Workshop Report: Information Doors — Where Information Search and Hypertext Link

Einat Amitay
Information and Communication Sciences, Macquarie University
CSIRO, Mathematical and Information Sciences, NSW, Australia
einat@ics.mq.edu.au

The Information Doors workshop was held in conjunction with ACM Hypertext and Digital Libraries conferences, on May 30th 2000. The theme of the workshop was textual presentation of search results.

Online search results are, no doubt, a form of hypertext created on-the-fly. Search results pages are also probably the most frequently seen hypertext form of writing nowadays. However, the research community tends to identify the presentation search results with Information Retrieval research. This workshop considered search results as a form of hypertext, entertaining presentations and discussions about the nature of this dynamically created textual point-of-departure.

The workshop was a great success in terms of attendance and participation. For many, it has been an opportunity to raise issues that are usually marginal to core Information Retrieval research, and even to main stream HCI (Human-Computer Interaction).

The idea for the workshop came from a literature survey about search results that was written for a research proposal. It appeared that text for itself was not considered a worthwhile research focus. Graphical visualisation seemed to be a preferred solution by most research systems, and when the results were textual, it was evident that researchers did not think them to be worthwhile reporting. Text was always the simplest default and was used to compare against. However, since the beginning of search visualisation research, most of the comparisons between graphical and textual displays of online search and browse interfaces actually show that textual display of search results is not inferior to graphics, and sometimes even better.

At the workshop there were four presentation themes. The first two presentations were a broad overview of the field, one from the academic point of view, and the other from the industrial point of view. John Cugini from the Information Technology Laboratory at National Institute of Standards and Technology (NIST), presented a number of NIST projects that addressed the generation, evaluation, and presentation of search results. His paper characterised the presentation problem, and outlined of the components of an evaluation system. John Cugini's involvement in the search visualisation community allowed us a glimpse into the important research that is being carried out in this field.

Sally Kleinfeldt and Jaideep Baphna, from Dataware Technologies Inc, presented a commercial perspective on textual search results. Sally Kleinfeldt gave a detailed overview of the industrial effort to enhance users' search experience. She gave examples such as Relevance Feedback, Concept Mining, best passages display, field-based summarisation, trend analysis, results clustering by category, results clustering by content, etc. She reported that feedback from customers clearly indicates that the search experience could be improved for most applica-

Full versions of the papers and contact details can be downloaded online from:
http://www.ics.mq.edu.au/~einat/info_doors/

tions by using commercially available techniques that are based on purely textual information. She also suggested that the best approach is to combine the following elements: Use summarisation techniques, such as best passage, that give a better indication of a document's relevance. Organise the results page according to some appropriate criterion that clusters categories of related results. She concluded by hinting that using concept mining to help users improve their queries is a promising technique in need of further exploration.

The next couple of talks examined different textual displays. Robert McArthur and Peter Bruza from the Distributed Systems Technology Centre at University of Queensland (DSTC), presented a technique and a tool for query refinement, suggesting that it would be easier and more productive for users to process the refinements than to peruse a list of document summaries. They reported that previous research has shown that the use of interactive query formulation does promote effective retrieval. Their research addressed the problem of how to present query refinements to facilitate effective perusal by the user, so that the most relevant refinements are located quickly and easily. Their system uses a method of breaking down the mess of refinements for the user by binning the phrases. They used six bins on the Web for over 12 months, and have anecdotal evidence that some users appreciate and use these bins to choose where to look for useful refinements. Their system also ranks the refinements, so that only the 'best' refinements are shown. They developed a ranking formula based on the multiplication of the frequency count of the refinement and the inverse document frequency of the non-query terms.

Offer Drori, from SHAAM Information Systems (Israeli Ministry of Finance) & The Hebrew University of Jerusalem, presented experimental evaluations of different methods for displaying a list of textual search results. He compared between users' interaction with titles, keywords, key phrases, highlighted terms, etc. His research also demonstrated the difficulty in comparing non-native speakers of English to native speakers of English, when measuring interaction with a by-default English environment. This presentation raised a discussion about the validity of online experiments vs. off-line ones, an issue that is currently debated by many researchers.

The pair of presentations that followed addressed the more general issue of arranging the information space in such a way that the relations between the different entities would create a coherent display. Douglas Grundman and Andrea Michalek, from Infonautics Corporation, suggested displaying generated search results information in a manner that is easily assimilated by users. Their work assumes that structuring the result list is most powerfully and naturally achieved by pre-structuring the information being presented. Their system, a series of search engines and crawlers, one of which is `www.company.sleuth.com`, uses the fact that in the same domain, information is arranged in similar ways and when this structure is already known, it becomes much easier to structure its presentation automatically.

Matt Carmack from Novell, Inc, and Deryle Lonsdale, from BYU Linguistics Department, proposed a framework for integrating a principled theory of information structure with traditional web-based search presentation techniques. They reviewed research that has been done in the technical writing industry on on-line information structure, highlighting some of the difficulties and challenges faced by developers and users of hypertext-based documentation. Then they introduced Peircean categories as a natural classification system for on-line information, showing how the Peircean categories explain the otherwise confusing intricacies of technical information phenomena.

The last presentation, by Anne Mahoney from the Perseus Project, Tufts University, was a very intriguing outline of a

system that uses textual search results embedded in an educational text to enhance the users' ability to acquire knowledge. The Perseus system introduces new links and text into viewed pages, tailoring implicit search results to cohere with the content and context of the displayed document. This presentation demonstrated the use of textual display when applied to existing database.

The discussions during and after presentations indicated that there are many questions left for debate and study, and that there is room for many more projects and applications. The lively discussions, pleasant atmosphere, and attendance from internationally mixed backgrounds provided an enjoyable and informing experience for all.❖

I would like to thank the program committee: Chaomei Chen (IS & Computing, Brunel University), Mary Czerwinski (Microsoft), Andrew Dillon (SLIS, Indiana University), Sue Dumais (Microsoft), Raya Fidel (SLIS, University of Washington), Gene Golovchinsky (FXPAL), Stephen Green (Sun Microsystems), Christina Haas (English, Kent State University), Johndan Johnson-Eilola (English, Purdue University), Chris Manning (CS & Linguistics, Stanford University), and Vibhu Mittal (Just Research) for their comments, moral support and encouragement while organising this workshop. And thanks also to Peter Nürnberg & David Lowe for their organisational effort and helpfulness. ❖

Workshop Report: 4th Hypertext Writers' Workshop The Messenger Morphs the Media

Deena Larsen
textra@chisp.net

The Hypertext writers workshop brought hypertext content and system developers together to explore hypertext writing strategies and discuss specific tools and techniques to create quality hypertexts.

The workshop broke into four tracks:

1. Write Now: A hypertext writing exercise to explore ways to write hypertext and demonstrate the concepts and thought techniques involved

2. Write With What?: A discussion designed to further expand existing lists of specific tool requirements—and determine where these are available or how we can get them

3. Write How?: A discussion designed to further expand our existing lists of strategies and techniques—and determine how these techniques are being used and could be used in future hypertext literature.

4. **Talk about Writing**: A discussion designed to explore the role of literary criticism in hypertext and determine how to adapt the tools of literary criticism to analyze and discuss hypertext literature.

After these sessions, participants met to define recurring issues and challenges and to plan on what we can do as a community to promote hypertext and hyperliteracy. The following items and ideas were mentioned in our discussions. However, this paper by no means suggests that all the participants held these ideas or agreed with these positions.

Recurring Issues

Defining Hypertexts

What a hypertext is has never been clearly laid out, and has changed drastically over the past few years to include blended media, imagery, sound, programming elements, structure, and

This Report can also be found at:http://www.wordcircuits.com/htww/ht00aft.html