The Return of the Probability of Relevance

Norbert Fuhr
The University of Duisburg-Essen
Duisburg, Germany
norbert.fuhr@uni-due.de

Abstract

The probability ranking principle (PRP) proves that ranking documents by decreasing probability of relevance yields optimum retrieval quality. Most research on probabilistic models has focused only on producing a probabilistic ranking, without estimating the actual probabilities. In this talk, we discuss models for three types of modern IR applications which rely on calibrated values of the probability of relevance.

1. Vertical search deals with the aggregation of documents with different types or media (such as, e.g., Web pages, news, tweets, videos, images) in response to a query. Based on the probabilistic estimation of the number of relevant documents per resource, the decision-theoretic selection model describes the optimum solution for this problem.

2. The optimum clustering framework provides not only the first theoretic foundation for document clustering, it also proves the clustering hypothesis. Its key idea is to base cluster analysis and evaluation on a set of queries, by defining documents as being similar if they are relevant to the same queries.

3. The interactive PRP generalizes the classical PRP for interactive retrieval. It characterizes each situation in interactive retrieval as a list of choices, where each choice is described as the effort for evaluating it, the probability that the user will accept it, and the benefit resulting from acceptance. By developing appropriate parameter estimation methods, we can describe interactive retrieval by Markov models, which allow for a number of predictions.

With these models, it becomes possible to implement approaches based on solid theoretic foundations, which are more transparent than heuristic approaches, thus allowing for theory-guided adaptation and tuning.

About the Speaker

Dr. Norbert Fuhr is a full professor in the Department of Computer Science at the University of Duisburg-Essen. He obtained his Ph.D in Computer Science from the Technical University of Darmstadt in 1986 where he served as an assistant professor. He became Associate Professor in the computer science department of the University of Dortmund in 1991, before taking up his current position in 2002.

He has published more than 300 papers in the fields of IR, databases and digital libraries. His current research interests are retrieval models, networked digital library architectures, user-oriented retrieval methods and the evaluation of digital libraries.

He has served as regular PC member of many major international conferences related to information retrieval and digital libraries, such as ACM-SIGIR, CIKM, ECIR, SPIRE, ICDL, ECDL, ICADL, FQAS. He was PC chair of ECIR 2002, IR track chair of CIKM 2005 and Co-Chair of SIGIR 2007. For the German IR-group GI-FGIR, he served as Chair from 1992-2008. He also is a member of the editorial boards of the journals Information Retrieval, ACM Transactions on Information Systems, International Journal of Digital Libraries, and Foundations and Trends in Information Retrieval.

In 2012, he received the prestigious Gerald Salton Award in recognition of his significant, sustained and continuing contributions to research in information retrieval.

The committee particularly emphasised his *"pioneering contributions to the theoretical foundations of information retrieval and database systems. His work describing how learning methods can be used with retrieval models and indexing anticipated the current interest in learning ranking functions, his development of probabilistic retrieval models for database systems and XML was ground-breaking, and his recent work on retrieval models for interactive retrieval has inspired new research. His rigorous approach to research and research methods is an outstanding example for our field."*