tasks are created mostly in an ad hoc manner, without theorizing task characteristics and how these characteristics might interact with the system features concerned. Yet with more and more interactive search features that demands users' feedback being introduced (e.g. visualization and clustering, relevance feedback etc.), it is easy to imagine search tools specialized in certain aspects of search process or certain types of search tasks. Their strengths might neither easily be translated into traditional evaluation metrics, nor be readily observable across all types of tasks.

On the other hands, as the search tools become more specialized, there is also growing awareness of the diversity of search tasks users are engaged in when accessing information (e.g. [7]). It follows that the first step to evaluate a novel search tool such as an exploratory search interface is to identify the types of search tasks it best supports. In other words, in the design of evaluation study, the interaction effects between search tasks and interfaces becomes a topic of great interest instead of random variance one seeks to minimize.

In this study we propose a quasi-experimental approach to the evaluation of two interfaces designed with exploratory searches in mind. It is hoped that the study will allow us to investigate the usefulness of the interfaces in different kind of search tasks.

2. SYSTEM EVALUATED

2.1. TERM SUGGESTION
One of the critical problems faced by the user searching for a heterogeneous and massive bibliographic database such as

PubMed is the difficulty of managing the returned results. Several system features of PubMed (e.g. default “explode” function and free-text indexing in title and abstract fields) that aim at facilitating end-user search tend to increase indexing exhaustivity and therefore favor search recall at the expense of precision. Faced with the unmanageable amount of returned results, users are often left with few options but to hastily browse the first few returned pages.

The consequence of information overload creates at least two barriers to a successful user-system communication. Firstly, there in no telling whether there might be documents relevant to users’ needs buried deep down in the returned set that never get the chance of being viewed. Secondly, the skimming of the surface of the returned set also gives incomplete feedback for the judgment of users’ query performance.

The breakdown of user-system communication is especially severe in search situations where users search for unfamiliar topics or only have vague information needs. Without timely cues from the system, users are unlikely to be able to interact effectively with the system and refine their searches.

To elicit more input from users, attempts have been made to dynamically extract metadata from the initial returned set for term suggestion purpose [2,3,4]. We adopt a similar approach that extracts MeSH terms from the returned set to help users refine their queries when searching PubMed.

2.2. PRESENTING AND SUGGESTING TERMS

It is crucial to organize and present the extracted MeSH terms in a manner that makes efficient browsing and selection possible. We plan to experiment and compare two ways of presenting the MeSH terms. The mutual information version (MI) will rank the extracted terms based on their mutual information with the original query. The other presentation method, the MeSH tree version (TREE) will organize the extracted terms using the 16 top-level semantic categories of MeSH. Terms co-occur with the initial query will be mapped against the MeSH tree with the non-occurring terms left out. Thus a filtered MeSH tree that includes only the extracted terms will be generated dynamically each time a query is submitted. The user will be given the option of browsing and selecting terms from the post-querying MeSH tree.

3. RESEARCH DESIGN

The study will be conducted in a naturalistic, longitudinal setting:

- A total of 18 regular PubMed users will be recruited
- Participants search at their workplaces for their own search topics
- Latin square design to reduce individual variance and order effects
- Participants instructed to conduct only “subject search”, as opposed to “known-item” search
- Each participant will conduct 9 search sessions in a period of 3 months using alternately the three interfaces: the baseline PubMed, the MI version and the TREE version

3.1. Real users and real search tasks

Instead of assigning the participants a uniform set of search tasks, we decide to let the participants conduct their own search tasks. Even though it has been demonstrated that well crafted, semantically open task descriptions were able to simulate enough participant involvements and yielded results comparable with those of the real tasks [8], the use of textual representation is problematic for our purpose of evaluating an interface designed to facilitate query construction. The textual representation of the task would likely interfere, if not offset, the query construction functionality the interfaces designed to provide. It was feared that the participants would rely heavily on terms in the task narrative, which not only makes the interfaces unneeded but also puts the validity of the task categorization in jeopardy.

3.2. A naturalistic and longitudinal study

One of the major methodological issues we need to tackle is the learning or familiarity effect when comparing customary features with a relatively novel one. The user might prefer or perform better with the tool s/he is most familiar with simply out of habit. We hope that a longitudinal study would be a sufficient solution to users’ learning effect because it enables us to observe the learning and assimilation of interface processes by the users, which is often unavailable in a strictly controlled environment. An extended period also gives us an opportunity of sampling different search tasks with relatively fewer participants so the relationships between and search tasks and usefulness of interfaces can be observed.

4. RESEARCH PROCEDURES

As the experiment begins, the participants will log in and conduct the searches at their workplaces instead of coming to the laboratory. The participant starts with each search session with filling out a pre-search questionnaire, followed by performing searches on the search task of their own, and concludes with answering a post-search questionnaire. All the questionnaires will be administered online to make remote monitoring possible. The participants’ interactions with the interfaces will be logged. In pre-search questionnaire, participants will be ask to characterize their search tasks on attributes such as topic familiarity, comprehensiveness of the results desired, and specificness of the search question. In post-search questionnaire, they will be asked to indicate their levels of satisfaction with the search results and search experience, which will be the main criteria variables.

5. DISCUSSION

The strength of this approach lies in the authenticity of the search problems and the naturalistic setting where the searches are performed. The measurement of users' search tasks allows us to explore their interaction with type of interfaces. Yet it is also recognized the lack of objective evidence about the effectiveness of the interface as measured by traditional evaluation metrics. This seems to be an inevitable consequence of sampling real information problems instead of assigning uniform tasks, as it precludes the use of comparable measures. One possible way of maintaining the balance between realism and comparable measurement in the future is utilizing the real life situation where multiple participants are assigned a uniform set of information problems such as classroom assignments or assignments for professional information analysts. Such a "group" or "team" real life task environment will afford us to use both subjective and objective measures of performance.

REFERENCES

1.Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science, 5* (1980),133-143.

2.Hearst, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ping Yee. Finding the flow in Web site search. *Communications of the ACM 45*,9 (2002), 42-49.

3.Lin, X. Visual MeSH. Proceedings of 22nd International Conferences on Research and Development in Information Retrieval (SIGIR'99, Berkeley, CA. pp 317-318).

4.Pollitt, A. S. The key role of classification and indexing in view-based searching. *International cataloguing and bibliographic control* 27,2 (1998), 37-40.

5.Taylor, R. S. Question-negotiation and information seeking in libraries. *College & Research Libraries, 29*,3, (1968) 178-194.

6.White, R. W., Drucker, S., Kules, B. and schraefel, m. c. (2006). Supporting exploratory search. *Communications of the ACM (Special Section)*, 49 (4): 36--39.

7.Vakkari, Pertti. Task-based information searching. *Annual Review of Information Science and Technology,*. 37 (B. Cronin, Ed.) Information Today: Medford, NJ, (2003), 413-464.

8.Borlund, P. Experimental components for the evaluation for interactive information retrieval systems. *Journal of Documentation*, 56, 1, (2000), 71-90.

Exploring and Investigating: Supporting High-level Search Activities

Gina Venolia, Meredith Ringel Morris, and Dan Morris
Microsoft Research
One Microsoft Way
{ gina.venolia | merrie | dan }@microsoft.com

ABSTRACT
Much work has been done to improve search as an isolated act, yet little has been done to understand search as it relates to higher-level patterns of behavior, or to develop user interfaces to support these patterns. In this paper, we analyze exploratory and investigative search processes that involve performing several related searches over a period of time. We then discuss requirements for interfaces to support these tasks, and describe three prototype systems.

Author Keywords
Exploratory search, sensemaking, persistent search.

ACM Classification Keywords
H.5.2 Information Interfaces and Representation: User Interfaces.

INTRODUCTION
Broder presents a taxonomy of web search tasks [2], characterizing them as *navigational*, *informational*, or *transactional*. For example a user may want to find the web page for the Toyota Prius car (navigational), find the gas mileage rating of that car (informational), or use a web-based tool to configure the options on the car and estimate costs (transactional). This model presents each query as an isolated task.

Various researchers have identified and described larger-scale patterns of search-related behavior. We have identified two such patterns, which we call *exploration* and *investigation*. In exploration, the user's goal is to increase his or her knowledge about a topic. In investigation, the goal is to reach a decision. In this paper we briefly survey related work, describe these patterns, and then briefly describe three UI prototypes which support them.

RELATED WORK
Several researchers have identified large-scale information-seeking patterns that involve search. Vélez et al. [4] present a model of *query refinement* in which a search task consists of a sequence of related queries, where the user incrementally modifies the query specification to achieve a desirable result set.

CHI 2007 Workshop on Exploratory Search and HCI.

A longer version of this paper is available as Microsoft Research Technical Report MSR-TR-2007-05.

Bharat [1] describes a pattern of information-seeking behavior that encompasses many searches over time. He describes a tool, *SearchPad*, which attempts to support this pattern of behavior. The tool records the sequence of queries and allows the user to flag important results as *leads*. The SearchPad UI shows the recent queries and, for each, its leads. The tool makes it easy to issue the same query to multiple search engines. Bharat describes how the lightweight and transient nature of this information makes it better suited to support complex searches than bookmarks.

Russell et al. define *sensemaking* [3] as "the process of searching for a representation and encoding data in that representation to answer task-specific questions." They describe a system of three related *learning loops*, wherein representations are created, instantiated, and modified. Sensemaking activities may incorporate multiple actors and take place over long periods of time.

CASE STUDY: ROOT-CAUSE ANALYSIS OF SOFTWARE DEFECTS
One task that relies heavily on in-depth exploration is *root-cause analysis*, or *RCA*, which is the process of finding the reasons for critical failures. We have interviewed one of the people at a large software corporation responsible for RCA of key software defects.

The process begins when the a particular incident is identified as worthy of investigation, e.g. a build break or a critical bug regression introduced by a patch. He chooses some keywords from the incident description and searches for them over the various repositories that he has at his disposal, including the bug database, the product-support knowledgebase, email repositories, project-specific databases, code check-ins, etc. Each repository has its own unique search interface, so this process is quite tedious. He refines his query to improve the result set. He then scans the result list to find potentially-relevant documents, which he opens and examines.

For each relevant document he creates an entry in a structured Microsoft Word document, clearly identifying its date, the document type and name, often a quotation from the document, and often his own notes about the document and/or questions for further investigation. These entries are sorted chronologically, forming a timeline. The Word document is his working "notebook," and is generally not shown to others.

In reading the source documents he may encounter additional terms, people, date ranges, etc. which lead to more searches, which in turn lead to more documents, which then lead to more entries in the timeline. In addition, he is creating and discarding hypotheses, which he retains in another section of his "notebook" document.

In addition to his electronic investigations, he interviews key participants in the incident. These interviews may result in the addition of entries to the timeline and clarifying comments for existing entries.

His final report is a separate document, in which he presents a narrative of the incident and a summary of its root causes. The former is based on the timeline and the latter on the hypotheses.

One of his investigations was of a particular build break. The timeline for this investigation contained about 50 entries representing documents or events that spanned 48 days. The final report identified 21 issues and three root causes. He reported that the investigation, including the electronic sleuthing, interviews, and writing, took him "a couple of weeks" of full-time work.

EXPLORATIONS AND INVESTIGATIONS

The pattern of behavior described for RCA is familiar to anyone who has made a purchasing decision or performed a literature review using online resources. Based on our study of RCA, as well as reflection on our own exploratory and investigative search habits, we propose a preliminary model of this pattern.

In the *exploration* pattern of behavior, the user's goal is to increase his or her knowledge about a topic. For example a user might explore the models of environmentally-friendly cars that are available and investigate their environmental benefits compared to standard cars.

An exploration involves finding documents that may be relevant to the topic, examining them, and then discarding irrelevant documents and retaining relevant ones. We call this process *triage*. Documents may be found by executing queries, browsing, or by a combination of the two. While queries may be formed by iterative query refinement, an exploration is often comprised of multiple independent queries. When examining documents a user may find keywords that are then used in further queries.

For simple explorations the relevant documents and keywords may be retained mentally. Complex explorations require external representation, in which the user may record and annotate the relevant documents. A user may invest significant time into an exploration, which may be divided over multiple sessions, so persistence of the accumulated state is important. A user may revisit key searches in an exploration to see "what's new."

The two most important aspects of explorations are that they *persist* and *consist of multiple independent queries*. The multiple independent queries that characterize explorations are distinct from the multiple *related* queries that characterize query refinement.

In an *investigation* the user's goal is to reach a decision. The investigation pattern extends the exploration pattern with a collection of hypotheses, which may be created, supported, and refuted in the course of the exploration. For example a user might perform an investigation to decide which hybrid car models to test-drive. The hypotheses in this case might be "the Prius is the right car for me," "the Civic Hybrid is the right car for me," etc. Each hypothesis is associated with the search results that *support* and *refute* it. An investigation may culminate in the generation of a report describing, at a minimum, the well-supported hypotheses.

This model incorporates concepts identified in the Related Work section. The search/refine loop is equivalent to query

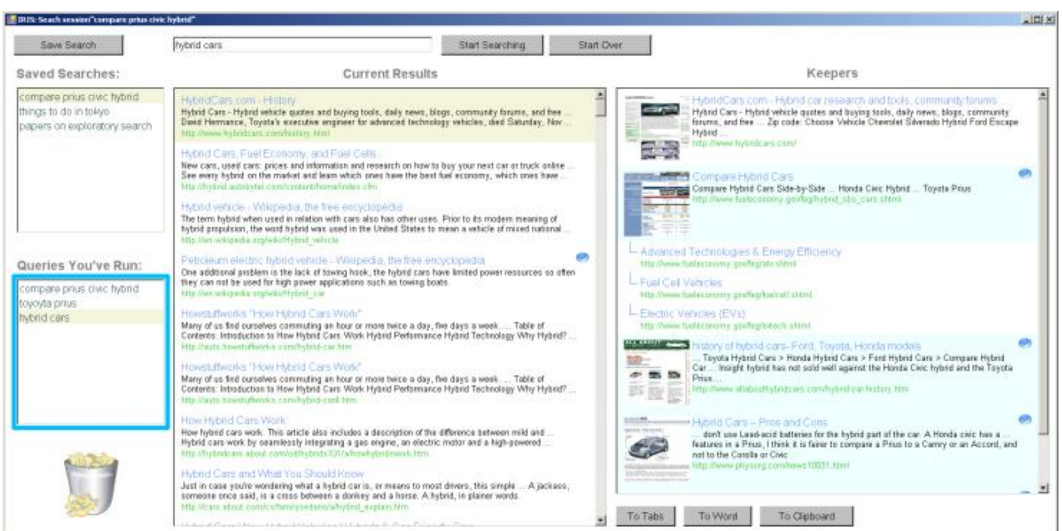

Figure 1: The Interactive, Rapid-Iteration Search (IRIS) prototype.

refinement [4]. Items retained as supporting evidence are similar to the "leads" that SearchPad provides [1]. Sensemaking [3] revolves around representations. In this model the core representation is a collection of documents marked as relevant. Additional aspects of representation come from annotation and hypotheses. Sensemaking's "residue" becomes additional keywords. The model may be seen as a specific framework for a subset of sensemaking activities.

Not all searches are part of an exploration or investigation. We mean to extend, not replace, Broder's taxonomy [2].

UI TO SUPPORT EXPLORATIONS

We have prototyped three UIs that implement some or all of these features. Each UI emphasizes different aspects of the design space.

The Interactive, Rapid-Iteration Search (IRIS) prototype (Figure 1) emphasizes the following features that we propose will enhance the exploratory search experience: *Re-ranking of results to reflect their occurrence among multiple queries*. Results that have been recalled numerous times but ignored can be automatically hidden to remove redundancy, and results that have been recalled by multiple queries but were not highly ranked in any one query are automatically promoted to higher rank. *Integration with navigation*. A user can view results in a separate browser without disrupting his search context, and pages that the user subsequently navigates to are automatically associated with the search results that originated the navigation. *Automatic summary preparation.* The user can prepare and distribute a summary of his search results, with information about the queries that produced each result, with a single click.

The S^3 (Storable, Shareable Search) prototype (Figure 2) emphasizes a document model for searches. A session (i.e., a series of queries related to a single task) can be saved to and loaded from file. This file (which includes data such as query terms used, sites visited, and comments) can even be shared with others, enabling asynchronous collaboration on an exploratory search task. Each site and comment is tagged with the username of the person who contributed it to the persistent search session.

The Search Grid prototype (Figure 3) also uses a document model. Its user interface is dominated by a grid where the rows are the search results, columns are the queries, and cells show a checkmark for all queries that produced a given result. The list of relevant and untriaged items is the most salient feature; queries are demoted to metadata on the items. Another interesting aspect of the Search Grid prototype is that each item presents commands (in the

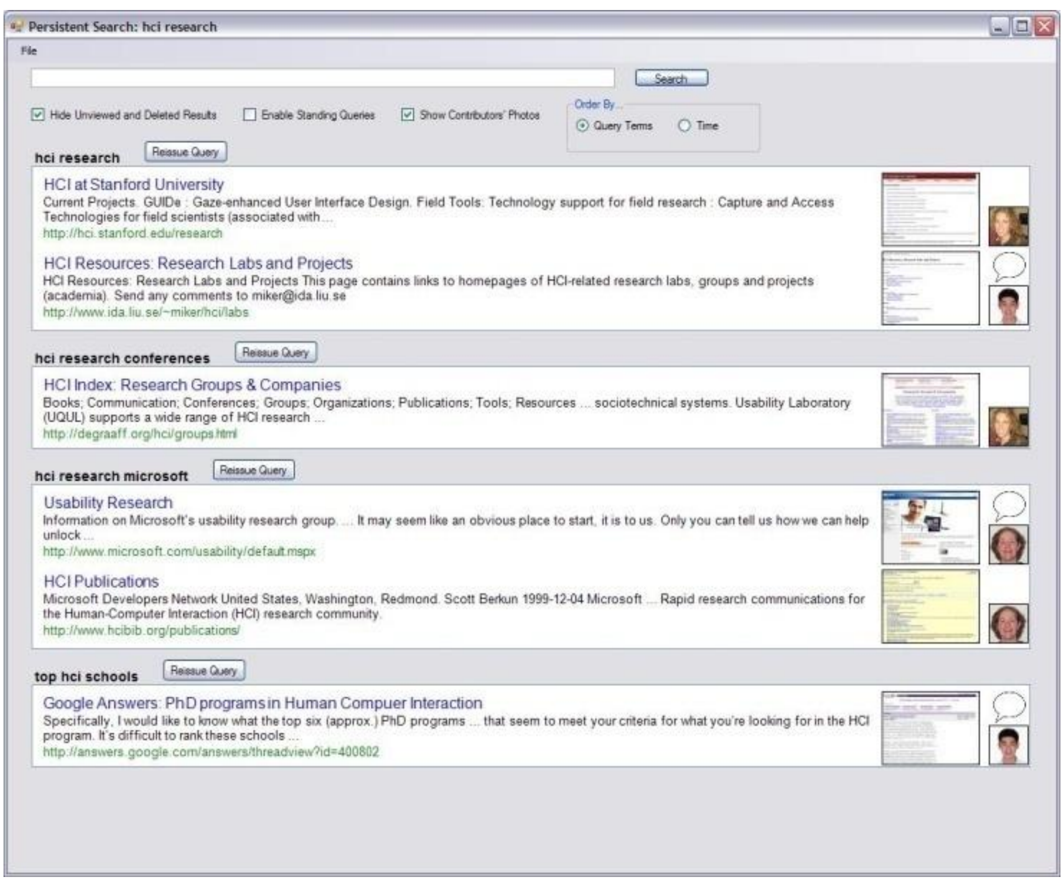

Figure 2: The S^3 prototype.

context menu) to create new queries related to the item, e.g. searching for pages that link to the item's URL or contain text similar to the item's. To some degree these item-relative queries, which we call *pivots*, bridge the gap between searching and browsing.

EXTENDING THE UI TO SUPPORT INVESTIGATIONS

Any of these prototypes could be extended to support investigations by adding these key UI features:

- Representation of hypotheses
- Association of each hypothesis with supporting and conflicting evidence
- Generation of a skeleton report by emitting a document listing the hypotheses and, for each, the related evidence

For example, in the Search Grid prototype each hypothesis would be represented as a column, containing a three-way toggle so the user could specify whether the item supported or refuted the hypothesis, or was unrelated to it.

NEXT STEPS

We are at a preliminary stage in our development of these concepts. There is much to do to validate our model of exploration and investigation and our UI solutions to support those activities. We plan to improve our prototypes and then perform lab studies to see how users approach explorations and investigations with existing tools and our new tools, to detect usability problems in our tools, and to compare performance and user experience to that achieved with traditional search tools. We expect the UI approaches to evolve through this process.

REFERENCES

1. Bharat, K. SearchPad: Explicit capture of search context to support Web search. In Proc. WWW 2000, 493-501.
2. Broder, A. A taxonomy of web search. In SIGIR Forum 36, 2 (Sep. 2002), 3-10.
3. Russell, D., Stefik, M., Pirolli, P., and Card, S. The cost structure of sensemaking. In Proc. CHI 1993, 269-276.
4. Vélez, B., Weiss, R., Sheldon, M. A., and Gifford, D. K. Fast and effective query refinement. In Proc. SIGIR 1997, 6-15.

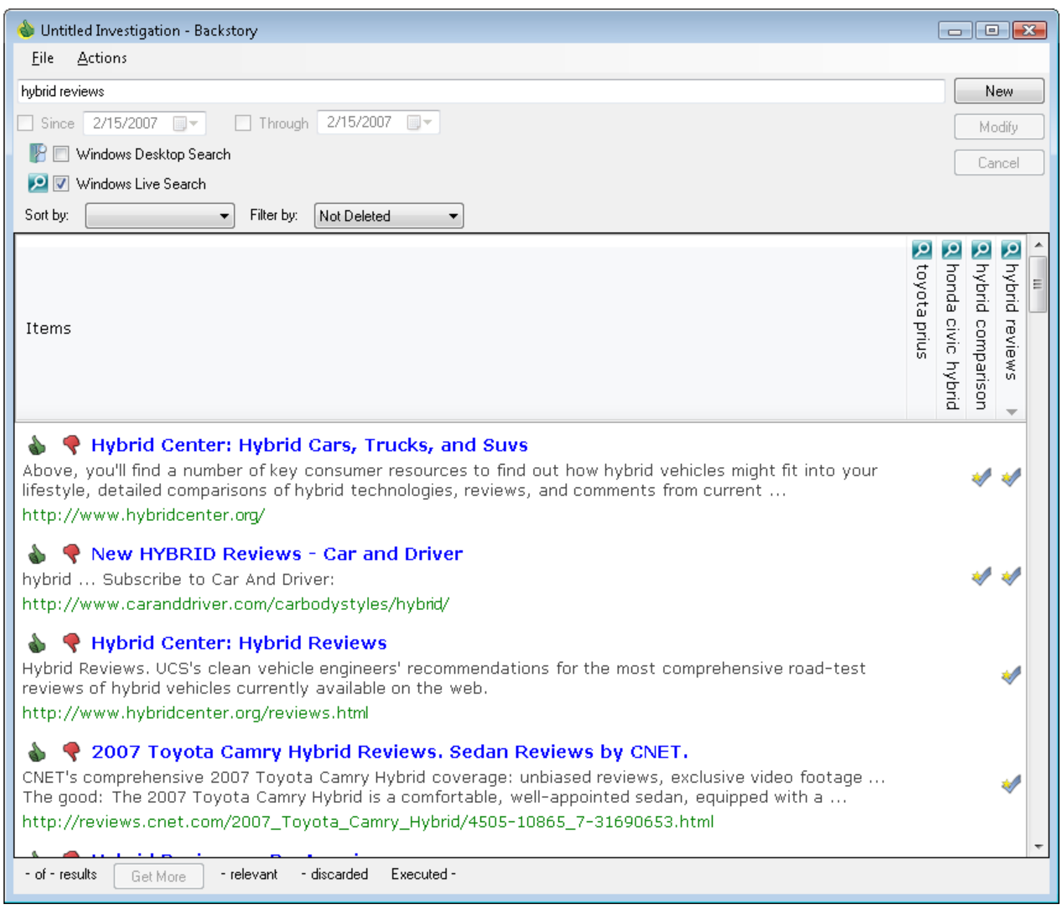

Figure 3: The Search Grid prototype.

Using Trees to Explore the Forest: An Interactive Digital Library for Natural Resource Managers

Mathew J. Weaver
EarthSoft, Inc.
mweaver@earthsoft.com

Marianne Lykke Nielsen
Royal School of Library
and Information Science (Denmark)
mln@db.dk

Lois Delcambre
Portland State University
lmd@cs.pdx.edu

Timothy Tolle
Natural Resources Consultant
timtolle@aol.com

ABSTRACT

Natural resource managers make decisions that impact numerous organizations, individuals, and the environment. Such decisions require a broad range of information gleaned from a variety of documents. We built a domain-specific digital library, called Metadata++, with a focus on specialized terminology including terms from a large number of well-established, well-known classification schemes and terminologies used by multidisciplinary experts. These experts frequently use the same terms with often subtle (and occasionally significant) differences in meaning. This conglomeration of terminology presents an interesting challenge and opportunity with regards to digital library design and user-driven searching. The Metadata++ digital library system allows the user to interactively browse and search the terminology as they index documents as well as browse and search for documents. In this paper, we focus on the Metadata++ user interface that supports user-driven exploration of the digital library. We also report on a user study that demonstrates how natural resources managers easily understood the path-based representation of terms for both indexing and searching.

AUTHOR KEYWORDS

Digital Library, user interface, interactive search

ACM CLASSIFICATION KEYWORDS

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6): Graphical user interfaces, interaction styles, user-centered design

INTRODUCTION

Suppose you are asked to decide whether or not to build a new campground next to a backcountry lake. Or perhaps you must evaluate how a proposed timber harvest would affect the wildlife within a particular watershed. These examples illustrate just a few of the many issues faced by natural resource managers as part of their daily responsibilities. Natural resource management is fundamentally interdisciplinary with almost every project involving various disciplines – such as soil, forestry, vegetation, hydrology, climatology, hydrology, wildlife and fish biology, recreation, and range land. Natural resource managers must be able to quickly and efficiently find information pertinent to pending decisions, often without a detailed knowledge of the specialized terminology used in the individual documents.

Our research[1] involves the design, construction, and evaluation of a digital library system [9] called Metadata++. One of our primary research objectives addresses a problem common to many multidisciplinary digital libraries: utilizing multiple controlled vocabularies that come from a variety of disciplines and often contain the same (or similar) terms. In Metadata++, terms (words or phrases used for indexing and searching) are represented and distinguished by their path in the hierarchy of vocabularies. The user (who is often unfamiliar with many of the vocabularies) can explore the hierarchy of terms and infer the meaning or connotation intended for a particular term based on its path in the hierarchy. We used Metadata++ in an extensive case study involving natural resource managers as well as exploratory research with medical professionals. Section 2 describes the Metadata++ user interface that supports user-driven exploration of the

[1] This work is supported in part by the National Science Foundation, grant number EIA 9983518. Any opinions, findings, conclusions, or recommendations expressed here are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

digital library. Section 3 outlines a user study conducted with the Metadata++ system that clearly indicates that users felt very comfortable with the path-based representation of terms. The final section presents conclusions and future work.

SEARCHING THE METADATA++ DIGITAL LIBRARY

The main goal of any library, including a digital library, is to warehouse and organize information and make it more accessible to a wide spectrum of users. Accessibility of information is based on how easily a user can search for, and find, information relevant to his or her needs. Nielsen [7] reports that users want and need a thesaurus – primarily for query formulation. Bates [1] suggests two different thesauri – one thesaurus for indexers and a different thesaurus for searchers. Metadata++ uses the same hierarchy for both tasks, providing consistency between tasks. The Metadata++ system provides a number of features that allow users to interactively explore the hierarchy of controlled vocabularies and associated documents:

- The user can browse the terminology directly (using the familiar "+/-" convention) in the *Browse* window.
- The user can right-click on any term to see multiple occurrences (the same term occurring in multiple vocabularies, each shown with their full path), synonyms, related terms, human-indexed documents, and computer-indexed documents.
- The user can search for terms at every level in the hierarchy, including the use of wildcard matches, using the *Find Term* function. Each term found is displayed with its complete path.
- The user can double-click on any term in the *Find Term* window in order to see it in context, within the full hierarchy of terms in the *Browse* window.
- The user can select path-based terms, to index a document or to formulate a search, in either the *Find Term* or the *Browse* window and drag it to the document or to the *Search* window.

Bates [1] makes some interesting observations in regards to searching and the nature of information systems. She states that it is impossible to accurately predict what specific aspects of a topic a search will pursue and which specific terms the searcher will use. She continues with this suggestion:

> "[We should] stop trying to design systems that will *target* the desired information through perfect pinpoint match on the one best term; rather, design systems to *encompass* the answer by displaying and making it easy to explore a variety of descriptive terms. Show searchers a wide range of descriptive terms and thereby implicitly educate them on the need to produce variety." ([1], pg. 361)

Instead of trying to guess what the user intended, Metadata++ makes it easy and intuitive for the user to interactively expand the search with a variety of terms. Instead of typing in free-text search terms, searchers must select path-based terms from the hierarchy; but the user may type any search string into the *Find* window to find possible path-based terms that may then be added to the search. Searchers may drag any term from the *Browse* window or the *Find* window and drop it on the *Search* window. Dropping the term on the *Search* window will automatically show the results for that term, which include both human-indexed documents (in green) and computer-indexed documents (in blue). Figure 1 shows the *Search* window after dropping the *AIR\air quality* term into the search. Research [5] [7] shows that selecting search terms from a controlled vocabulary is useful for expressing searches and improves search consistency.

Additional research shows [1] [3] that thesaurus-based search expansion is effective at improving search results. Various automatic thesaurus-based expansion algorithms exist, but further research [3] [6] suggests that interactive user-driven query expansion can be more effective and preferable when compared to automated query expansion. Metadata++ supports the user in interactively expanding the search. After right-clicking on any search term in the *Search* window, the searcher will see a popup context menu (illustrated in Figure 2). The first item in this menu will display a popup tree containing the multiple occurrences, synonyms, and related terms for the selected term. This

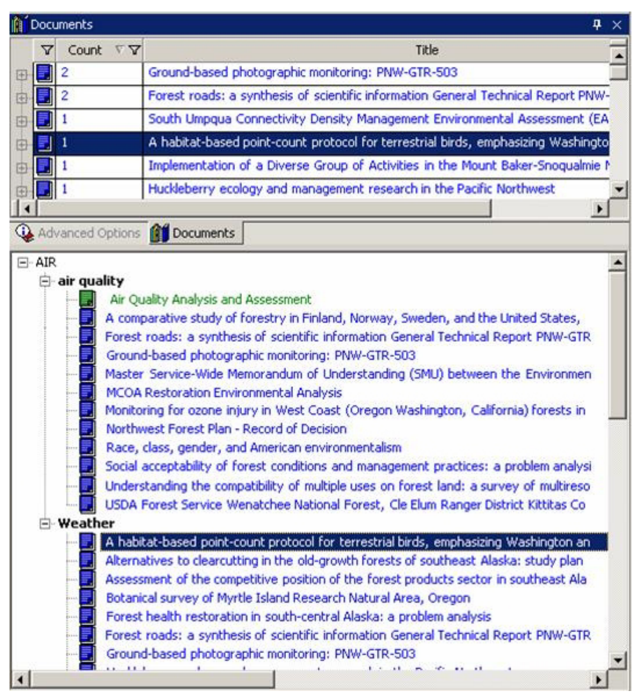

Figure 1: Screenshot of Search result document

popup tree provides suggested terms that may be of interest to the searcher based on their relationship to the current search term. If desired, the searcher may click on any term in popup tree and that term will automatically be added to the search. The second item in the context menu provides a convenient way for adding all multiple occurrences of a term. The third menu item works in a similar fashion; it provides a quick and easy way to expand a search to include all narrower terms of the current search term. In addition to adding search terms, the user may easily remove an unwanted search term (and any descendant search terms) by simply deleting it from the *Search* window (click on the term in the hierarchy and press the <DELETE> key).

Metadata++ uses the hierarchy of path-based terms to display the actual search result (as show in Figure 1). Petrelli et al. [8] determined that users expect to see a ranked list of documents and anything else is considered annoying. However other research [1] [4] indicates that showing search results in context of the subject hierarchy increases user satisfaction. By showing the search results in context of the hierarchy, the searcher can infer the relevance of any particular document by seeing where that document appears in the search results. For example, suppose the searcher added all of the descendant terms of a particular search term. Documents related to the immediate child terms may be considered more relevant than documents related to other descendant terms that are several generations below the original search term. A document related to more than one search term (and thus appearing in more than one place in the search result) might have a greater degree of relevance. Viewing the documents in context of the hierarchy gives a clearer picture of why and how any particular document relates to the search – as opposed to a ranked list where the user does not necessarily understand why or how documents are ranked.

In addition to the hierarchical search result, the user may also choose to view a simple list of documents. This list (as illustrated in the top of Figure 1) shows the same set of documents that appear in the hierarchical view sorted by count (the number of times the document appears in the hierarchical view). The list of documents can easily be sorted by count or title simply by clicking on the corresponding column header. If the searcher clicks on any document in the hierarchical view, that same document is automatically selected in the list view (as shown in Figure 1) and vice versa.

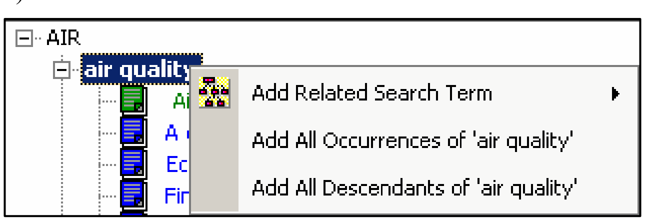

Figure 2: Cropped screenshot showing the context menu for Search term

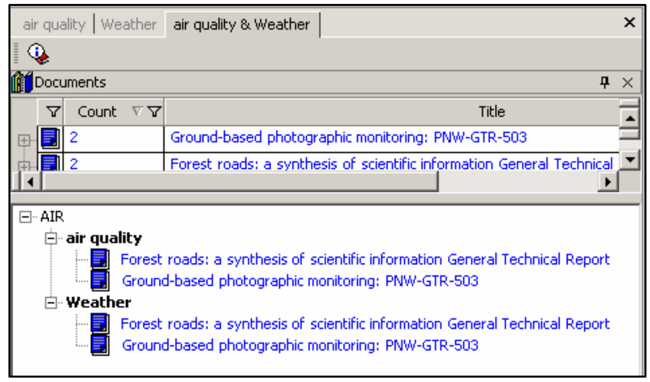

Figure 3: Cropped screenshot of a compound search

Compound Search

A search task within natural resource management typically involves more than one concept [2]. For example, a natural resource manager may be trying to ascertain the overall water quality within a particular region. This task would comprise two different search concepts: water quality and the region of interest. Users may perform compound searches in Metadata++ by interactively expanding each individual search concept (as described in the previous section) then combining the results of each concept into a compound search.

Figure 3 illustrates the results of a simple compound search involving *AIR\air quality* and *AIR\Weather*. We produced this screenshot by first creating a Search window and adding only the *AIR\air quality* term. Next, we created a second Search window and added the term *AIR\Weather*. We could have further refined each search concept by adding (or deleting) additional search terms. Finally, we intersected the two Search windows, which produced the visible tab shown in Figure 3, labeled "air quality & weather".

When doing an intersection, Metadata++ combines the hierarchical structure from both existing *Search* windows – then removes all documents that do not appear in both concepts. As shown in Figure 3, only those documents that were in both original *Search* windows (the first two tabs) appear in the compound *Search* window (the third tab). The current Metadata++ application only supports intersection – but the Metadata++ conceptual model supports two additional compound operators: *Union* (include all path-based terms and documents from both searches) and *Difference* (include all path-based terms from the first search, but remove all documents that appear in the second search).

METADATA++ USER STUDY

We completed a user study designed to evaluate the usefulness of the Metadata++ digital library system. The study objectives focused on evaluating the users' perception, understanding, and satisfaction with the Metadata++ model and application. We conducted the study on-site at two different USDA Forest Service offices – with four users participating from each office. Each of the four searchers was given the same two search tasks – actual search tasks based on real information needs. They received a written explanation of the task and were asked to review the task and outline the concepts they felt were important for that particular search task. They then used the Metadata++ system to select the terms that best represented the concepts they had outlined for the search task. We gathered data using written notes (from three observers), audio recording/transcription, application logging, and follow-up interviews.

Participants knowingly selected the paths that were most appropriate for the particular task. In some cases, they selected a path, then later removed the path and selected a different occurrence of the same term – one that more accurately described the concept that they wanted to express. In several cases the multiple occurrences of terms inspired the searchers to try out a variety of search terms for each searching task [1]. Participants expressed how multiple occurrences illustrate viewpoints they had not previously considered – and provided ways to restrict or expand the search. Participants were comfortable with the interactive compound search. In addition, they requested additional features including the ability to "black list" certain documents during a search (prevent specified documents from showing up in search results regardless of search expansion) and having the application highlight new documents that show up with each interactive step during search expansion.

CONCLUSIONS & FUTURE WORK

We have built a digital library system, Metadata++, which uses hierarchical controlled vocabularies to support user-driven exploratory searching. Our user study shows that natural resource managers comprehend and appreciate the features of the Metadata++ application. Future work includes implementing a mechanism for the intuitive display and expression of compound searches (specifically intersection and difference).

References

[1] Marcia J. Bates. Subject Access in Online Catalogs: A Design Model. *Journal of the American Society for Information Science*. 37 (6):357-376, 1991.

[2] Balbinder Banga, Eric Landis, Timothy Tolle, Lois Delcambre, Fred Phillips. User needs assessment for the adaptive management portal. In *National Conference on Digital Government Research*. Los Angeles, California, Pages 329-334, 2002.

[3] Hsinchun Chen, Tobun D. Ng, Joanne Martinez, Bruce R. Shatz. A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System. *Journal of the American Society for Information Science*. 48 (1):17-31, 1997.

[4] Susan Dumais, Edward Cutrell. Optimizing Search by Showing Results in Context. In *ACM SIGCHI Conference on Human Factors in Computing Systems*. Seattle, Washington, Pages 277-284, 2001.

[5] Mirja Iivonen. Consistency in the Selection of Search Concepts and Search Terms. *Information Processing & Management*. 31 (2):173-190, 1995.

[6] Bernard J. Jansen. Seeking and implementing automated assistance during the search process. *Information Processing & Management*. 41 (4):909-928, 2005.

[7] Marianne Lykke Nielsen. Task-based evaluation of associative thesaurus in real-life environment. In *Proceedings of the ASIST Annual Meeting*. Providence, Rhode Island, Pages 13-18, 2004.

[8] Daniela Petrelli, Preben Hansen, Micheline Beaulieu, Mark Sanderson, George Demetriou, Patrick Herring. Observing Users – Designing Clarity: A Case Study on the User-Centred Design of a Cross-Language Information Retrieval System. *Journal of the American Society for Information Science and Technology*. 55 (10):923-934, 2004.

[9] Weaver, Mathew Jon. Enhancing a Domain-Specific Digital Library with Metadata Based on Hierarchical Controlled Vocabularies. Doctoral Dissertation, OGI School of Science & Technology, Oregon Health and Science University, 2005.

Subjectivity: Its Role in Exploratory Search Processes and Evaluation

Barbara M. Wildemuth
School of Information & Library Science, University of North Carolina at Chapel Hill
100 Manning Hall, CB #3360, Chapel Hill, NC 27599
wildem@ils.unc.edu

ABSTRACT
Subjectivity is inherent in exploratory search. It is evident in the attributes of the searcher before beginning the search (e.g., the searcher's domain knowledge). In the form of the searcher's interests, knowledge, cognitive/learning style, and personality, it affects the search process – its overall trajectory. And it comes into play at the end of the search, when the searcher makes judgments about the success of the search. Researchers and designers interested in exploratory search will need to develop measures of searchers' beliefs about a search's success and attitudes toward the search outcomes. These measures can then be leveraged to more directly evaluate the quality of the search process and to design personalized systems that improve a searcher's performance.

INTRODUCTION

As White et al. [32] note, "defining what constitutes an exploratory search is challenging" [p.38]. Nevertheless, they provide an initial definition, in concert with that provided by Marchionini [21], that, "in exploratory search, users generally combine querying and browsing strategies to foster learning and investigation" [32, p.38]. In the "real world" (i.e., outside the boundaries of TREC-like evaluation paradigms), each exploratory search is conducted by an individual person or a small group of people. In this sense, the process of conducting an exploratory search is inherently subjective.

This inherent subjectivity can be addressed in many different ways. In some research traditions, the subjective experience is highlighted as the preferred focus of a particular social phenomenon. In these traditions, the researcher attempts to understand the phenomenon of interest from the participants' perspectives, seeing it through their eyes. This approach rests on the assumption that reality is, in fact, socially constructed and cannot be understood apart from the understandings of those participating in its construction [5]. However, for others (includeing most system designers), subjectivity introduces "fuzziness" or complexity into our understanding of a particular reality [27, p.5], making it difficult to generalize study findings beyond the study participants.

The subjectivity inherent within a particular search trajectory has implications both for understanding the search process and for evaluating the effectiveness of a particular IR system in supporting the search process. Most importantly, the inherent subjectivity of the search process may work against both the standardized evaluation of a system's effectiveness and, possibly, even the design of systems that are useful to a large number of people. This paper will explore the implications of subjectivity at three points in an exploratory search: the starting point for the search, the trajectory of the search being undertaken, and the conclusion of the search. In addition, implications for evaluation methods and for system design will be discussed.

THE STARTING POINT FOR THE SEARCH

A variety of individual characteristics may affect a search at its beginning, but perhaps the most prominent of these is the domain knowledge of the searcher. The focus of exploratory search on learning and investigation makes it very likely that the searcher's domain knowledge will influence retrieval success, as well as the search process.

Somewhat surprisingly, domain knowledge has not been consistently found to affect retrieval success. Studies finding a positive effect of domain knowledge on retrieval success include those by Jacobson and Fusani [19] and Marchionini [22]. However, a number of studies have found no such relationship [1, 4, 13, 26, 33]. Thus, whether domain knowledge affects the recall and precision of a retrieved set is still an open question.

A few studies have looked directly at the effect of domain knowledge on the process of searching, identifying the effects of domain knowledge on the tactics used in search strategy formulation and reformulation [34], the amount of time spent monitoring a search and the frequency with which terms were combined [18], the focus/attention of the searcher and his or her expectations for the outcomes of a search [23], and the searcher's ability to choose appropriate search terms [2, 18, 29, 30, 31]. Other studies have examined the effects of domain knowledge on navigation in hypertext systems, and found that it affected use of links and topic abandonment [7], as well as search efficiency [24].

Wildemuth's [34] study can be used as an example of an approach to understanding the effects of pre-existing domain knowledge on the search process. The searches conducted by the study participants were not defined as exploratory searches, but they might be considered as examples of this type, since students were using searches in a factual database in microbiology to help them solve simulated clinical problems (i.e., they were learning about the material in the database in preparation for their course exam and were simultaneously investigating the specific clinical problem described in the simulation). Figure 1 (columns labeled 1-3) illustrate the three tactics most often used for searches before the course, when the students' domain knowledge was very low. The tactics shown in columns 1 and 2, as well as a tactic including only the selection of a new concept and a display of the results, were used at the end of the course (when domain knowledge was particularly high). Six months after the course ended, when domain knowledge had dropped somewhat, the study participants continued to use the tactic shown in column 1 and also used the tactic shown in column 4 (a slightly less efficient tactic than that developed in their searches just after the course).

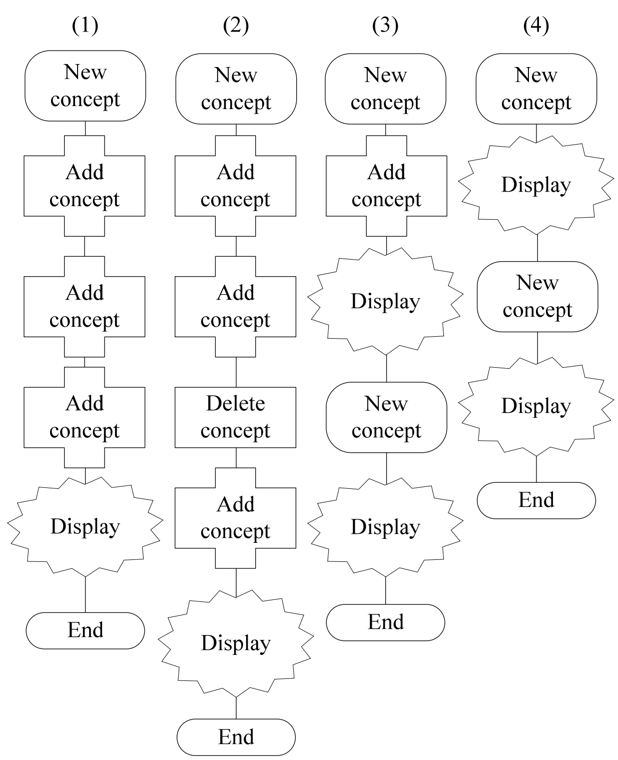

Figure 1: Tactics used by medical students (adapted from Figures 2 and 6, Wildemuth, 2004)

Wildemuth's [34] study, like most of the other studies cited above, measured domain knowledge before the search process was begun, and used that as a variable in later analyses of the search process or search results data. This approach is based on the assumption that domain knowledge is fairly stable and will not vary from the beginning to the end of the search episode. However, the definition of exploratory search implies that this assumption may be false. During exploratory search, it is expected that one of the outcomes of the search process is a change in the searcher's knowledge of the domain brought about through encounters with new information during the course of the search process. Thus, domain knowledge can be expected to have even more pronounced effects on exploratory search than on look-up searches, and these effects will be manifest at the beginning of the search process and throughout the search process trajectory.

THE TRAJECTORY OF THE SEARCH

As noted above, the searcher's domain knowledge as the search begins is likely to have an effect on the trajectory of a particular search. In addition, domain knowledge may be altered during a search. Consider an example. After attending a workshop on research ethics, a doctoral student saw an ABC "Primetime Live" show and became interested in Milgram's experiments in the 1960's. She picked up the transcript of the show from *Lexis Nexis*, but decided to look further. A quick *google* search sent her to *Wikipedia*, where she learned a bit more background and followed several links to other *Wikipedia* articles. The *google* search also sent her to a number of other sites about Milgram and his experiments on obedience. She then went back to *Psychological Abstracts* and found a review of a book by Arthur Miller that discusses the Milgram experiments as a case study in methodological controversy. This search concluded with her going to the library and checking out Miller's book. A more fine-grained look at this example of an exploratory search would include her interactions with (i.e., the tactics used in) each formal information source (*Lexis Nexis*, *google*, *Psychological Abstracts*, the library catalog). The searcher's knowledge about the Milgram experiments was meager at the beginning of the process, but by the time she saw the review of Miller's book, she had a good understanding of the original experiments and people's objections to them. Her level of domain knowledge evolved during the search process, enabling her to make "stopping" decisions in *google* and *Psychological Abstracts* when she knew enough to move in a different direction. The search trajectory was likely affected by the student's interests in developing her own research career in the social sciences, her own pre-existing domain knowledge, her ability to learn during a search, and aspects of her personality (e.g., the curiosity and persistence found in many doctoral students). If this story had a different main character, it is likely that it would have a different plot. This difference reflects the influence of subjectivity on the trajectory of a search.

Search trajectories have been investigated to some extent. Several of the studies cited above include explicit analysis of the steps taken during a search. Additional studies have examined search tactics more generally. For example, Rieh and Xie [25] identified eight distinct patterns of search modification sequences when they examined 313 Web searches. Clark et al. [9] integrated the hypertext navigation patterns identified by Canter et al. [6] – path, ring, loop, and spike – with the general browsing strategies that people use (e.g., scanning would involve a mix of deep spikes and short loops). Curzon, Wilson, and Whitney [10] examined older adults' use of both print-based information resources and the Web, to see the ways in which each type of source was used. They found that this group had well-rehearsed tactics for use in traditional sources which did not always transfer gracefully to the Web.

In addition to general studies of search processes, two recent studies have investigated the effects of individual differences on the search trajectory. Ford, Miller and Moss [15] found that several aspects of cognitive style and the searcher's ability to deal with cognitive complexity affected use of a Boolean search strategy versus a best-match search strategy; they also found that age and gender were related to choice of search strategy. Graff [17] also found that cognitive style and age were related to the number of pages visited in two different hypertext architectures, though they were not related to the browsing strategy (surface versus deep) used. Based on these results, additional studies are needed to identify which individual characteristics have the strongest influence on the search process.

THE CONCLUSION OF THE SEARCH

As a search is concluded, the searcher will make judgments about whether it has been successful. For exploratory searches, one would expect that those judgments would be based on whether the searcher learned enough to be satisfied and/or whether the question being investigated was satisfactorily resolved. In other words, the searcher's judgment about search success will be based on the information *used* in accomplishing his or her goals, not just on the information *retrieved* during the course of the search.

For researchers, the challenge posed by these judgments is that they are based on goals that are likely to be changing throughout the course of the search – particularly for exploratory searches. In the story in the previous section, the student began with the goal of learning more about the Milgram experiments. By the time she started reading Miller's book, her goal was to learn more about how methodological controversies in the social sciences are resolved. If one takes an extreme position, one would say that a search can *not* be considered successful *unless* the search goals have changed as a result of learning that occurred during the search process. Thus, it is not clear which goal should be used in an evaluation, or whether some type of "goal-free" evaluation [28] should be conducted.

IMPLICATIONS FOR EVALUATION AND DESIGN

From the discussion above, it is clear that traditional measures of retrieval success (e.g., precision and recall) will not be appropriate for evaluating the success of exploratory searches. Yet some type of reliable and valid measurements must be taken if we are to compare the effectiveness of one system in supporting exploratory searches with the effectiveness of another system.

A first step may be to rely on users' perceptions of success as a basis for evaluation. There are at least two possibilities here: to ask users for their judgments of the success of a search or to ask users about their satisfaction with the outcomes of the search. While both of these measures might rely on some type of questionnaire, each has a different focus [12]. The first is asking people about their beliefs, or their assessments of what they think is true or false. Specifically, it would ask people for their judgments about whether the search was successful or not, and is simply a subjective measure of system performance. The second type of questionnaire is asking people about their attitudes, or how they feel about the search outcomes. This type of measure has an affective component not present in a questionnaire focused on beliefs.

At present, the field has no reliable and valid questionnaire that can be used for either of these measures;[1] development of such questionnaires would be a significant contribution to the field. In the meantime, some measures of aspects of user beliefs or attitudes might be applied, including Davis' [11] measure of perceived usefulness, Doll and Torkzadeh's [14] measure of satisfaction, or other measures adapted from their use in the fields of information systems or human-computer interaction.

If we can find a way to evaluate search outcomes, then we could leverage those measures to better understand which search *processes* are most effective. For example, in a given study, the search tactics used by the most successful searches could be analyzed and compared with the tactics used in less successful searches. Based on these types of comparisons, we may be able to identify particular patterns that lead to better outcomes, thus enabling direct evaluation of search processes. In addition, we could evaluate users' satisfaction with the search process, including measures of satisfaction with the user-system interaction [8] or measures of flow [16] and other affective responses.

Being able to evaluate the success of a search also provides leverage for improving retrieval system design. Given the subjective character of exploratory search, it seems likely that personalization will play a key role in the design of systems intended to support such searches. The user

---

[1] An alternative – an auction based on the perceived value of a system – has been proposed by Ben-Bassat, Meyer, and Tractinsky [2006]. They argue that bids placed in such an auction are highly related to user performance.

attributes that will be the most effective levers of personalization will be discovered most readily only when we have a means for telling a successful search from an unsuccessful search. Thus, developing valid evaluation methods is critical to progress in the field.

REFERENCES

1. Allen, B. Topic knowledge and online catalog search formulation. *Library Quarterly 61*, 2(1991), 188-213.
2. Allen, B. Logical reasoning and retrieval performance. *Library & Information Science Research 15* 1(1993), 93-105.
3. Ben-Bassat, T., Meyer, J., and Tractinsky, N. Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transactions on Computer-Human Interaction 13*, 2(2006), 210-234.
4. Borgman, C.L., et al. Children's searching behavior on browsing and keyword online catalogs: The Science Library Catalog Project. *Journal of the American Society for Information Science, 46*, 9(1995), 663-684.
5. Bradley, J. Methodological issues and practices in qualitative research. *Library Quarterly 63,* 10(1993), 431-449.
6. Canter, D., Rivers, R. and Storrs, G. Characterizing user navigation through complex data structures. *Behaviour & Information Technology 4,* 2(1985), 93-102.
7. Carmel, E., Crawford, S., and Chen, H. Browsing in hypertext: A cognitive study. *IEEE Transactions on Systems, Man, and Cybernetics 22*, 5(1992), 865-884.
8. Chin, J.P., Diehl, V.A., and Norman, K.L. Development of an instrument measuring user satisfaction of the human-computer interface. *CHI '88 Proceedings,* (1988), 213-218.
9. Clark, L., Ting, I., Kimble, C., Wright, P. and Kudenko, D. Combining ethnographic and clickstream data to identify user Web browsing strategies. *Information Research 11*, 2(2006), paper 249. Available at http://InformationR.net/ir/11-2/paper249.html.
10. Curzon, P., Wilson, J., and Whitney, G. Successful strategies of older people for finding information. *Interacting with Computers 17,* 6(2005), 660-671.
11. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13,* 3(1989), 319-340.
12. Dillman, D.A. *Mail and Telephone Surveys: The Total Design Method.* John Wiley & Sons, New York, NY, 1978.
13. Dimitroff, A., and Wolfram, D. Searcher response in a hypertext-based bibliographic information retrieval system. *Journal of the American Society for Information Science 46,* 1(1995), 22-29.
14. Doll, W.J., & Torkzadeh, G. The measurement of end-user computing satisfaction. *MIS Quarterly 12*, 2(1988), 259-276.
15. Ford, N., Miller, D., and Moss, N. Web search strategies and human individual differences: Cognitive and demographic factors, internet attitudes, and approaches. *Journal of the American Society for Information Science & Technology 56,* 7(2005), 741-756.
16. Ghani, J.A., Supnick, R., and Rooney, P. The experience of flow in computer-mediated and in face-to-face groups. *Proc., 12th Int. Conf. on Information Systems,* (1988), 229-237.
17. Graff, M. Individual differences in hypertext browsing strategies. *Behaviour & Information Technology 24,* (2005), 93-99.
18. Hsieh-Yee, I. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science, 44,* 3(1993), 161-174.
19. Jacobson, T., and Fusani, D. Computer, system, and subject knowledge in novice searching of a full-text multifile database. *Library & Information Science Research 14,* 1(1992), 97-106.
20. Marchionini, G. Exploratory search: From finding to understanding. *Communications of the ACM 49,* 4(2006), 41-46.
21. Marchionini, G. Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science 40,* 1(1989), 54-66.
22. Marchionini, G., Dwiggins, S., Katz, A., and Lin, X. Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library & Information Science Research 15,* 1(1993), 35-69.
23. McDonald, S., and Stevenson, R.J. Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers 10,* 2(1998), 129-142.
24. Rieh, S.Y., and Xie, H.(I.) Analysis of multiple query reformulations on the Web: The interactive information retrieval context. *Information Processing & Management 42,* (2006), 751-768.
25. Saracevic, T., and Kantor, P. A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science 39,* 3(1988), 197-216.
26. Schwartz, H., and Jacobs, J. *Qualitative Sociology: A Method to the Madness.* Free Press, New York, NY, 1979.
27. Scriven, M. Goal-free evaluation. In House, E.R. (ed.), *School Evaluation: The Politics and Process.* McCutchan, Berkeley, CA, 1973.
28. Shute, S.J., and Smith, P.J. Knowledge-based search tactics. *Information Processing & Management 29,* 1(1993), 29-45.
29. Vakkari, P. Subject knowledge, source of terms, and term selection in query expansion: An analytic study. *Advances in Information Retrieval: 24th BCS-IRSG European Colloquium on IR Resarch, Proceedings* (2002), 110-123.
30. Vakkari, P., Pennanen, M., and Serola, S. Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management 39,* 3(2003), 445-463.
31. White, R.W., Kules, B., Drucker, S.M., and schraefel, m.c. Supporting exploratory search. *Communications of the ACM 49,* 4(2006), 37-39.
32. Wildemuth, B.M., de Blied, R., Friedman, C.P., and File, D.D. Medical students' personal knowledge, searching proficiency, and database use in problem solving. *Journal of the American Society for Information Science, 46* (1995), 590-607.
33. Wildemuth, B.M. The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science & Technology 55*, 3(2004), 246-258.