

WEB INFORMATION GATHERING TASKS

A Framework and Research Agenda

Anwar Alhenshiri, Carolyn Watters, Michael Shepherd and Jack Duffy
Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, NS, Canada

Keywords: Web Information Retrieval, Web Tasks, Information Gathering, User Behaviour.

Abstract: This paper provides in-depth analysis of Web information gathering tasks. Research has focused on categorizing Web tasks by creating a high-level framework of user goals and activities on the Web. Yet, there has been very limited emphasis on improving the effectiveness of Web search for information gathering under the concept of a complete task. This paper provides a framework in which subtasks underlying the overall task of Web information gathering are considered. Moreover, the paper provides research recommendations for techniques concerning collecting and gathering information on the Web.

1 INTRODUCTION

Web information retrieval has been studied in the light of request-response for a relatively significant period of time. The user submits a query trying to convey their information need to the Web and in return, they receive a response from the search engine in the form of document hits. In many occasions, a search activity may necessitate that the user continues interacting with the search engine to achieve a higher-level Web task (Kules, et al., 2008). Research has studied user tasks in order to identify a task framework that would help with understanding user interactions with the Web (Byström and Hansen, 2005). Web tasks have been classified into fact finding, navigation, performing a transaction, and information gathering (Broder, 2002; Kellar, et al., 2007). The latter represents a great portion of the overall tasks on the Web, between 51.7% (Broder, 2002) and 61.5% (Rose and Levinson, 2004). Information gathering tasks imply several steps and sequences within each step, longer search time than other types (Mackay and Watters, 2008), and looking at several sources of information to achieve the overall task (Terai, et al., 2008). This type of task is common when a user is completing a report or a project using information sources published on the Web.

Current Web search and gathering techniques provide limited support for the characteristics and procedures involved in the information gathering

task. Web search is a one-session process—in most cases—where the Web search engine provides no means for connecting one search activity to the rest of the activities in the task. Since information mismatching and overloading are two significant problems regarding how search engines gather information (Tao and Li, 2009), it becomes the user's role to locate, compare, and manage the required information in the task. A Web search engine sees the sequences of a task as separate interaction steps. It also provides no means for re-finding information (Tauscher and Greenberg, 1997), which is an activity that represents one third of the user interactions during information gathering tasks according to Kellar and Watters (2006). Moreover, search engines do not usually provide support for representing task results according to the type of information being sought in the task. Consequently, there is a very limited understanding by the design of current search engines of the fact that a search operation may not be just a one-time query, but rather a more complete and sophisticated task.

This article provides a framework in which subtasks underlying the information gathering task are identified. In addition, based on research in the literature regarding information gathering subtasks, the paper provides practical recommendations for Web tools intended for information gathering. The paper is divided as follows. Section 2 describes the related work. Research work concerning information gathering on the Web is discussed in this section. A

framework for the subtasks that comprise the overall task of Web information gathering is also provided in this section. Section 3 provides a discussion of the research findings indicated in Section 2. Section 4 presents practical recommendations for future studies regarding the concept of Web information gathering tasks and improving the effectiveness of tools intended for this type of task. Section 5 concludes the paper and highlights future research directions.

2 WEB INFORMATION GATHERING

This section discusses research work related to the concept of information gathering on the Web. The aspects of information gathering that are researched in the literature are discussed first. The discussion follows by providing a framework for subtasks that comprise the overall task of information gathering. Techniques intended for improving particular aspects of each subtask are discussed along with the type of subtask being explored.

2.1 Research Rationale

Information gathering tasks involve collecting information possibly of different types from different sources to achieve an overall goal identified in the task. Information gathering tasks are mostly search-based as shown by Kellar, et al. (2006). In addition, information gathering is recognized as the most frequent task in re-finding information on the Web (Kellar, et al., 2006). Information gathering tasks have been studied as a part of user interactions with the Web for searching and navigation as discussed by Kules, et al. (2008) and Alhenshiri, et al. (2010, 1). However, there has been little effort to connect the concepts of finding, re-finding, comparing, goal identification, and decision making for the purpose of investigating improvements to information gathering tools on the Web.

Research has examined those aspects in isolation without specific focus on evaluation within the context of a complete task. Yamada and Kawano (2009) used sections in Web pages located for an information gathering task to extract links to other pages. The target pages are considered a part of the user plan for the task and suggested to the user to continue gathering information. In a similar approach, Bagchi and Lahoti (2009) used hyperlink

connectivity among Web pages to assist users in gathering information on the Web. They argued that providing links to pages currently being viewed by the user can facilitate the process of information gathering. However, the only part of the information gathering task considered in these two studies was locating the intended information, i.e. finding. Dearman, et al. (2008) investigated the subtask of the information gathering task that concerns information sources. Re-finding information on the Web was also investigated either with respect to locating previously found results (Tauscher and Greenberg, 1997), or monitoring Web sources of information (Kellar, et al., 2007). Issues with how users deal with information gathering and how they manage their time for the task were discussed in the work of Murphy (2003). Finally, decision making was investigated and considered as an intermediate step in information gathering tasks (Yamaguchi, et al., 2004).

In addition to those aspects that are involved in information gathering, user interactions with the Web have been studied in many directions under different objectives. Rose and Levinson (2004) attempted to identify a framework for user search goals using ontologies in order to understand how users interact with the Web. He and Goker (2000) and Jansen, et al. (2007) attempted to identify boundaries among user search sessions to be potentially able to decide on the user search goal in each session. Both studies intended to improve the effectiveness of the Web search process by providing more suitable results to the user's goal. Broder (2002) studied different user interactions during Web search and identified three types of tasks, namely: transactional, informational, and browsing. Similarly, Kellar, et al. (2007) classified Web tasks into navigation, information gathering, and fact finding. These categorizations provided a framework for the high-level types of tasks users perform on the Web. Consequently, such classifications can further be exploited to improve the process of task accomplishment on the Web for each type of task.

With respect to research regarding how users gather information on the Web, several questions remain open for further investigation. The concept of information gathering remains unclear with regard to the effectiveness of the tools used for gathering and comparing Web information and the challenges the user encounters during the gathering process. In addition, most of the conducted research in Web information retrieval attempted to improve aspects of the subtasks underlying information

gathering without considering the contribution of the context of a whole task to the gathering process. Studying the subtasks that comprise the overall task of information gathering may permit for better understanding of this type of task. In addition, it may permit for further improvements in the field of information retrieval since a great portion of users' search activities on today's Web are considered parts of broader tasks. During information seeking, the user information need is shown to be motivated by a higher-level work task (Byström and Hansen, 2005; Kules, et al., 2008). Before tackling the issue of targeting possible improvements to the effectiveness of Web information gathering tools, it is crucial to understand the components (subtasks) of the information gathering task and some examples of the research that has been conducted so far for investigating certain aspects in each subtask. The following section illustrates the subtasks involved in the Web information gathering task.

2.2 Subtasks in the Web Information Gathering Task

The information gathering task can be studied effectively by investigating the subtasks comprising the overall task. Research in the literature reveals those subtasks through how scholars investigated Web information gathering. Based on the definition provided earlier and the different aspects of Web information gathering that have been studied and investigated in the literature, the subtasks involved in Web information gathering, which are shown in the provided framework in Figure 1, can be summarized in the following:

2.2.1 Interpreting the Task

Web information gathering tasks can be of varied complexities. Interpreting the task is a concern to the user and the tools used in the task. For users to start performing information gathering, they have to make a decision about the information required in the task, the plan desired for performing the task, and the tools to be used for accomplishing the task (Yu and Lau, 2005; Terai, et al., 2008). Interpreting the task includes identifying the information required to be retrieved in the task, the sequences and steps required to achieve the task, and the information given in the task as a priori (Bell and Ruthven, 2004). The user's interpretation determines the tools needed in the task and their effectiveness. On the Web side, current Web information gathering tools, including search engines, do not take into

account user differences. In addition, the type of task performed on the Web cannot be identified easily by relying only on search queries. Unless additional information are provided by the user to the search interface—such as in the form of a user profile—the search engine cannot take into account the type of user or the type of task being performed. With information gathering tasks, the difficulty in identifying the task and the information required in the task increases because of the different possible information and sources the task may require the user to locate on the Web.

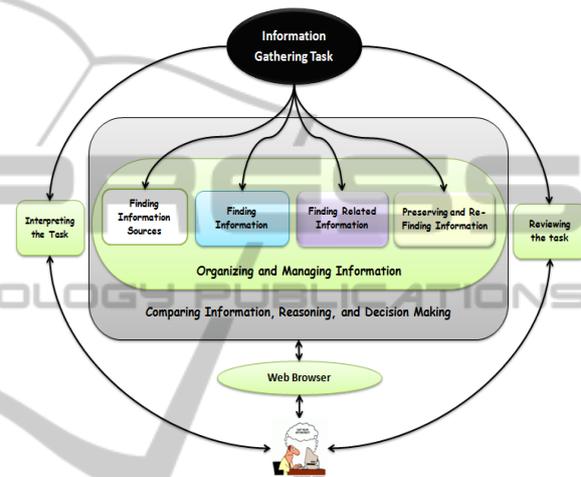


Figure 1: A framework for subtasks in the information gathering task.

2.2.2 Finding Sources of Information on the Web

The Web search engine is the tool predominantly used for this subtask (Teevan, et al., 2004; Kellar, et al., 2006). The user conveys their information need to the search engine in the form of a search query and receives a set of information sources that match the search query but not necessarily satisfies the intended information need (Manning, et al., 2008). A study comparing users search behaviour shows that 55% of users' search behaviour involves keyword search to locate sources of information instead of typing-in a URL into the Web browser (Teevan, et al., 2004). In addition, 57% of internet users use search engines daily (Hsieh-Yee, 2001). Therefore, the search engine is recognized as the most used tool for this subtask. The rest of the subtasks in information gathering are performed by the user on the Web browser using different utilities.

With regard to finding sources of information, research has focused on improving the relevancy of Web search results to match the user's information

need (Manning, et al., 2008). There are several aspects of the Web search process that have been investigated including indexing (Srihari, et al., 2000), query matching (Kawano, 2000, Spink, et al., 2001), search results ranking (Zhuang and Cuserzan, 2006; Zitouni, et al., 2008; Wang, et al., 2009), and search results presentation (Alhenshiri, et al., 2010 (1), Teevan, et al., 2009). The latter aspect is concerned with interacting with the user and allowing them to perceive interesting information. Search engines usually provide high recall of relevant documents, but the results are incorrectly ranked and presented to the user. Consequently, the effectiveness in finding the intended sources is usually concerned with how the results are presented to the user (Alhenshiri and Blustein, 2010). In the literature, there are several suggested improvements with regard to results presentation. The two main concepts regarding investigating such improvements are visualization and clustering. However, there is little focus on the specific aspects of visualization and clustering that would particularly improve gathering sources of Web information.

2.2.3 Finding Information on the Web

The result hits provided by search engines represent sources of possible information of interest to the user. The following subtask in information gathering is locating task-relevant information among the provided sources. This stage in information gathering has been researched in several directions. On the Web browser side of the subtask, results presentation has been rigorously investigated for providing recommendations for effective search interfaces. Different forms of textual presentations (Alonso and Baeza-Yates, 2003), visual presentations (Bonnell, et al., 2005, 2006), and a mix of both textual and visual presentations (Mukherjea and Hara, 1999; Kunz and Botsch, 2002; Rivadeneira and Bederson, 2003; Brown, et al., 2003; Suvanaphen and Roberts, 2004) have been investigated. Clustering of search results according to different criteria was also considered (Carpineto, et al., 2009).

Nonetheless, this subtask is usually studied as a part of the previously discussed subtask in which there is no obvious separation between locating an information source and locating information of interest on that source. The separation is actually apparent. For example, users usually cannot make decisions only by relying on the list of hits provided by the search engine. Finding sources of information is actually a different subtask from finding

information because of trust and familiarity issues with Web sources (Alonso and Baeza-Yates, 2003). Teevan, et al. (2009) showed that presenting Web documents using visual snippets that consisted of the most important image on the page (i.e. the page logo) accompanied with text found in titles on the page was favoured over text only summaries. Presenting more features of the page in the result set was more effective because users recognized the nature of the document and were able to make more effective decisions. The visual snippets and the visualized glyphs in the work of Alhenshiri, et al. (2010, 1) presented actual information about the sources located by the search engine. In the comparison study conducted by Alhenshiri, et al. (2010, 1), participants who used Google opened more pages on the browser and submitted more queries in order to achieve the information gathering task. The results showed that users were less confident about the sources located by Google because they were only able to see the text summaries.

2.2.4 Finding Related Information

Finding related information to the already identified information in the sources provided by a search engine is a subtask that is common in information gathering. The user finds a source of information and continues looking for task-related information in one of two ways. First, when clustering is involved in the presentation of Web documents, the user may look for similar documents to the one of interest by relying on clusters of related documents (Carpineto, et al., 2009). The second approach is by following anchors on the page of interest for the purpose of finding similar information (Karim, et al., 2009; Alhenshiri, et al., 2010 (1)). For example, Google provides clustering in the “*see similar*” feature underneath some of the result hits. The search engine Clusty (www.clusty.com) performs unsupervised clustering and presents categories of topics on a sidebar. Yahoo directories are an example of human-clustered hierarchy of Web documents intended for finding related information to categories of interest. Clustering on the Web is a concept intended for better topical coverage which may assist the user in information gathering tasks. On the Web browser, following anchors on a page and which link to other pages may indicate similar content (Bederson, et al., 1996; Karim, et al., 2009; Alhenshiri, et al., 2010 (1)).

Finding related information is a subtask that is usually intended for gathering further information

and comparing already gathered information for reasoning and decision making. Consequently, it can be considered a separate subtask from collecting sources and information on the Web. The study conducted by Alhenshiri, et al. (2010, 2) showed that users followed the link hierarchy on the located Web information sources in order to make confident decisions about the task results. Similarly, Karim, et al. (2009) developed a technique that gathers hyperlinks on a page and provides those links accompanied by viewing popularity statistics at the bottom of the page. Those anchors helped users to decide whether or not to follow a certain navigation path for finding related information. In information gathering tasks, locating information is usually followed by looking for more relevant information to the task topic for comparisons and decision making. Research has shown that different search and navigation interfaces achieved different effectiveness results (Alhenshiri, et al., 2010, 2). Consequently, locating information related to Web sources and already gathered information is an important subtask that should be further investigated in Web information gathering tasks.

2.2.5 Comparing Information

Comparing information located for the purpose of the task happens on the browser side of the retrieval process. The user performs such comparisons in different ways—yet mostly by reading text on the presented Web pages (Roberts, et al., 2002). The comparison process is meant for making decisions about the types of information required in the task (Zilberstein and Lesser, 1996). In current Web search techniques, comparing information requires reading a lot of text and scrolling over multiple sources of information (Spink, et al., 2001). Visualization is suggested to help with this process by providing multiple features of the presented Web documents to assist the user in making faster and more effective decisions (Nguyen and Zhang, 2006; Wiza, et al., 2004). Clustering Web information by providing meaningful labels may also assist users comparing sources of information. This subtask is involved in all of the subtasks comprising the overall information gathering task.

Comparing information is an important subtask in information gathering that has been investigated in isolation. Suvanaphen and Roberts (2004) designed a search interface that allows users to compare sets of results rendered to multiple queries. The objective was to permit users to observe similarities and differences among the result sets,

reduce the cognitive effort that would result from switching from one result set to another, and enable them to browse more effectively. Similarly, Havre, et al. (2001) introduced Sparkler, a technique that visualizes the results of multiple queries generated as alternatives to a user query. The interface also shows the contribution of each query alternative/component to the overall relevance of documents in the result set. The usability test showed that users preferred Sparkler to the row presentation due to the ability to observe the differences between the initial query and its alternatives in the result set using the visual presentation of Sparkler. Comparing information is a common subtask in Web information gathering. Enhancing the effectiveness of how users perceive and compare information requires further investigations in the context of a complete information gathering task with a defined task goal.

2.2.6 Preserving and Re-finding Information

Information gathering tasks usually happen over the course of multiple sessions (Spink, 1996; Mackay and Watters, 2008). According to Sellen, et al. (2002), 40% of information gathering tasks took more than one session. Therefore, some subtasks such as finding related information and comparing information located for the task may require preserving some or all of the information that were retrieved in previous sessions. Research regarding re-finding information on the Web has investigated several techniques in the Web browser including the back button, the browser history, and the list of favourites and bookmarks. In addition, alternative methods with similar behaviour to the aforementioned techniques were investigated including the mouse flick gesture for the back and front buttons (Moyle and Cockburn, 2003), the use of *Bookmaps* for visualizing the browser history and bookmarked pages (Mountaz, 2000), and the use of *Landmarks* for visual presentation of the browser history (Mackay, et al., 2005).

Preserving search results of previous sessions to be involved in later activities has also been studied in the work of Teevan (2008). However, it remains unknown which technique is the most effective with regard to information gathering tasks. This is so because visualization studies, such as in the work of Yamaguchi, et al. (2004) and Mackay, et al. (2005), measured how effective the presentation was in permitting the user to only find previously preserved documents. The effectiveness of involving re-finding in comparing information within an

information gathering task has not yet been investigated. In addition, re-finding Web documents for re-visitation requires more investigations regarding not only ranking the mix of fresh Web results and the previously preserved ones, but also with consideration to the results presentation.

2.2.7 Organizing and Managing Information

Organizing and managing information during Web information gathering tasks is an important subtask. Research has focused on investigating how users manage their information for re-finding (Jones, et al., 2003; Mackay, et al., 2005) and how they view and manage desktop information in general (Knoll, et al., 2009). Important reasons behind giving up on certain personal information management tools were discussed in the work of Jones, et al., (2008). Strategies users follow to manage Web information in order to be able to relocate and reuse previously found information are discussed in the work of Jones et al. (2003).

The work of Jones, et al. (2003) showed that users—while gathering Web information—follow different preserving strategies to re-find and compare information later. Most users gather information over multiple sessions (Spink, 1996; Mackay and Watters, 2008), which indicates the need for management strategies for preserving and re-finding such information for reuse. The variety of finding, re-finding, organizing, and management strategies and approaches users follow while seeking and gathering Web information can be related to the fact that current Web tools lack important reminding, integration, and organization schemes.

Jones, et al. (2008) found that users abandon the use of an information management tool for one or more of five closely related reasons which are: visibility, integration, co-adoption, scalability, and return to investment. These reasons need to be further investigated in the case of Web information gathering. The Web may reveal further reasons why users use certain tools over others, why they do not use the same tools, what tools do most users actually use to keep track of their gathered information, and how they maintain the consistency of their located information. Other questions may include what tools are actually supportive to information organization and management during information gathering, if any? Research has little consideration to factors that would improve how Web users collect, manage, compare, and organize their information for information gathering tasks.

2.2.8 Reviewing the Task

During information gathering, reasoning, and decision making may occur at any time depending on the task, the user expertise, and the tools used in the task (Adar, et al., 2008). The process of accomplishing the overall Web information gathering task is affected by the user's short term memory, the number of sequences required in the task, and the type of information being searched. These factors necessitate that the user revisits and reviews the task to make sure that the requirements are accomplished and to make a decision about the completion of the task. This subtask is an important factor that has to be further investigated in the presence of other subtasks in Web information gathering in a controlled environment. Information gathering tools and how information is provided to the user to collect, compare, and make decisions about the task should be further investigated.

3 DISCUSSION

Research has, so far, identified information gathering as a very common activity on the Web and which has its own characteristics. Information gathering is a task concerned with collecting information of various types from different sources to satisfy a higher-level goal (Kellar and Watters, 2006). Information gathering usually takes more time than other tasks (Mackay and Watters, 2008), happens over the course of multiple sessions, and has no specific tools that take the whole task into consideration. Research has investigated several aspects in Web information gathering. However, there has been no consideration of the context of the overall task in the investigation studies.

Visualization and clustering are two important factors that have been investigated with regard to improving the effectiveness of Web search tools. Nonetheless, investigation has only been applied to certain aspects of the subtasks in the information gathering task as discussed above. In Web information gathering, the concept of a complete task should be further considered. The tools used in the task and the challenges the user encounters while trying to locate sources of information, compare information and sources, re-locate information, and find more related information should be investigated.

4 RESEARCH RECOMMENDATIONS

Regarding research intended for investigating improvements to each of the subtasks discussed above, and hence to the overall task of Web information gathering, several important practical recommendations are summarized in the following points.

4.1 Gathering Web Information Sources

Gathering sources of information should be investigated by using visualization and clustering while emphasizing issues of trust and familiarity with the sources being gathered and the tools used in the gathering process. Using visualized features of Web documents can help users make effective decisions about the source of information being presented. However, cluttered presentations through the use of certain visualization layouts may degrade the effectiveness of the interface since users are practically used to raw text-based presentations (Alonso and Baeza-Yates, 2003).

4.2 Gathering Web Information

Gathering Web information should be investigated with regard to the presentation of Web information. How many features the user can perceive at once on the display is a crucial factor in satisfying the user information need. Moreover, efficiency is a very important factor since most users tend to look at very few items in the search results list (Spink, et al., 2001). Consequently, the type of presentation that would assist users to find interesting information and locate such information efficiently and effectively should be investigated. The presentation should involve aspects of visualization and textual presentation allowing the user to choose the view that suites the user information need and the topic of the task.

4.3 Finding Related Information

Finding related information to the already gathered sources and information can be improved by utilizing clustering and visualization. Clustering can assist users trying to locate Web information related to sources gathered using search engines. In addition, providing overviews of the hierarchy of Web domains can assist users with gathering information by navigation. Moreover, a Web

information gathering task may be concerned with collecting information that belongs to different sources and topics. Clustering may play a significant role in improving this process. However, investigation is needed with regard to the most effective clustering criteria, i.e. genre-based and/or topic-based clustering. In addition, the type of presentation of the clustered results that would benefit collecting related information should be further studied, i.e. visual clustering and/or tabular text-based clustering.

4.4 Comparing Web Information

With regard to comparing different types and pieces of information located for a task, visualization can play an important role by providing multiple features of the presented Web documents to assist the user in making faster and more effective decisions. Clustering Web information by providing meaningful labels may also assist users comparing sources of information. Previous research investigated the issue of comparing information in Web search (Havre, et al., 2001; Suvanaphen and Roberts, 2004). However, investigations usually excluded the context of a complete task. Web information is compared for making decisions about the relevancy of results provided for individual search queries. The need is to investigate tools that can be used in reasoning and decision making within the context of Web information gathering tasks.

4.5 Re-finding Web Information

Re-finding information for comparison and decision making has not, so far, been investigated in Web information gathering. In the information gathering task, this issue should be further studied in the context of a complete task. Research has studied re-finding for the purpose of identifying efficient and effective techniques in presenting preserved Web information. Nonetheless, the need is to further reinvestigate such techniques in the context of information gathering and identify features that would help users find, compare, and manage task information.

4.6 Organizing Web Information

On the Web, research has only considered the case of managing and organizing information for re-finding (Jones, et al., 2003). How users organize and manage information during Web information gathering has had minimum consideration. Since

information gathering on the Web may take several sessions, involve looking at information from different sources, and involve comparing information that may belong to varied topics, investigating organizational and management strategies users follow on the Web is necessary. Such investigations would reveal design characteristics regarding tools needed for improving the process of Web information organization and unleash challenges users encounter with current Web tools.

4.7 Interpreting and Reviewing the Task

Interpreting and reviewing the task are important subtasks in information gathering. Research should further investigate these factors within the context of a complete task by investigating how effective the tools used in the task are in limiting the task progress overhead. This can be done by investigating visualization, clustering, and re-finding as discussed above. Moreover, annotation is a concept that can be investigated. Annotation may assist users with managing and comparing the task information especially in the case of a multi-session information gathering task. Research shows that users sometimes find it difficult to look back at preserved bookmarks and documents in the browsing history to re-find information (Mountaz, 2000). Annotation may improve the process of re-finding by searching annotations applied to preserved documents.

5 CONCLUSIONS

This paper presented some of the research that has been conducted regarding gathering Web information. A framework of the subtasks that comprise the overall task of information gathering was developed and illustrated. Some of the research that has investigated different aspects in each of the identified subtasks was also discussed. The paper provided practical recommendations in the area of research concerning how users gather information on the Web. Future work will investigate some aspects of results presentation through the use of visualization and clustering for seeking improvements regarding Web information gathering tasks. In addition, the concept of re-finding will be studied in the light of visualization and clustering in addition to aspects of annotation for improving Web information gathering tasks of the multi-session nature.

REFERENCES

- Adar, E., Teevan, J., and Dumais, S. 2008. Large Scale Analysis of Web Revistation Patterns. *In Proceedings of the 2008 ACM Conference on Human Factors in Computing Systems*, Florence, Italy, 1197-1206.
- Alhenshiri, A., Shepherd, M., Brooks, S., and Watters, C. 2010 (1). Augmenting the Visual Presentation of Web Search Results. *In Proceedings of the 5th International Conference on Digital Information Management*, Thunder Bay, ON, Canada, to appear.
- Alhenshiri, A., Shepherd, M., Watters, C., and Bliemel, M. 2010 (2). Information Gathering within Websites: Visualized Links for Navigation (VLN). *The 3rd International Workshop on Patent Information Retrieval*, Toronto, Canada, to appear.
- Alhenshiri, A., Blustein, J. 2010. Utilizing Visualization for Improving Web Search Effectiveness. *In Proceedings of the i-Society2010 Conference*, London, UK, to appear.
- Alonso, O., and Baeza-Yates, R. 2003. Alternative Implementation Techniques for Web Text Visualization. *In Proceedings of the 1st Latin American Web Congress*, California, USA, 202-204.
- Bagchi, A., and Lahoti, G. 2009. Relating Web Pages to Enable Information-Gathering Tasks. *In Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*, Torino, Italy, 100-118.
- Bederson, B. B., Hollan, J. D., Stewart, J., Rogers, D., Druin, A., and Vick, D. 1996. A Zooming Web Browser. *In SPIE Multimedia Computing and Networking'9*, vol. 2667, 260-271.
- Bell, D. J., and Ruthven, I. 2004. Searchers' Assessments of Task Complexity for Web Searching. *In Proceedings of the 26th European Conference on Information Retrieval*, Sunderland, UK, 57-71.
- Bonnel, N., Cotarmanac'h A., and Morin, A. 2005. Meaning Metaphor for Visualizing Search Results. *In Proceedings of the 9th International Conference on Information Visualization*, London, England, 467-472.
- Bonnel, N., Lemaire, V., Cotarmanac'h, A., and Morin, A. 2006. Effective Organization and Visualization of Web Search Results. *In Proceedings of the 24th IASTED International Multi-Conference on Internet and Multimedia Systems and Applications*, Innsbruck, Austria.209-216.
- Broder, A. 2002. A Taxonomy of Web Search. *ACM SIGIR Forum*, vol. 36, issue 2, 2-10.
- Brown, L. D., Hua, H., and Gao, C. 2003. A Widget Framework for Augmented Interaction in SCAPE. *In Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 02 - 05, 2003). *UIST '03*. ACM Press, New York, NY, 1-10. DOI=<http://doi.acm.org/10.1145/964696.964697>.
- Byström, K., and Hansen, P. 2005. Conceptual Framework for Tasks in Information Studies. *Journal of the American Society of Information Science and Technology*, vol. 56, issue 10, 1050-1061.

- Carpineto, C., Osiński, S., Romano, G., and Weiss, D. 2009. A Survey of Web Clustering Engines. *ACM Computing Surveys*, vol. 41, issue 3, Article No. 17.
- Dearman, D., Kellar, M., and Truong, K. N. 2008. An Examination of Daily Information Needs and Sharing Opportunities. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. San Diego, CA, USA, 679-688.
- Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. 2001. Interactive Visualization of Multiple Query Results. In *Proceedings of the 2001 IEEE Symposium on Information Visualization*, San Diego, California, USA, 105-112.
- He, D., and Goker, A. 2000. Detecting Session Boundaries from Web User Logs. *Paper Presented at 22nd Annual Colloquium of IR Research, Cambridge UK*.
- Hsieh-Yee, I. 2001. Research on Web Search Behavior. In *Library and Information Science Research*, vol. 23, 167-185.
- Jansen, B. J., Spink, A., Blakely, C., and Koshman, K. 2007. Defining a Session on Web Search Engines. *Journal of the American Society for Information Science and Technology*, vol. 58, issue 6, 862-871.
- Jones, E., Bruce, H., Klasnja, P., & Jones, W. (2008). I Give Up! Five Factors that Contribute to the Abandonment of Information Management Strategies. *68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2008)*. Columbus, OH, USA.
- Jones, W., Bruce, H., & Dumais, S. (2003). How do People Get Back to Information on the Web? How Can They Do It Better? 9th IFIP TC13 International Conference on Human-Computer Interaction. Zurich, Switzerland.
- Karim, J., Antonellis, I., Ganapathi, V., and Garcia-Molina, H. 2009. A Dynamic Navigation Guide for Web Pages. In *CHI 2009*.
- Kawano, H. 2000. Overview of Mondou Web Search Engine Using Text Mining and Information Visualizing Technologies. In *Proceedings of the International Conference on Digital Libraries*, Kyoto, Japan. 234-244.
- Kellar, M., Watters, C., and Shepherd, M. 2006. The Impact of Task on the Usage of Web Browser Navigation Mechanisms. In *Proceedings of the Graphics Interface Conference*, Quebec City, QC, Canada, 235-242.
- Kellar, M., and Watters, C. 2006. Using Web Browser Interactions to Predict Task. In *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, 843-844.
- Kellar, M., Watters, K., and Shepherd, M. 2007. A Field Study Characterizing Web-based Information-Seeking Tasks. *Journal of the American Society for Information Science and Technology*, vol. 58, issue 7, 999-1018.
- Knoll, S., Hoff, A., Fisher, D., Dumais, S., & Cutrell, E. (2009). Viewing Personal Data Over Time. *CHI 2009 Workshop on Interacting with Temporal Data*. Boston, USA.
- Kules, W., Wilson, M. L., Schraefel, M. C., and Shneiderman, B. 2008. From Keyword Search to Exploration: How Result Visualization Aids Discovery on the Web. *Technical Report, School of Electronics and Computer Science, University of Southampton*.
- Kunz, C., Botsch, V. 2002. Visual Representation and Contextualization of Search Results: List and Matrix Browser. In *Proceedings of the International Conference on Dublin Core Metadata Applications*, Florence, Italy, 229-234.
- Mackay, B., Kellar, M., and Watters, C. 2005. An Evaluation of Landmarks for Re-finding Information on the Web. In *Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems*, Portland, Oregon, USA, 1609 - 1612.
- Mackey, B., and Watters, C. 2008. Exploring Multi-session Web Tasks. In *Proceedings of the 2008 ACM Conference on Human Factors in Computing Systems*, Florence, Italy, 4273-4278.
- Manning, C. D., Raghavan, P., and Schütze, H. 2008. *Introduction to Information Retrieval*. Cambridge University Press.
- Mountaz, H. 2000. A User Interface Combining Navigation Aids. In *Proceedings of the 11th ACM Conference on Hypertext and Hypermedia*, San Antonio, Texas, United States, 224 - 225.
- Moyle, M., and Cockburn, A. 2003. The Design and Evaluation of a Flick Gesture for 'back' and 'forward' in Web Browsers. In *Proceedings of the 4th Australasian User Interface Conference (AUIC2003)*, Adelaide, Australia, 39-46.
- Mukherjea, S., Hara, Y. 1999. Visualizing World-Wide Web Search Engine Results. In *IEEE International Conference on Information Visualization*, London, UK, 400-405.
- Murphy, J. 2003. Information-Seeking Habits of Environmental Scientists. *A Study of Interdisciplinary Scientists at the Environmental Protection Agency in Research Triangle Park, North Carolina*, Issues in *Science and Technology Librarianship*, Retrieved February 28, 2010, from <http://www.istl.org/03-summer/refereed.html>
- Nguyen, T. and Zhang, J. 2006. A Novel Visualization Model for Web Search Results. *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, Number 5, 981-988.
- Rivadeneira, W., and Bederson, B. B. 2003. A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces, *University of Maryland, HCIL*.
- Roberts, J. C., Boukhelifa, N., and Rodgers, P. 2002. Multiform Glyph Based Web Search Result Visualization. In *Proceedings of the 6th International Conference on Information Visualisation*, London, England, 549-554.
- Rose, D., and Levinson, D. 2004. Understanding User Goals in Web Search. In *Proceedings of the 13th*

- International Conference on World Wide Web*, New York, NY, USA, 13-19.
- Sellen, A., Murphy, R., and Shaw, K. 2002. How Knowledge Workers Use the Web. In *Proceedings of the 2002 SIGCHI Conference on Human Factors in Computing Systems*, Minneapolis, MN, 227-234.
- Spink, A. 1996. Multiple Search Sessions Model for End User Behaviour: An Exploratory Study. *Journal of the American Society for Information Science*, vol. 47, issue 8, 603-609.
- Spink A., Wolfram D., Jansen M., Saracevic T. 2001. Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science and Technology*, vol. 52, issue 3, 226-234.
- Srihari, R. K., Zhang, Z., and Rao, A. 2000. Intelligent Indexing and Semantic Retrieval of Multimodal Documents. *ACM Information Retrieval*, vol. 2, issue 2-3, 245-275.
- Suvanaphen, E., and Roberts, J.C. 2004. Textual Difference Visualization of Multiple Search Results Utilizing Detail in Context. In *Proceedings of the Theory and Practice of Computer Graphics Conference*, Bournemouth, UK, 2-8.
- Tao, X., and Li, Y. 2009. Concept-Based, Personalized Web Information Gathering: A Survey. In *Proceedings of the 3rd International Conference on Knowledge Science, Engineering, and Management*, Vienna, Austria, 215-228.
- Tauscher, L., and Greenberg, S. 1997. How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies – IJHCS*, vol. 47, issue 1, 97-138. *Academic Press. Special Issue on World Wide Web Usability*.
- Teevan, J. 2008. How People Recall, Recognize, and Reuse Search Results. *ACM Transactions on Information Systems*, vol. 26, issue 4. Article No. 19.
- Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. 2004. The Perfect Search Engine is not enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems*, Vienna, Austria, 415-422.
- Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., Andre, P., and Hu, C. 2009. Visual Snippets: Summarizing Web Pages for Search and Revisitation. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, Boston, MA, USA, 2023-2032.
- Terai, H., Saito, H., Egusa, Y., Takaku, M., Maiwa, M., and Kando, N. 2008. Differences between Informational and Transactional Tasks in Information Seeking on the Web. In *Proceedings of the 2nd International Symposium on Information Interaction in Context*, London, UK, 152-159.
- Wang, A. G., Jiao, J., and Fan, W. 2009. Searching for Authoritative Documents in Knowledge-Based Communities. In *Proceedings of the 13th International Conference on Information Systems (ICIS 09)*, Phoenix, AZ, USA.
- Wiza, W., Walczak, K. and Cellary, W. 2004. Periscope: a System for Adaptive 3D Visualization of Search Results. In *Proceedings of the 9th International Conference on 3D Web Technology*, 29-40.
- Yamada, S., and Kawano, H. 2009. Information Gathering and Searching Approaches on the Web. *Journal of New Generation Computing*, 195-208.
- Yamaguchi, T., Hattori, H., Ito, T., and Shintani, T. 2004. On a Web Browsing Support System with 3D Visualization. In *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers and Posters*, New York, NY, USA, 316-317.
- Yu, Y. T., Lau, M. F. 2005. A Comparison of MC/DC, MUMCUT and Several other Coverage Criteria for Logical Decisions, *Journal of Systems and Software*.
- Zhuang, Z., and Cuserzan, S. 2006. Re-ranking Search Results Using Query Logs. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, Arlington, Virginia, USA, 860-861.
- Zilberstein, S., and Lesser, V. 1996. Intelligent Information Gathering Using Decision Models, Technical Report, 96-35, Computer Science Department, *University of Massachusetts*, Retrieved February 22, 2010, from http://www.agent.ai/doc/upload/200407/zilb96_2.pdf
- Zitouni, H., Sevil, S., Ozkan, D., and Duygulu, P. 2008. Re-ranking of Web Image Search Results Using a Graph Algorithm. In *Proceedings of the 19th International Conference on Pattern Recognition*, Tampa, FL, USA, 1-4.