
Information Retrieval, 3, 171–172, 2000

Introduction

The papers in this special issue reflect research and development that has been conducted between 1995 and 2000 at the Swiss Federal Institute of Technology (ETH) Zurich and at its spin-off company Eurospider Information Technology AG. These five years probably represent the most significant period in the history of Information Retrieval (IR), where IR evolved from a rather academic discipline into an area where a much larger audience is involved. The papers in this issue are based on original publications that are cited in the bibliography.

The first paper on "New Approaches to Spoken Document Retrieval" by Wechsler and Schäuble shows what has been achieved on speech retrieval. In 1990 Peter Schäuble learned about the SPHINX speech recognition system. He predicted that within a few years speech recognition would be mature enough to perform retrieval on audio recordings that are no longer restricted to a narrow domain. When starting this new research area, we assumed that spoken document retrieval is not simply speech recognition plus information retrieval. Martin Wechsler discovered experimentally that surprisingly good retrieval effectiveness is achieved despite many recognition errors. This was the starting point for Elke Mittendorf to establish a new theoretical framework to study the effects of recognition errors on the retrieval effectiveness. Her findings are presented in "Information Retrieval Can Cope with Many Errors."

The advent of Multimedia Retrieval in the 90's has not only led to solutions but also to new research issues. One of them is concerned with determining the optimality of retrieval results. The increasing degree of distribution and the growing variance in transmission and inspection times of multimedia documents has led to situations where the classical probability principle is not helpful. In the paper "The Probability Ranking Principle Revisited" Wechsler and Schäuble describe a generalization of this classical principle.

A successful application of IR to retrieve solar radio spectrograms is presented by Csillaghy, Hinterberger, and Benz. The paper is based on Andre Csillaghy's doctoral dissertation where he developed a new indexing method that extracts indexing features that are relevant in the context of astronomy.

The method described in "Using the Cooccurrence of Words for Retrieval Weighting" by Mittendorf, Mateev, and Schäuble extends word based indexing features by pairs of words that cooccur in a text window. The paper not only shows that this extension improves the retrieval effectiveness, it also describes how this extension fits into the framework of probabilistic retrieval.

More and more retrieval applications are constrained by the lack of global inverse document frequencies (idfs). The global idfs may not be available because of a federated architecture or because of a heterogeneous document collection where subcollections are indexed by different indexing features such as graphical, audiovisual, and textual features. In the paper "Retrieving Information From a Distributed Heterogeneous Document Collection,"

Christoph Baumgarten tackles this problem with a new retrieval model and he even shows that it outperforms the non-distributed case.

Multilingual Information Retrieval (MLIR) allows users to formulate queries in their preferred language in order to retrieve relevant documents from a collection containing documents in multiple languages. As a consequence of globalization, there is increasing pressure for access to information without language or cultural barriers. The paper by Braschler and Schäuble entitled “Using Corpus-based Approaches in a System for Multilingual Information Retrieval” describes the state of the art of corpus-based MLIR.

Peter Schäuble

Elke Mittendorf