REPORT
on
ACM SIGIR WORKSHOP ON MATHEMATICAL/FORMAL METHODS IN INFORMATION RETRIEVAL
MF/IR 2001
New Orleans, Louisiana, U.S.A.
September 7-12, 2001

The purpose of the MF/IR 2001 workshop was, on the one hand, to continue and enhance the results obtained so far, and on the other hand, to present, discuss, analyse, integrate the newer/newest results. Therefore MF/IR 2001 aimed at promoting discussion and interaction among those with theoretical and applicative research interests in mathematical/formal aspects of Information Retrieval, and also at being a forum for the presentation of both theoretical and applicative results (e.g., foundational issues; description and/or integration of models; retrieval applications; mathematical/formal techniques, properties and structures in IR; existing and/or new theories and theoretical aspects).

The following areas were addressed: Information Retrieval, Information Filtering, Information, Mining, Indexing and Retrieval, Hypermedia, World Wide Web Retrieval, Digital Libraries, Evaluation, Human Computer Interaction, User Modelling, where the different entities involved (e.g., documents, queries, indexing, retrieval, relevance, effectiveness, users, etc.) were modelled using any of, but not necessarily limited to, the following approaches: Classical Sets, Fuzzy Sets, Rough Sets, Vectors, Linear Space, Similarity, Functions, Probability, Functional Analysis, Algebra, Topology, Metric Spaces, Boolean Logic, Non-standard Logics, Fuzzy Logic, Quantum Logic, Matroid Theory, Graph Theory, Theory of Computation, Recursion Theory, Information Theory, Artificial Intelligence.

The following papers and results were presented.

Aggregated Representation for the Focussed Retrieval of Structured Documents. Gabriella Kazai, Queen Mary and Westfield College, University of London, UK; Mounia Lalmas, Queen Mary and Westfield College, University of London, UK; Thomas Rölleke, Queen Mary and Westfield College, University of London, UK — It was argued that structured documents allow to focus retrieval to the best entry point, i.e., to a relevant document component.

Document Instantiation for Relevance Feedback in the Bayesian Network Retrieval Model. Luis M. de Campos, Universidad de Granada, Spain; Juan F. Huete, Universidad de Granada, Spain; Juan M. Fernández-Luna, Universidad de Jaén, Spain — A new relevance feedback method was introduced based on Bayesian Networks, and experiments were reported.

Learning to Match and Cluster Entity Names. William Cohen, Whizbang Labs, USA; Jacob Richman, Carnegie Mellon University, USA. — Several techniques were presented and discussed which allowed for adaptively modifying similarity measures.

Document-Query Duality Meets Maximum Likelihood: The Answer is 3/5/8?. David Bodoff, Hong Kong University of Science and Technology — It was argued that in order to know how much to modify a document or query in response to relevance feedback, error models for them were needed, and then a maximum likelihood approach could be used to decide what should be modified.

Unitary Operators on the Document Space. Eduard Hoenkamp, Nijmegen Institute for Cognition and Information, The Netherlands — It was shown how LSI was based on a computationally more advantageous unitary transformation.

Theoretical Evaluation of IR Models Using Symbolic Means. Peter D. Bruza, University of Queensland, Australia; D. Song, University of Queensland, Australia — An overview of symbolic-based theoretical evaluation and analysis of IR models was provided.

On Applying Formal Grammar and Languages, and Deduction to Information Retrieval Modelling. Sandor Dominich, University of Veszprem, Hungary — It was shown that a set of documents in normal form is recursive allowing for designing an additional validation processor, and that the formal correctness of the query does imply its positive answerability.

Adaptive Feedback Methods in an Extended Boolean Model. Jongpill Choi, Ajou University, Korea; Minkoo Kim, Ajou University, Korea; Vijay V. Raghavan, University of Lousiana at Lafayete, USA — An adaptive way to improve the retrieval performance in an extended Boolean model was proposed.

International News Connection: A Real-time Online News Filtering and Classification System. Zhipping Zheng, University of Michigan, USA — International News Connection system was presented as an application of formal IR methods.

On Undirected Representation of Bayesian Networks. S.K.M. Wong, University of Regina, Canada; C.J. Butz, University of Regina, Canada; D. Wu, University of Regina, Canada — A new probabilistic networks, called Hierarchical Markov Network, was introduced, and it was shown that this network was equivalent to the Bayesian Network from which it was algorithmically obtained.

Bitmap Indexing-based Clustering and Retrieval of XML Documents. Jong P. Yoon, University of Louisiana at Lafayette, USA; Vijay Raghavan, University of Louisiana at Lafayette, USA; Venu Chakilan, University of Louisiana at Lafayette, USA — A bitmap indexing technique to cluster XML documents was presented.

To allow for maximum interaction, each presentation was scheduled for 30 minutes including questions. A 45 minutes discussion allowed further interaction between participants. Following the previous MF/IR 2000 Workshop, Athens, Greece, the exchange of ideas was very stimulating.

The MF/IR 2001 organisers were: Sandor Dominich (University of Veszprem, Hungary; Buckinghamshire Chilterns University College, U.K.), Mounia Lalmas (Queen Mary, University of London, England, U.K.), Keith van Rijsbergen (University of Glasgow, Scotland, U.K.), who, on behalf of MF/IR 2001.

They would like to thank all participants for presentations and discussions as well as ACM SIGIR for making this event possible.

Sandor Dominich
Mounia Lalmas
Keith van Rijsbergen