Image Search and Information Visualization Research at UW-Milwaukee's Multimedia Software Laboratory

James Dabrowski, Cheng Thao, and Ethan V. Munson
Multimedia Software Laboratory
Department of EECS
University of Wisconsin-Milwaukee

1 Introduction

This document summarizes the 2001 research activities in UW-Milwaukee's Multimedia Software Laboratory under a subcontract from the Center for Adaptive Data-Driven Information Processing. These activities are focused in two research areas, image search on the Web and information visualization, which are discussed in Sections 2 and 3, respectively. Section 4 describes some other activities supported by the subcontract and Section 5 describes planned activities for 2002.

2 Web Image Search

Traditionally, image search has been confined to image databases, which may hold thousands of images. While a database with thousands or even tens of thousands of images is quite large, it will only hold a small number As the Web is growing at an exponential rate, people are interested in ways of searching Web pages and other Web media. The nature of the Web image search problem is different than that of the tradition image search problem. Traditional image search searches images within a small local database of images where images may be captioned and images may have certain resolution and may be categorized. The Web images are not organized in any way, and images are found in several formats and with a variety resolution and quality. Tradition image search approaches are textual search and content search. Textual search relies on the caption, which is manually provided to support text search of images, and content search, which make use of the color, shape, and texture of the image. Some approaches make use of both caption and image content.

In previous work by Tsymbalenko [14, 15], we showed that several types of metadata in the HTML source of Web pages can be used to find images that match textual queries. Useful metadata clues included the text of the document's title element, the name of the file in which the image was stored, and the value of the image's ALT attribute. Median precision statistics (over 12 different queries) were 55.0% for the document title, 83.0% for the image file name and 87.5% for the ALT attribute. In another article [8], we argued that image search on the Web should always start with textual methods because users prefer making textual queries, because images are almost always accompanied by explanatory text on the Web, and because text-based methods avoid the costs associated with downloading and analyzing the images themselves.

Tsymbalenko's work had several significant limitations. First, queries were only single words, while many normal Web queries have two or three words. Second, the retrieved images were only rated for relevance by one person. Finally, the document analysis looked at only a subset of conceivable metadata clues.

In order to address these limitations, we have begun a replication and expansion of Tsymbalenko's research. This study will have several important changes:

- Relevance of images will be determined by the rating of three judges.
- The system will support multi-word queries.
- A larger number of queries will be tested and more pages will be downloaded per query.

A prototype image search and download tool to support this research has been implemented in Java. This tool supports multi-word queries and can communicate with multiple search engines. It sends the query to a chosen search engine and retrieves the maximum number of URLs that the search engine will permit. Then it downloads the pages referenced by the URLs and all the images that they contain, storing the Web pages and downloaded images on the file system. Information about the pages and the images is stored in a database using an XML format. In addition, the URLs in the Web pages that point to the images are patched so that they point to the local copies of the images. This URL patching ensures that relevance raters will be able to see each image in context if this is necessary in order to judge its relevance.

We have also implemented a separate metadata clue analysis tool that is used to analyze the downloaded documents and record statistics about which clues are present in each document. This clue analysis tool duplicates all of the eight clues used by Tsymbalenko and adds many others, including clues based on text formatting tags and on the layout semantics of tables. The table layout clues are of particular interest, because they do not appear to have been studied previously.

Finally, development has begun on a prototype Web-based relevance rating tool that can automatically collect ratings from multiple raters. Once complete, this should greatly simplify the collection of ratings for the large number of images that we expect to download for this study.

Once the relevance rating process is complete, statistical techniques such as discriminant analysis or logistic regression will be used to identify formula for estimating relevance automatically.

3 Information Visualization

On part of the statement of work for CADIP research at UW-Milwaukee called for research on the usability of the Shaw and Ebert's Stereoscopic Field Analyzer (SFA) [5]. SFA was designed for experiments with novel scientific visualization interfaces. It could be controlled by an unusual two-handed control interface that used magnetic tracking devices and used various different graphical techniques for efficient and effective display large numbers of data points. While SFA looked promising, its dependence on a very specific hardware/software combination made it unstable. Ultimately, the project was abandoned when it became apparent that getting SFA to run stably enough to conduct actual usability studies with inexperienced users would be impossible. However, many of the research issues that would have been addressed by the SFA project are important, so we have begun to explore the field of information visualization in more detail.

Our work on information visualization is just in its formative stages. Thus, the remainder of this section presents a survey of key ideas in the area and suggests some avenues for further research.

Information visualization is a young domain that began to be seen as a separate field of study shortly after 1990, evolving from scientific visualization. Scientific visualization studies how advanced graphics technology can facilitate scientific investigations and makes extensive use of three-dimensional interaction techniques because so much scientific data is spatially-oriented. In contrast, information visualization tends to focus on data that lacks an obvious spatial analogy, such as the results of information retrieval queries or of data mining analyses for the business domain.

For its first ten years, numerous information visualization projects were completed each showing yet another way of displaying complex data in ways that could make that information more useful to users. Recently, researchers have begun to categorize these approaches in an effort to bring further understanding to the field [2].

One information visualization technique is to provide the user with tools for animating an information display by controlling a parameter that modifies which portion of the data set is shown. Ahlberg and Shneiderman [1, 12] developed the concept of *dynamic queries*. Their prototype visualization helped its users understand real estate listings in the Washington DC area. Rather than presenting just a textual listing of the homes, Ahlberg's system displayed a map of the area on which available homes were displayed as dots. Users could then use interactive sliders to filter out homes based on criteria including price, number of bedrooms and location. Specific information on a particular home was available by clicking on the home's dot. This approach allows users to interactively animate the display in order to take advantage of the visual system's sensitivity to changes. It also makes it easy to reduce the amount of information on a busy display down to a level more easily coped with. The dynamic query technique has also been applied to U. S. census [4], films [1], and personal histories [9].

A second category of information visualization tools, called "Focus + Content," focus on providing the user with a detailed view of the data at hand plus additional information about the context in which that data occurs. The best known example of this class of interfaces is the Fisheye view developed by Furnas [6]. Other examples include the TableLens [10] and the Hyperbolic Browser [7]. In all of these approaches, the material of immediate interest to the user is centered in the screen and takes up most of the space. Other related information is displayed in the periphery of the screen, with the most closely related information being placed closer to the center. Three-dimensional graphics techniques are used to distort the display of the periphery so that the space occupied by a unit of information becomes progressively smaller as a unit's content becomes less relevant to the primary material. The distortion of the periphery typically makes that less relevant information appear to be receding away from the user.

Another general category of visualizations are those that display textual information. For the most part, research in this area has focused on ways of displaying large corpuses of documents so that users can see relationships between the documents. One example of this is the Galaxy of News [11] in which hundreds of thousands of documents are displayed in a three-dimensional space where documents which are related to one another (the news topic is the same or similar) are clustered together on the screen. The purpose here is to allow the user to gain an understanding of the size of the document space and filter out irrelevant documents or zoom in on documents of particular interest.

The final area of focus in the field of information visualization is the creation of virtual workspaces. In this area, the computer screen is set up to simulate an entire environment in which the user works with objects. The workspace and the objects act as metaphors for the task that the user is trying to accomplish — for example, the saving and browsing of groups of related web-pages. A popular example of this approach is Xerox's Web-Forager project [3].

While this categorization of different types of information visualization techniques is useful, the field still lacks a clear conceptualization of itself and to where it will proceed in the future. All of the research described above presented *spot solutions* [13] to information visualization problems. They are specific applications designed to display one specific type of data for one specific purpose. Thus, it is clear that the field still lacks general principles of information visualization.

To date, the best advice for displaying information to the user has been given by Edward Tufte, who says simply that the display should fit the cognitive task [16]. This advice is very basic and is a good starting point for thinking about information design, but it is not sufficiently specific to be used reliably by naive designers.. It does not answer any of these questions:

- Are there certain ways of displaying certain types of data that will aid the user in completing specific cognitive tasks?
- Are there categories of cognitive tasks, all of which can be equally well supported by single type of display, such as a scatter-plot?
- Are there other cognitive tasks for which a "focus + context" view of the data is more appropriate?

These are the questions that we believe need to be answered by information visualization research.

We believe that we can begin to develop a taxonomy of cognitive tasks and then identify specific visualizations that best support each type of tasks. We are also interested in extending our previous work on style sheets into the information visualization domain. Given a taxonomy of tasks and of best-practice visualizations for task categories, it should be possible to provide users with style sheets that embody each of those best practices. Then users will be able to view data in powerfully informative ways without having to choose the details of the visualizations that they use.

4 Other activities

In addition to the research activities described in this report, Dr. Munson has been active in establishing a document engineering research community. He has chaired the steering committee for the new ACM Symposium on Document Engineering

(DocEng 2001), which will meet for the first time in Atlanta on November 9 and 10, 2001 in conjunction with the ACM CIKM 2001 conference. DocEng 2001 received over fifty submissions, of which eighteen were accepted. Unfortunately, the events of September 11 caused several authors to cancel their travel plans. Even in the face of these problems, the Symposium can be viewed as a success for stimulating document engineering research.

5 Research plans for 2002

For the Web image search project, our plan is to construct a Web-based system for rating the relevance of images to a query. We will have three raters rate images collected using multi-word queries. The queries will be drawn from a number of query types including people, places, and phenomena. Some people and place names will be well-known (meaning the names of figures in the celebrities or celebrities or famous places) or less well known. Once we have human relevance ratings to act as a gold standard of relevance, we will use statistical techniques to develop formulas for computer-generated relevance ratings.

For the information visualization project, we plan to proceed on two paths. We plan to develop visualization techniques for the image search data that we collect. These will be spot solutions for this problem, but they should give us some insight into the process of designing visualizations. Separately, we will investigate the suitability of style sheet technology for the visualization task.

References

[1] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Conference Companion of CHI '94: ACM Conference on Human Factors in Computing Systems*, page 433, 1994.

[2] S. Card, J. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, California, 1999.

[3] S. K. Card, G. G. Robertson, and W. York. The Web-Book and the Web Forager: An Information Workspace for the World Wide Web. In *Proceedings of CHI '96: ACM Conference on Human Factors in Computing Systems*, pages 111 – 118, New York, 1996. ACM Press.

[4] G. Dang, C. North, and B. Shneiderman. Dynamic queries and brushing on choropleth maps. In *Proceedings of IEEE International Information Visualization Conference 2001*, 2001.

[5] D. Ebert, C. Shaw, A. Zwa, E. Miller, and D. A. Roberts. Two-handed interactive stereoscopic visualization. In *IEEE Visualization '96 Conference*, 1996.

[6] G. W. Furnas. The fisheye view: A new look at structured files. Technical report, AT&T Bell Laboratories, Murray Hill, NJ, 1981.

[7] J. Lamping and R. Rao. The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing*, 7(1):33–55, 1996.

[8] Ethan V. Munson and Yelena Tsymbalenko. To search for images on the web, look at the text, then look at the images. In *Proceedings of the First International Workshop on Web Document Analysis*, pages 39–42, 2001.

[9] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. In *Proceedings of CHI '96: ACM Conference on Human Factors in Computing Science*, pages 518–525, New York, 1996. ACM Press.

[10] R. Rao and S. Card. The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proceedings of CHI '94: ACM Conference on Human Factors in Computing Systems*, pages 318–322 and 481–482, New York, 1994. ACM Press.

[11] E. Rennison. Galaxy of news: An approach to visualizing and understanding expansive news landscapes. In *Proceedings of UIST '94: ACM Symposium on User Interface Software and Technology*, pages 3 – 12, New York, 1994. ACM Press.

[12] B. Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, 1994.

[13] R. Spence. *Information Visualization*. Addison-Wesley, New York, 2001.

[14] Yelena Tsymbalenko. Using HTML metadata to find relevant images on the Web. Master's thesis, University of Wisconsin-Milwaukee, May 2000.

[15] Yelena Tsymbalenko and Ethan V. Munson. Using HTML metadata to find relevant images on the Web. In *Proceedings of Internet Computing 2001, Volume II, Las Vegas*, pages 842–848. CSREA Press, June 2001.

[16] E. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.