

Preface

This volume contains a selection of papers presented at the 13th International Symposium on String Processing and Information Retrieval (SPIRE), held October 11-13, 2006, in Glasgow, Scotland.

The SPIRE annual symposium provides an opportunity for both new and established researchers to present original contributions to areas such as

- string processing - dictionary algorithms, text searching, pattern matching, text compression, text mining, natural language processing, and automata-based string processing
- information retrieval languages, applications, and evaluation - IR modeling, indexing, ranking and filtering, interface design, visualization, crosslingual IR systems, multimedia IR, digital libraries, collaborative retrieval, Web related applications, XML, information retrieval from semi-structured data, text mining, and generation of structured data from text
- interaction of biology and computation - sequencing and applications in molecular biology, evolution and phylogenetics, recognition of genes and regulatory elements, and sequence driven protein structure prediction.

The accepted publications in the conference were selected from 102 papers submitted from over 20 countries in response to the Call for Papers. A total of 26+5 submissions were accepted as full and short papers respectively. Of those a number were invited to be expanded for inclusion in this special issue of the journal. Each of the papers underwent two rounds of reviews by external reviewers and by the Editors of the Special Issue, thus eventually ensuring a high-level of correctness and significance of the scientific contribution.

The papers selected cover the range of SPIRE subjects: “Output-Sensitive Autocompletion Search” by Bast, Mortensen and Weber who present a novel data structure that allows auto-completion of user queries to be executed rapidly. The structure is of similar size to an inverted file. Esuli, Fagni and Sebastiani in “Boosting Multi-label Hierarchical Text Categorization” present a variation of the AdaBoost.MH algorithm that deals with categorization of documents into hierarchical categories. The algorithm is shown to be successful both in terms of efficiency and accuracy. Providing means to better deal with ranking of documents based on multiple sources of evidence is described in “An Outranking Approach for Information Retrieval” by Farah and Vanderpooten. Here the authors argue that previous approaches to ranking that use multiple attributes fail to handle the imprecision often present in some attributes. They describe their novel procedure for dealing with forms of attribute and show through experimentation that their approach works well compared to common alternatives.

Moving from Information Retrieval focused research to String Processing, Fredriksson and Grabowski in “Efficient algorithms for pattern matching with general gaps, character classes and transposition invariance” address a particular class of pattern matching problem in which gaps are allowed in the pattern being searched, including so-called negative gaps. Finally, “A Compressed Self-Index using a Ziv-Lempel Dictionary”, by Russo and Oliveira describes a compressed self-index which improves on previous approaches by reducing search time complexity from quadratic to linear time in the size of the searched pattern.

We would like to thank Ricardo Baeza-Yates, who, on behalf of the Steering Committee, invited us to organize SPIRE 2006 and supported us at every step of the way. Thanks also are due to the program committee members who helped us with the additional reviewing.

Sheffield, December 2007

Fabio Crestani, Paolo Ferragina, Mark Sanderson