Theoretical evaluation of IR models using symbolic means

P.D. Bruza and D. Song
Distributed Systems Technology Centre
University of Queensland, 4072 Australia
{bruza,dsong}@dstc.edu.au

Experimental approaches are widely employed to benchmark the performance of an information retrieval (IR) system. Measurements in terms of recall and precision are computed as performance indicators. Although they are good at assessing the retrieval effectiveness of an IR system, they fail to explore deeper, theoretical aspects of the associated models. The emergence of logic-based IR in the late eighties has spawned a body of work in which IR models are subjected to theoretical scrutiny (see [Bruza & Huibers 1994, Huibers 1996, Hunter 1996, Bruza et al 2000, Song 2000, Wong et al 2001]). These works have focussed on providing a symbolic, rather than numeric characterization of the matching function via the use of rules based around the "aboutness" relation. Several sets of rules characterizing the aboutness relation have appeared in the literature. Even though there is no agreement, a given set of rules provides a context in which IR models can be compared via representation theorems. In a nutshell, IR models can be compared on the basis of which aboutness properties they support. Alternatively, a single IR model can be analyzed by the aboutness properties it supports, thus gaining some potential insight into how it may behave in practice. This form of analysis has been termed *inductive evaluation* [Bruza & Huibers 1994] and more recently *functional benchmarking* [Song 2000, Wong et al 2001].

When it was initially proposed, inductive evaluation was promoted as a means of circumventing the vagaries of experimental IR [Bruza & Huibers 1994], but has theoretical evaluation of IR models using symbolic means lived up to this promise? This article will essentially provide an overview of symbolic-based theoretical evaluation and analysis, present the findings and detail the lessons that have been learnt. More specifically, this article

- will present a short history of aboutness research through the last three decades, particularly focussing on the research that has been inspired by logic-based IR.
- will provide examples of rules characterizing the aboutness aboutness relation symbolically and relate them to similar rules found in nonmonotonic logic. A major topic of discussion will be monotonicity of aboutness- this describes symbolically how aboutness behaves under information composition. It will be argued that IR is a fundamentally conservatively monotonic with respect to aboutness. The relation between the rules and recall/precision and document length will be discussed.
- will define a functional benchmarking suite based on our latest investigation in aboutness rules and provide examples of applying it to the inductive evaluation of well-known IR models: Boolean, vector-space and probabilistic models. These models will be discussed in terms of to what extent they support monotonicity. It will be shown that the vector space and probabilistic models, though nonmonotonic with respect to aboutness, do mimic conservative monotonicity under certain circumstances.
- will discuss the advantages and disadvantages of symbolic-based theoretical evaluation and analysis. The major strengths of this approach include:
  - *Enhanced perspective:* the functionality implied by the numeric formulation of a matching function in question can be teased out
  - *Transparency:* the assumptions made by a functional benchmark is explicitly stated
  - *New insights:* inductive evaluation has highlighted the import of monotonicity in retrieval functions, and its effect on retrieval performance. Designers of new matching functions could then provide functions that are conservatively monotonic with respect to the composition of information. More sensitive IR systems would then result.
- will detail the problems with modelling probabilistic IR models using symbolic means.

Bibliography

Bruza, P.D. and Huibers, T.W.C. 1994. Investigating aboutness axioms using information fields. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval.* Dublin, Ireland, 112-121.

Bruza, P.D., Song, D., and Wong, K.F. 2000. Aboutness from commonsense perspective. *Journal of American Society for Information Science*, 51(12):1090-1105.

Huibers, T.W.C., 1996. *An Axiomatic Theory for Information Retrieval*. Ph.D. Thesis, Utrecht University, The Netherlands

Hunter, A. 1996. Intelligent text handling using default logic, In Proceedings of the Eighth IEEE International Conference on Tools with Artificial Intelligence (TAI'96), 34-40, IEEE Computer Society Press.

Song, D. 2000. *A Commonsense Aboutness Theory for Information Retrieval Modeling.* Ph.D. Thesis, The Chinese University of Hong Kong.

Wong, K.F., Song, D., Bruza, P.D., & Cheng, C.H. (1998). Application of aboutness to functional benchmarking in information retrieval. *ACM Transactions on Information Systems.* (In press).