Employing Natural Language Summarization and Automated Layout for Effective Presentation and Navigation of Information Retrieval Results

Simon Lok
CGUI Lab, Columbia University
New York, NY 10027
lok@cs.columbia.edu

Min-Yen Kan
NLP Group, Columbia University
New York, NY 10027
min@cs.columbia.edu

ABSTRACT
In this paper, we propose the employment of a combination of multi-document summarization[2] and automated layout[3] as a post-processing step in document retrieval. We examine the use of an interactive textual fisheye that employs automated layout techniques to present a generated natural language summary as a replacement to the standard ranked list. Our presentation system is novel because it employs a natural language summarization system to generate informative as well as indicative summaries of varying lengths to create an effective layout given the constraints of a specific amount of available screen space. In addition, the system leverages the topical structure of the generated summary to create a textual fisheye for navigating and interacting with the presentation where the actual length of the text is varied to provide different levels of detail rather than the more conventional change in rendering parameters such as font size.

Keywords
Automated Layout, Natural Language Generation

1. INTRODUCTION
The ranked-list is the de-facto standard method of presenting the results of an information retrieval (IR) system. All of the major World-Wide-Web search engines (e.g. Google, AltaVista, Lycos) as well as most IR systems found in standalone software follow this paradigm. This technique has become so pervasive that it can be compared with the ubiquity of the Window Icon Menu Pointer (WIMP) user interface paradigm in that an entire generation of up-and-coming computer users will have experienced nothing else. Although the display of a ranked list is the most obvious way to present the results of an information retrieval query (after all, the result is almost by definition going to be document-rank-tuples), this format seldom directly fulfills the need of the user.

In this paper, we discuss the use of a combination of natural language summarization and automated layout techniques in a system that attempts to provide a user with an efficient presentation of the actual information the user desires rather than simply displaying a list of documents that may contain the information. To accomplish this, we feed the textual summaries along with meta-data describing the text generated by a combination of informative and indicative natural language summarization to an automated layout engine. In addition, we provide the user with an interactive user interface the enables the user to quickly navigate and focus on particular aspects of the retrieved information that leverages the topical nature of the natural summaries generated by the natural language system.

2. IMPLEMENTATION
We created Centrifuser, a summarization system to meet the needs of browsers and searchers in highly structured domains. In a nutshell, Centrifuser relies on an infrastructure of document topics. First, documents are converted into topic trees and tree structure-based similarity calculations are used to find which topics are similar and which are different. The browser summarization module takes the similar topics and performs sentence extraction of salient sentences for the synopsis. It additionally utilizes relationships between the topics to generate navigation links to related topics. The searcher module uses text generation techniques to create text that describes high-level differences between documents.

Equipped with document topic trees for each of the documents in the result set, a composite topic tree for the text type, and the query mapped to topic nodes in the document and composite trees, Centrifuser produces summaries suited for browsing and searching. Each topic within the summary is variable length as it is the result of merging together various pieces of the source documents. This variable length property is exploited by the AIL (Automated Interface Layout) system to not only automatically generate a presentation to fit on a given amount of real-estate but also support an interactive informative summary topic browsing system.

The AIL system was created to automatically generate effective presentations of the topic-tree based summaries of Centrifuser and is part of the PERSIVAL[4] Digital Libraries Initiative Phase 2 project at Columbia University. It operates using many of the well-established principles of automated layout systems that have come before it (e.g. spatial constraints to enforce design principles like margins) in addition to the novel technique of leveraging the ability of Centrifuser to generate output of different lengths. AIL provides fish-eye browsing of the presentation where the level of detail is controlled by changing the actual length of the text instead rather than modifying rendering parameters like font size.

The topic-based informative summaries of Centrifuser are the underpinning of the layout generated by AIL. The informative summaries formed by merging different parts of the source documents that have been identified to be about the same topic are rendered as paragraphs on the display. These informative summaries are accompanied by indicative summaries that consist of references to numerous source documents in addition to a block of text describing the differences between them. In addition, Centrifuser generates a table of semantic links between each informative summary and

one or more indicative summaries that contains references to the most relevant articles to the topic of the informative summary.

Each informative summary can be generated in a variety of lengths. Initially, Centrifuser will provide the topic that is the closest match to the user query in a longer length than the other topics in the informative summaries. This summary is allocated the most screen real estate and placed first. In addition, AIL will place the indicative summary that is most related to the highest ranked informative summary directly underneath. The indicative summary allows the user to drill down to the actual source documents that were used to generate the informative summary. We use the RemoteJFC system developed at Columbia University to allow AIL to take control of a window on any machine with network connectivity for displaying the source articles. This allows the source article to be displayed on the same machine that is running AIL as shown in Figure 4, or on a separate machine such as a tablet computer or PDA that the user is holding in their hand. This would allow the user to use an AIL system running on a shared system with a large wall size display that might be in a fixed position while keeping the results of their search on a handheld machine that they can carry with them.

Topics that are closely related to the initial topic are provided by Centrifuser in shorter lengths and placed next on the screen after the topic of highest rank. Indicative summaries related to these summaries are currently ignored by AIL. Topics that are even less related are provided in even shorter lengths and placed last. In many cases, the amount of screen real estate will not be sufficient to display all of this text. For example, there may simply not be enough pixels on the screen if the user is viewing the presentation on a handheld or tablet whereas there might be enough if the user is sitting at a workstation with a large high-resolution display. In this case, AIL will ask Centrifuser for shorter representations of the text and display those if possible. If the text will still not fit, AIL will choose to leave out the topics that are low in rank.

The initial layout presented to the user will have the informative summary that most closely matches the query being displayed in great detail with topics that are not as related being display in less detail. The user can tell AIL to change the focus by clicking on topics other than the closest match. This will cause AIL to ask Centrifuser to change the topic of closest match to the one that the user picked. The entire layout process that has been described above will then be executed and a new presentation with the new focus will be created. However, rather than simply replacing the current presentation, AIL animates the screen and changes smoothly from one layout to the next in attempt to help the user keep visual context on what they are looking at. The result is a fish-eye[1] navigation system for the results of an information retrieval system. However, unlike traditional fish-eye systems that change the rendering parameters like font size to change the level of detail, AIL changes the actual length of the text being displayed by leveraging Centrifuser.

3. ACKNOWLEDGMENTS

4. REFERENCES

1. G. Furnas. "Generalized Fisheye Views." *Proc. CHI '86,* ACM Press, New York, 16–23.

Figure 1: A series of screenshots showing AIL animating from one layout to the next

2. M.Y. Kan, K.R. McKeown and J.L. Klavans. "Applying natural language generation to indicative summarization." *Proc. of 8th European Workshop on Natural Language Generation*, Toulose, France, 2001.

3. S. Lok and S. Feiner. "A Survey of Automated Layout Techniques for Information Presentations." *Proc. of the 1st. Int. Symp. on Smart Graphics*, Hawthorne, NY, 2001, 61–68.

4. K. McKeown, S.-F. Chang, J. Cimino, S. Feiner, C. Friedman, L. Gravano, V. Hatzivassiloglou, S. Johnson, D. Jordan, J. Klavans, A. Kushniruk, V. Patel, and S. Teufel. "PERSIVAL: A System for Personalized Search and Summarization over Multimedia Healthcare Information." 1st. *ACM/IEEE Joint Conf. On Digital Libraries*, Roanoke, VA, 2001, 331–340.