Designing Interactive IR combining Adaptation to User Tasks and Strategies with Non-topic Document Analysis

by

Preben Hansen SICS/CRIT preben@sics.se

Jussi Karlgren SICS/HUMLE jussi@sics.se

Swedish Institute of Computer Science
Box 1263, S-164 28 KISTA, Stockholm, Sweden

GENERAL SOLUTIONS ARE NOT SUFFICIENT

Research in information retrieval and document analysis has traditionally concentrated on building general, task-independent representations about the content of documents based on word occurrence statistics of different kinds: performing a sort of shallow semantic analysis, in effect. The primary research goal has been to create an optimal representation for the general case.

Throughout the history of information retrieval, however, the research community has been aware of the fact that the interaction of information seeking users and the tools to access information sources is important in itself. Information can be sought for various reasons and with various ideas of how to determine what documents are relevant.
(Belkin, 1993b; Luhn, 1957).

This research plan outlines a framework within which
a) to find more knowledge from texts and their users than a shallow approximation of text topic. Texts have, besides content, STYLE and ECOLOGY, both which can be automatically identified from a text base and its usage statistics and used for text categorization and
b) users have information seeking STRATEGIES that can be recognized through user studies and supported through interface design.
Finding ways to describe and evaluate the problems of search behavior and navigation through a hypermedia/hypertext system are important.

USERS DIFFER as do TASKS and STRATEGIES

Our starting point is that users are different; they have different backgrounds; their tasks vary; their goals and reasons for using the system are various and difficult, if not impossible, to anticipate completely in advance. Since the introduction of the World-Wide Web, the easy and user-driven accessibility to information through hypertext systems and hypertext documents has increased the number of users, and we are getting to the point where the sophistication and accessibility of document bases and the diversity of users is making the heterogeneity of users very clear to information providers. (Belkin, 1993b;Fox et al, 1993).

TEXTS ARE MORE THAN TOPIC

Word occurrence statistics can only be taken to a certain limit - the assumption that words are representations of concepts and that these concepts are the only factors that make a document relevant or interesting is a gross simplification. Systems deployed today asymptote at similar performance levels; it is clear that the simple and theoretically clean framework limits the amount of information that can be extracted in real life tasks.

Text genre and stylistic variation is easily recognizable, down to the level of individual variation between authors within the same genre using computationally non-complex methods. Users of document bases may well be aware of which genres they search for: popular science, overviews, technical descriptions, program manuals, long texts, short texts. (Biber, 1989; Walker, 1991; Belkin, 1993a; Belkin, 1994; Karlgren and Cutting, 1994).

The ecology of texts can be as important to users as the content and genre. Who reads a document is an important factor in deciding which to choose among several candidates. Social filtering, the technique of using text ecology or text usage as a discriminant, is easily applied to areas as diverse as Usenet News texts, Usenet News conferences, music albums, video movies, and novels. (Karlgren, 1990; Karlgren 1994; Resnick et al, 1994; Hill et al, 1995; Shardanand and Maes, 1995).

PROJECT OUTLINE

In conclusion, we find that beyond the technical design challenges of Digital Libraries and other information retrieval systems, there is a need to address other aspects e.g. non-topical text analysis; information seeking strategies; user interface design; and user tasks and navigation.

For SICS' retrieval system for technical and research reports, we will follow a user-centered design methodology, interviewing current users and perusing usage logs, and building an interface to a document database. The method will be integrated to provide a model of end-user searching and navigation in a hypertext system.

Users can be categorized by different background and domain e.g. industrial partners, or researchers from academic sites; user activities fall broadly into high level strategies, such as searching for specific information, with a specifically framed information need or browsing for general information, with a less explicitly framed request. Assumptions of user goals and preferences can be made from background information, provided users are informed, and are given control over the assumptions made.

We will build information retrieval tools which will support high level strategies in different ways. We will use techniques from previous SICS projects on adaptive hypermedia where an information system adapts to its perception of user task and background, and displays information of different type and quantity accordingly. In this case, the perceived background and task of the user will not change the information itself as in the case of adaptive hypermedia, but primarily the tool setup and default tool parameter settings offered to the user. Tools for the user will not only include standard tools for content search, but also tools for genre identification and social filtering.

REFERENCES

Nicholas J. Belkin. 1993a. Interaction with Texts: Information Retrieval as Information-Seeking Behavior. Proceedings of IR'93, the first meeting of the IR special interest group of the GI, Ravensburg.

Nicholas J. Belkin. 1993b. BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval. Information Processing & Management Vol. 29, No 3, pp 325-344, 1993.

Nicholas J. Belkin. 1994. "Design Principles for Electronic Textual Resources: Investigating Users and Uses of Scholarly Information". In Antonio Zampolli, Nicoletta Calzolari, and Martha Palmer (eds.) Current Issues in Computational Linguistics: In Honour of Don Walker, Dordrecht: Kluwer.

Douglas Biber. 1989. A typology of English texts, Linguistics, 27:3-43.

Fox, E. A., D. Hix, L. Nowell, D. J. Brueni, W. Wake and L. Heath. 1993. Users, User Interface, and Objects: Envision, a Digital Library. JASIS Vol. 44 (8), 480-491, 1993.

Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. 1995. Recommending and Evaluating Choices in a Virtual Community of Use. SIGCHI'95 (Denver Colorado, May 7-11, 1995) Human Factors in Computing System Proceedings 1995. New York: ACM SIGCHI, pp. 194-201.

Jussi Karlgren. 1990. An Algebra for Recommendations. SyslabWorking Paper 179, Department of Computer and System Sciences, Stockholm: Stockholm University.

Jussi Karlgren. 1994. Newsgroup Clustering Based On User Behavior - A Recommendation Algebra, SICS Technical Report T94:04, Swedish Institute for Computer Science, Stockholm.

Jussi Karlgren and Douglass Cutting. 1994. Recognizing Text Genres with Simple Metrics Using Discriminant Analysis, Proceedings of COLING 94, Kyoto. (In the Computation and Language E-Print Archive:cmp-lg/9410008).

Hans Peter Luhn. 1957. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. IBM Journal of Research and Development1 (4) 309-317. (Reprinted in H.P. Luhn: Pioneer of Information Science, selected works. Claire K. Schultz (ed.) 1968. New York: Sparta.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergström, John Riedl. 1994. Group Lens: An Open Architechture for Collaborative Filtering of Netnews, Procs. CSCW 94, Chapel Hill.

Upendra Shardanand, Patti Maes. 1995. Social Information Filtering: Algorithms for Automating "Word of Mouth" SIGCHI '95 (Denver Colorado, May 7-11, 1995) Human Factors in Computing System Proceedings 1995. NewYork: ACM SIGCHI, pp. 210-217.

Donald E. Walker. 1991. The Ecology of Language. Proceedings of the International Workshop On Electronic Dictionaries. Japan ElectronicDictionary Research Institute. Also in Antonio Zampolli, Nicoletta Calzolari, and MarthaPalmer (eds.). 1994. Current Issues in Computational Linguistics: In Honourof Don Walker, Dordrecht: Kluwer.