Query Fingerprinting: Visual Representation of Text-Based Information

Philip R. Johnson
Indiana University - Purdue University Fort Wayne

Follow this and additional works at: http://opus.ipfw.edu/stu_symp2002
Part of the Computer Engineering Commons

Recommended Citation

This Presentation is brought to you for free and open access by the IPFW Student Research and Creative Endeavor Symposium at Opus: Research & Creativity at IPFW. It has been accepted for inclusion in 2002 IPFW Student Research and Creative Endeavor Symposium by an authorized administrator of Opus: Research & Creativity at IPFW. For more information, please contact admin@lib.ipfw.edu.
QUERY FINGERPRINTING: VISUAL REPRESENTATION OF TEXT-BASED INFORMATION
Philip R. Johnson
Sponsor: Beomjin Kim
Department of Computer Science
Indiana University Purdue University Fort Wayne

Recent research estimates that, worldwide, about one exabyte of data is generated annually. When the amount of information continues to grow at this rate, more data will be generated in the next three years than during all previous human history. Finding relevant information from such a vast workspace is a difficult and time-consuming process. To assist users in accessing these volumes of data effectively, various information retrieval systems have been developed that provide rich interaction and abstract views of the information.

This research has developed an information visualization methodology called Query Fingerprinting (QF) in which text-based information is transformed into an iconic visual representation. To reduce the complexity of the document, a series of preprocessing tasks, such as paragraph segmentation, stop-word filtering, and suffix stemming are performed. Processed documents are then transformed into visual abstractions using three attributes: color, intensity, and size. This visualization maps given queries to primary colors, and the concentration of each query in a paragraph defines the intensity of the corresponding query’s color. The mixture of queries’ colors is projected onto a sequence of strips that simultaneously shows paragraph lengths, query distribution, and query relationships. This compact, but informative, visual abstraction allows users to compare multiple documents intuitively and swiftly while alleviating the need to reference the underlying text. The QF method also provides interactive zoom-in functionality for further navigation to support users in their decision processes, and suggests other significant words to refine the search scope.

An experimental study has been conducted on the Congressional text archive. The pioneer study showed that users could find related information significantly faster using the QF system than using the summary-based system, while maintaining the level of accuracy. A post-experiment survey illustrated that the QF system was fairly intuitive to use and useful in assisting them make their decision when searching for information. This on-going research is under expansion to be functional as an Internet search engine, which will eventually save users time and decrease network traffic by reducing the number of references to unrelated sites.