Structured Text Representations and Relevance Judgments: An Exploratory Study

Padmini Srinivasan

School of Library & Information Science
University of Iowa
Iowa City, Iowa, USA -52242
padmini-srinivasan@uiowa.edu

This paper presents results of an exploratory study on using a structured text and query model for analyzing relevance judgments made by users on retrieved texts [1] *The structured model moves beyond keyword based text representation and retrieval. In particular it allows texts and queries to be represented by concepts embedded within the structural framework of complex objects. These objects preserves important text-based links between the selected concepts. The current study reveals a tendency for user judged relevance potential to decrease as the structure-based similarity between query and text decreases. This indicates that the structure-based retrieval model has the potential to conduct more precise retrieval than keyword based models. This supports our overall research goal of developing and investigating the structured retrieval model.*

INTRODUCTION

The study is part of ongoing research on structure-based text retrieval wherein both texts and queries may be represented using complex objects. In the 1993 SCAMC session we presented the details of this model and its motivation [1] We have also discussed associated retrieval strategies and a prototype implementation in other work [4]. Here we focus on the ability of the model to explain (and therefore possibly predict) relevance judgments made by users.

METHODOLOGY & CONCLUSIONS

The analysis was conducted on a subset of the dataset used by Hersh et al. [3]. For each of 12 queries a maximum of 20 relevant and 20 non relevant texts were analyzed resulting in a total of 174 analyzed texts. In particular we focused on those queries where the majority of non relevant articles had the query concepts but not in central roles.

Four distinct steps were followed: 1) Extract structured representations from queries. 2) Extract structured representations from retrieved abstracts. 3) Compare query and text structures and describe the relationship between the two. 4) Compare query-text relationships (obtained in the previous step) to relevance judgments made by users.

The data analysis indicates that there is a tendency for relevance to decrease as the relationship between the query and text weakens. The analysis of these query-text relationships involves conceptual comparisons within a structured framework. To conclude, the tentative conclusions made here provides further support for continued development and investigation of the underlying structure-based text retrieval model. Our next step is to perform similar analysis on larger sets of queries and texts.

REFERENCES

[1] Hankom, L. L. and Srinivasan, P. Quality Retrieval of the Empirical Literature: A Structured Approach. *Proceedings of the SCAMC,* 1994.

[2] Haynes, R., McKibbon, K., Walker, C., Ryan, N., Fitzgerald, D., Ramsey, M. Online Access to MEDLINE in Clinical Settings. *Ann Intern. Med.* (1990), 112: pp. 78-84.

[3] Hersh, William R. Hickam, D. H., Haynes, B., McKibbon, K. A. A Performance and Failure Analysis of SAPHIRE with a MEDLINE Test Collection. *JAMIA* 1, 1 (1994), pp. 51-60.

[4] Rama, D. V. and Srinivasan, P. An Investigation of Content Representation using Text Grammars. *ACM Transactions on Information Systems*, January 1993

[1]The author is very grateful to William R. Hersh, MD for sharing the MEDLINE dataset which originated in an earlier study by Haynes & colleagues [2] and also for sharing the error analysis data generated in [3].