

IMPROVING HOW USERS ORGANIZE WEB INFORMATION FOR GATHERING TASKS

Anwar Alhenshiri, Michael Shepherd and Carolyn Watters

Faculty of Computer Science, Dalhousie University, 6050 University Ave., Halifax, NS, Canada

Keywords: Web information retrieval, Gathering tasks, Organization and management.

Abstract: This paper presents a Visual Information Management System (VIMS) for the Web. In this prototype system, the task of Web information gathering was taken into consideration with respect to how users locate information for the task, organize task information, preserve and re-find task information, and compare information for effective reasoning and decision making. VIMS was designed and built based on recommendations from previous studies in a larger research. A user study was conducted to evaluate VIMS. The results of the study show promising indications of the improvements achieved in VIMS with respect to Web information gathering tasks.

1 INTRODUCTION

Research has studied user tasks in order to identify task frameworks that would help with understanding user interactions with the Web. Web tasks have been classified into fact finding, navigation, performing a transaction, and information gathering (Broder, 2002, Kellar et al., 2007). The latter type accounts for a large portion of the overall tasks on the Web, representing between 51.7% (Broder, 2002) and 61.5% (Rose and Levinson, 2004).

A Web information gathering task is a composite of subtasks/activities users perform while interacting with the Web for accomplishing a goal described in the task. User activities during Web information gathering may involve finding sources of Web information (Web documents), searching for information on the sources located for the task, finding related information to the already located sources and information, comparing information for reasoning and decision making, organizing task information, and preserving and re-finding information (Alhenshiri et al., 2010b).

Since information mismatching and overloading are two significant problems while search engines gather and present information (Teevan, 2008), 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 activities in a task as separate interaction steps. It also provides no means for re-finding information, which is an

activity that represents one third of the user interactions during information gathering tasks according to Mackay and Watters (2008). Moreover, search engines do not usually provide support for representing task results according to the type of information being sought in the task.

With regard to information gathering tasks, information organization has been barely studied in the context of the Web. In this article, a Web information gathering and organization prototype (VIMS) is presented. VIMS exploits visualization, visual clustering, and several Web information preserving, re-finding, and organization strategies for Web information gathering tasks. The paper is presented as follows. Section 2 discusses related research work. Section 3 presents the research motivations. Section 4 describes VIMS' design in details. Section 5 presents the evaluation user study. Section 6 discusses the research findings. Section 7 concludes the paper.

2 RESEARCH RATIONALE

2.1 Visualizing Web Information

In addition to different prototypes that have been investigated (Bonnell et al. 2006; Teevan et al., 2009), there are several search tools on the Web that use visualization, such as the search engine Viewz (www.viewzi.com). Visualization of Web search

results has also been investigated in several layouts including the use of *hyperbolic trees*, *Self-Organizing Maps*, and *thematic maps* such as in the visual search engine *Kartoo* (www.kartoo.com). Most of these approaches were intended for improving how users find sources of Web information. Exploring multiple features of Web documents such as their content similarities, page thumbnails, URLs, and document summaries in a visualized approach should be investigated in Web information gathering tasks. These features—when visualized properly—can help users find sources of information on the Web, find information in such sources, compare information, and make more effective and efficient decisions.

2.2 Clustering Web Information

In Web information retrieval, clustering has been investigated in several prototypes such as in the work of Alhenshiri et al. (2010a). Clustering has also been implemented in conventional search engines such as Clusty (www.clsuty.com), Gceel (www.Gceel.com), and Google (in their “*see similar*” feature and *Google Wonder Wheel*). Although the performance of users with list presentations of Web documents is comparable to their performance with clustering-based presentations, user preference usually comes in favour of clustering-based methods (Carpineto et al., 2009). In addition, there are indications that clustering can even be more effective (Turetken and Sharda, 2005). With the variety of information that is gathered on the Web, clustering can play a significant factor in Web information gathering tasks.

2.3 Preserving and Re-finding Web Information

Research has focused on enhancing re-finding Web information locally on the Web browser. However, re-finding strategies such as the back button, favourites, and bookmarks can maintain limited numbers of information sources (documents), and they are useful only during a particular Web session. Therefore, searching the Web for re-finding, also known as re-searching (Teevan, 2008), has been studied for assisting users in locating results of interest from previous sessions. Research shows that a great deal of Web search visitations is for revisiting (Teevan, 2008). Consequently, Re-finding is a common activity in Web information gathering tasks accounting for 53.27% according to Mackay

and Watters (2008). For information gathering tasks of multi-session nature, which may require a multi-topic search, re-finding can play a significant role in the effectiveness of tools designed for this type of task. Re-finding should be focused not only on preserving active Web pages in the browser but also on preserving Web search results in the context of a complete task.

2.4 Organizing Web Information

Research has focused on investigating how users manage their information for re-finding (Elsweiler and Ruthven, 2007; Jones et al., 2003; Mackay et al., 2005). 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). Most users gather information over multiple sessions (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 idea that current Web tools lack important reminding, integration, and organization schemes (Cutrell et al., 2006). How users organize and manage information during Web information gathering has had little consideration. Since Web information gathering tasks may take several sessions, involve looking at information from different sources, and require comparing information that may belong to varied topics, investigating organizational and management strategies users follow on the Web is necessary.

3 RESEARCH MOTIVATIONS

To further exploit the concepts of information visualization, visual clustering, re-finding, and organization, VIMS was designed. VIMS attempts to utilize visualization and clustering to allow users to find, compare, and relate information to the already located sources of information more effectively. It also aims to create a more effective storing and re-finding environment. Re-finding is done not only by searching a list of documents, but also by using keyword search to re-find individual documents, sessions, and entire tasks previously stored by the user. Moreover, VIMS is intended to provide effective organizational schemes for information during information gathering tasks.

4 VIMS DESIGN

The VIMS interface was designed using Java swing components and the *prefuse* visualization toolkit (<http://prefuse.org/>). The *CloudGarden GUI builder 'Jigloo'* was also used in the implementation of VIMS. The VIMS interface is shown in Figure 1. Four models were employed in the design and implementation of VIMS: searching, clustering, organizing, and re-finding models.

It has been shown that Web information gathering tasks are largely search based (Alhenshiri et al., 2011). VIMS combines the powers of the Google and Yahoo Web search engines. Moreover, VIMS uses intuitive visual clustering to render its search results and its preserved task and session documents. Clustering is performed based on one of four criteria: network domain, country of origin, genre, and content similarity (*topical clustering*). Clusters are labelled using *cluster-internal labelling*. The title of the document closest to the centroid of the cluster is used as the label of the cluster.

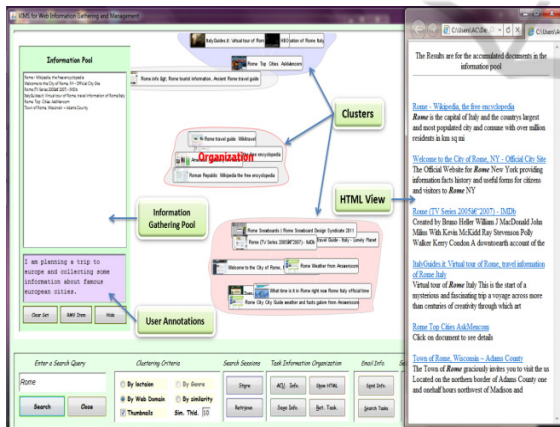


Figure 1: VIMS Interface.

To assist users with organizing task information, VIMS allows the user to store partial information during a task by preserving current session information. This is done by either preserving active visual views of the current display or by selectively preserving particular documents among the search results. Preserved documents are grouped under a task title (name) and sorted by date for later retrieval. The user can continue working on the same task over multiple sessions while adding and eliminating documents. The user can also add annotations to the preserved task information along the way towards completing the task. In addition, search results and task information (documents) can be viewed either visually or in HTML format (as a

list of hits). The study in Alhenshiri et al. (2010a) showed that users prefer to have both views during information gathering.

For serving the process of re-finding information, VIMS allows users to store complete sessions and individual documents. It also allows search within sessions and within tasks by either selecting from a list of tasks/sessions or by keyword search to further assist the user. The keyword search matches the task name given previously by the user and the annotations preserved alongside the task. Moreover, VIMS allows users to email task information including accumulated documents, task subject and date, and task annotations. The emailing strategy was recommended in the work of Jones et al. (2003). However, VIMS adds the organization of a task to the subject matter by submitting all the aforementioned information items. With VIMS, the user can follow the preserving and re-finding strategy that suits their needs and accommodates the task requirements.

5 EVALUATION

To evaluate VIMS, a small-scale study in a field setting was conducted. While the number of participants was limited, the results reflected how early adopters perceive VIMS as a Web information gathering and organization tool. In this study, engagement (enjoyment) factors only were considered in evaluating VIMS. The effectiveness and efficiency of VIMS were not considered by comparing VIMS to other systems; however, how users perceived these factors was taken into consideration in the survey items used for evaluating VIMS.

5.1 Study Participants and Design

Twenty participants took part in the study. They were students from Dalhousie University. VIMS was given to the participants in field settings. The choice of the field setting approach was due to the fact that the researchers wanted to accumulate first impressions about VIMS from early adopters, and to allow users to provide their own judgment where no control of the experiment was in place.

5.2 Study Procedure

Every participant was given a Web information gathering task to perform at the time and place of their convenience. Every task was divided into two

parts to encourage participants to keep information during the first part and to re-find information during the second part.

Information gathering tasks were created to satisfy the criteria described in the work of Kules and Capra (2006). Every participant was given one task in addition to a brief description of how VIMS works. The description illustrated functionalities of the interface and how each feature behaves. There was no training for participants on how to use VIMS. Participants were pointed to the website where they could download VIMS. They were also given the description file to read.

After a two week period, an evaluation form was completed and returned to the principal investigator. The form involved questions for participants that concerned several items including: a) the user satisfaction with VIMS; b) how helpful the visual clustering of search results was to the user; c) how interesting it was to use VIMS to complete the given task; d) how the user perceived preserving and re-finding Web information and Web documents using VIMS; e) how the user rated the presentation of Web search results on VIMS; f) whether or not the self-emailing feature was useful to the user during the task; g) whether the user used VIMS to accumulate and manage task information in the information pool provided in VIMS; h) what the user liked most about VIMS and what they thought should be improved; and i) any other comments the user wanted to add. *Likert* scales were used for answering the survey questions.

5.3 Study Results

With respect to user satisfaction with VIMS, 75% of the participants showed complete satisfaction. Twelve percent indicated that they were neither satisfied nor dissatisfied with VIMS. The remaining participants (13%) showed dissatisfaction with VIMS. They indicated that they either did not like the limit of 120 result hits on VIMS, or that they prefer a list of hits in a textual presentation such as in Google's. The limit of a maximum of 120 results per display was to prevent clutter, a purpose that was not discussed with the participants. The *z-test* shows a significant difference between the proportion of participants who were satisfied with VIMS and those who were dissatisfied ($z=2.84$, $p<0.003$).

VIMS' helpfulness with the given task was also evaluated as one of the engagement factors considered in the study. Most users (80%) used all the features provided in VIMS. The results show that 69% of the users found VIMS to be helpful while

12% found VIMS to be of no help. Participants who found VIMS not to be helpful indicated that they needed further training on some of the features in VIMS and how to use those features. The remaining participants (19%) indicated that they either did not use VIMS for the given task or that they did not use all of VIMS' features.

One of the most interesting organization features in VIMS is how users preserve and re-find documents and information about a particular task or session on the Web. All participants used and evaluated this feature. The results show that 85% of the participants think that VIMS provided effective preserving and re-finding capabilities. The remaining 15% thought that the features were neither effective nor ineffective. None of the participants rated these features negatively. The *z-test* shows that there was a significant difference between the proportion of participants who found VIMS to be effective and those who did not evaluate the feature of preserving and re-finding task information ($z=4.11$, $p<0.0001$).

One of the features in VIMS is the ability to send information to one's email. Information may include the name and date of the current task, documents located for search queries over one or multiple sessions during the task, and user annotations on the task. The results showed that only 50% of the users had the chance to use this feature. One interpretation of the results in this case is that users tended to finish the whole task in one session. In future studies, this feature will be tested during information gathering over multiple sessions in a more controlled environment. Of the remaining 50% of users, only 10% found the feature to be of no usefulness. In long term information gathering tasks such as writing a report, which may be conducted on multiple machines over multiple sessions, this feature is expected to be more effective and more useable.

Another important feature that was evaluated in VIMS is the task organization capability. This is the ability to accumulate task information including documents found for search queries in addition to annotations added by the user regarding the task at hand. The feature allows the user to add and remove documents from the information pool visualized on the display (see Figure 1) over multiple sessions while using the task name and date of creation for access. The results show that 70% of the participants used and evaluated the feature. Of those, 73% think that the feature was either helpful or very helpful. Only 10% think that it was of no help during the task. Of the participants, 17% did not evaluate this feature. The *z-test* shows a significant difference

between the proportion of participants who found VIMS' organizational feature to be helpful and those who found it of no help ($z= 3.83, p< 0.00001$). The results are shown in Figure 2.

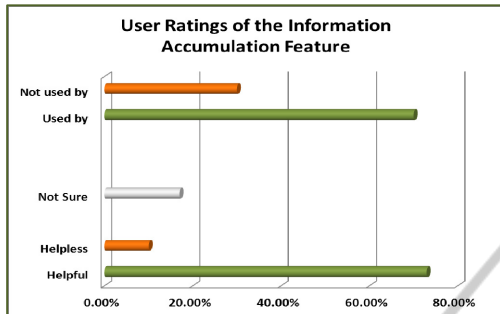


Figure 2: User Ratings of the Info. Accumulation Feature.

The last engagement factor the study tested was the presentation of Web search results to the user. As shown by Alhenshiri et al. (2011), Web information gathering tasks are largely search based. Hence, it was necessary to evaluate the presentation of Web search results on VIMS with the use of visualization and clustering. The results showed that 67% of the participants think that the VIMS presentation was effective. Twenty five percent think that the visual presentation in VIMS is as effective as a list-of-hits presentation. Only 8% of the participants think that the presentation of search results was somewhat cluttered and confusing. The z -test shows a significant difference between participants who found the presentation of Web documents on VIMS to be effective and those who did not ($z=1.87, p<0.04$).

Finally, the survey asked participants about features they liked or found effective in VIMS. The participants left 39 comments. The comments belonged to four main categories which are: a) organizing and managing task results, b) presenting and exploring search results, c) preserving and re-finding Web task information, and d) clustering task results. Most of the comments were in regard to the organization of task information. Organizing and managing task information included storing, re-finding, annotating, and comparing task information. The results of this item are shown in Figure 3. Examples of user comments are:

- *'the colour coding of clustering was really good'*
- *'The presentation of the search results grouped by the country it came from was kind of neat'*
- *'I loved how I can keep accumulating task information over multiple sessions'.*

Users left important feedback regarding how VIMS should be improved. User comments covered different issues but were hard to categorize into specific and distinct groups. Nonetheless, most of the comments concerned interface issues such as the choices of colors and the size of the glyphs that represent Web documents. Clutter was also an issue that was mentioned several times. The total number of comments regarding this criterion was 11. Some users (10%) indicated that they would like to see clustering based on different criteria happen on the fly instead of selecting the clustering criterion prior to sending search queries. Of the participants, one user indicated that the search box should be enlarged; one participant suggested that the user should have the ability to see the ranks of documents inside each cluster; and another participant indicated that there should be more clustering criteria. None of the comments occurred more than twice in the provided set of comments.

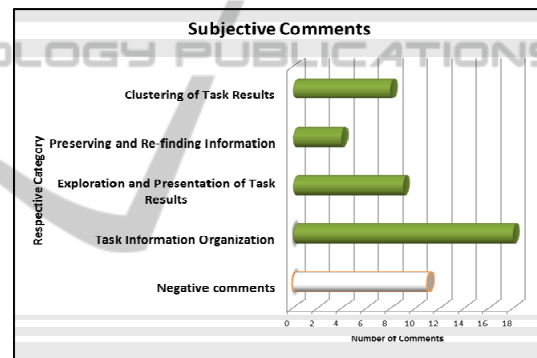


Figure 3: Categorized Subjective Comments.

6 DISCUSSION

VIMS is a prototype system for testing information organization and management features intended for improving Web information gathering tasks. A relatively small-scale user study was conducted to examine the feasibility of a larger research study and to highlight possible improvements to VIMS. In the evaluation approach, participants were meant to provide feedback about VIMS as recommended in the works of Scott (2009), and to suggest further possible improvements. Users were given two weeks to work on VIMS and return the evaluation forms. The longer evaluation period was intended to further ensure that the study participants had enough time to use and examine VIMS. The aim of the study was to identify usability, acceptability, and user satisfaction issues with VIMS.

The results of the study indicate that VIMS provided: a) an excellent level of user satisfaction; b) significant help to users with their tasks; c) effective information organization features; d) effective and useful presentation of task results; and e) effective preserving and re-finding features. Not only did users provide specific answers to questions related to evaluating each feature, but they also provided their own comments that showed the degree to which VIMS improved how Web information gathering is performed. The results of the study reflect how early adopters perceived VIMS.

The purpose of the prototype testing using only students as participants was to decide on the feasibility of a larger study to evaluate VIMS and to derive practical design recommendations. Further studies in this research will use experiments with larger populations. A larger complete factorial user experiment will be conducted to examine the organizational and management criteria identified in VIMS.

7 CONCLUSIONS

This paper presented VIMS, a prototype system for improving how users manage and organize Web information during gathering tasks. The current state of Web information gathering necessitates studying challenges users encounter during this type of task. VIMS was designed based on three previous studies (Alhenshiri et al., 2010a; Alhenshiri et al., 2010b; Alhenshiri et al., 2011). Our previous studies revealed several questions regarding which visualization, clustering, re-finding, and organizing factors would improve the process of Web information gathering. The initial evaluation of VIMS showed good indications of its usability and user satisfaction with VIMS as a tool for Web information organization.

REFERENCES

- Alhenshiri, A., Brooks, S., Shepherd, M., and Watters, C. 2010a. Augmenting the Visual Presentation of Web Search Results. In *proceedings of the 5th International Conference on Digital Information Management*, Thunder Bay, ON, Canada, 101-107.
- Alhenshiri, A., Shepherd, M., Watters, C., and Duffy, J. 2010b. Web Information Gathering Tasks: A Framework and Research Agenda. In *proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR2010)*, Valencia, Spain, 131-140.
- Alhenshiri, A., Shepherd, M., and Watters, C., 2011. User Behaviour during Web Search as a Part of Information Gathering. In *proceedings of the Hawaii International Conference on System Sciences (HICSS 2011)*, Hawaii, USA.
- Bonnell, 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.
- 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.
- Cutrell, E., Dumais, S. T., and Teevan, J. 2006. Searching to Eliminate Personal Information Management. In *Communications of the ACM (Special Issue: Personal Information Management)*, 49(1), 58-64.
- Elsweiler, D., and Ruthven, I. 2007. Towards Task-based Personal Information Management Evaluations. In *proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, The Netherlands, 23-30.
- 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 the 9th IFIP TC13 International Conference on Human-Computer Interaction*, Zurich, Switzerland.
- 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.
- Kules, B. and Capra, R.: What Do Exploratory Searchers Look at in a Faceted Search Interface? In: 9th ACM/IEEE-CS Joint Conference on Digital Libraries, Austin, TX, USA, 2009.
- 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.
- 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.
- Scott K. M. 2009. Is Usability Obsolete? *Interactions*, vol. 16, issue 3, 6-11.
- 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., 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.
- Turetken, O., and Sharda, R. (2005). Clustering-based Visual Interfaces for Presentation of Web Search Results: An Imperical Investigation. *Information Systems Frontier*, 7(3), 273-297.

