# Contents

Eighth Symposium on String Processing and Information Retrieval — SPIRE’2001

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>vii</td>
</tr>
<tr>
<td>Program Committee</td>
<td>viii</td>
</tr>
<tr>
<td>Additional Reviewers</td>
<td>viii</td>
</tr>
<tr>
<td>Organizing Committee and Sponsors</td>
<td>ix</td>
</tr>
<tr>
<td>Regular and Invited Papers</td>
<td></td>
</tr>
<tr>
<td>Invited Paper: Of Maps Bigger than the Empire</td>
<td>2</td>
</tr>
<tr>
<td>A. Apostolico</td>
<td></td>
</tr>
<tr>
<td>Distributed Query Processing Using Partitioned Inverted Files</td>
<td>10</td>
</tr>
<tr>
<td>C. Badue, R. Baeza-Yates, B. Ribeiro-Neto, and N. Ziviani</td>
<td></td>
</tr>
<tr>
<td>Relating Web Characteristics with Link Based Web Page Ranking</td>
<td>21</td>
</tr>
<tr>
<td>R. Baeza-Yates and C. Castillo</td>
<td></td>
</tr>
<tr>
<td>Compaction Techniques for Nextword Indexes</td>
<td>33</td>
</tr>
<tr>
<td>D. Bahle, H. Williams, and J. Zobel</td>
<td></td>
</tr>
<tr>
<td>A Subquadratic Algorithm for Cluster and Outlier Detection in Massive Metric Data</td>
<td>46</td>
</tr>
<tr>
<td>E. Chávez</td>
<td></td>
</tr>
<tr>
<td>Speeding-up Hirschberg and Hunt-Szymanski LCS Algorithms</td>
<td>59</td>
</tr>
<tr>
<td>M. Crochemore, C. Iliopoulos, and Y. Pinzon</td>
<td></td>
</tr>
<tr>
<td>Semantic Thesaurus for Automatic Expanded Query in Information Retrieval</td>
<td>68</td>
</tr>
<tr>
<td>M. González and V. Strube de Lima</td>
<td></td>
</tr>
<tr>
<td>Exact Distribution of Deletion Sizes for Unavoidable Strings</td>
<td>76</td>
</tr>
<tr>
<td>C. Heitsch</td>
<td></td>
</tr>
<tr>
<td>On Using Two-Phase Filtering in Indexed Approximate String Matching with Application to Searching Unique Oligonucleotides</td>
<td>84</td>
</tr>
<tr>
<td>H. Hyyrö</td>
<td></td>
</tr>
<tr>
<td>On-Line Construction of Symmetric Compact Directed Acyclic Word Graphs</td>
<td>96</td>
</tr>
<tr>
<td>S. Inenaga, H. Hoshino, A. Shinohara, M. Takeda, and S. Arikawa</td>
<td></td>
</tr>
<tr>
<td>Musical Sequence Comparison for Melodic and Rhythmic Similarities</td>
<td>111</td>
</tr>
<tr>
<td>T. Kadota, M. Hirao, A. Ishino, M. Takeda, A. Shinohara, and F. Matsuo</td>
<td></td>
</tr>
<tr>
<td>A Model for the Representation and Focussed Retrieval of Structured Documents Based on Fuzzy Aggregation</td>
<td>123</td>
</tr>
<tr>
<td>G. Kazai, M. Lalmas, and T. Rölleke</td>
<td></td>
</tr>
</tbody>
</table>
Preface

SPIRE’2001 is a Symposium on String Processing and Information Retrieval, now in its eighth edition. This series of meetings originated in South America (Belo Horizonte and Recife, Brazil, in 1993 and 1996; Valparaíso, Chile, in 1995 and 1997) and were originally called WSP (Workshop on String Processing).

Starting in 1998, at Santa Cruz, Bolivia, the focus of the workshop was broadened to include the area of information retrieval due to its increasing relevance and its inter-relationship with the area of string processing. SPIRE’1999, at Cancun, Mexico; SPIRE’2000, at A Coruña, Spain; and SPIRE’2001 continued this trend.

One of the main goals of SPIRE is to facilitate the cross fertilization between those different but related fields. I believe this objective is being met. This year we received a number of high quality contributions from the areas of string processing, information retrieval, and on problems resulting from their combination.

SPIRE’2001 was held in Laguna de San Rafael, Chile. The symposium included the presentations of 20 full papers and 5 poster papers (from 33 submitted), as well as 3 invited talks by Alberto Apostolico (Purdue University, USA, and University of Padova, Italy), Alistair Moffat (University of Melbourne, Australia) and Yuri Zieman (Unveil Inc., USA). This year we created the category of poster papers to encourage the dissemination of promising work that still requires further development.

I wish to thank all the members of the Program Committee and the additional reviewers for their careful and mostly timely evaluation of the submissions. I would also like to thank the local organization staff for their hard and enthusiastic work, and in particular to Ricardo Baeza-Yates for compensating my inexperience as a Program Committee Chair.

Gonzalo Navarro
SPIRE’2001 Program Committee Chair
Santiago, Chile
Program Committee

Amihood Amir, Bar-Ilan University, Israel
Alberto Apostolico, Purdue University, USA; University of Padova, Italy
Ricardo Baeza-Yates, Universidad de Chile, Chile
Josep Blat, Universitat Pompeu Fabra, Spain
Nieves Brisaboa, Universidad de A Coruña, Spain
Edgar Chávez, Universidad Michoacana, Mexico
Maxime Crochemore, Université de Marne-la-Vallée, France
Russell Deaton, University of Arkansas, USA
Pablo de la Fuente, Universidad de Valladolid, Spain
Efthimis Etthimiadis, University of Washington, USA
Paolo Ferragina, University of Pisa, Italy
Edward Fox, Virginia Tech., USA
Max Garzón, University of Memphis, USA
John Kececioglu, University of Arizona, USA
João Paulo Kitajima, Universidade de Campinas, Brazil
Eduardo Laber, Universidade Federal do Rio de Janeiro, Brazil
Mun-Kew Leong, BIGontheNet, Singapore
Massimo Melucci, University of Padova, Italy
Ruy Milidiu, Pontificia Universidade Católica de Rio, Brazil
Alistair Moffat, University of Melbourne, Australia
Sung Hyon Myaeng, Chungnam National University, Korea
Mario Nascimento, University of Alberta, Canada
Gonzalo Navarro, Universidad de Chile, Chile (Chair)
Arlindo Oliveira, INESC, Portugal
Berthier Ribeiro-Neto, Universidade Federal de Minas Gerais, Brazil
John Rose, University of Tokyo, Japan
Marie-France Sagot, Institut Pasteur, France
João Setubal, Universidade de Campinas, Brazil
Masayuki Takeda, Kyushu University, Japan
Jorma Tarhio, Helsinki University of Technology, Finland
Esko Ukkonen, University of Helsinki, Finland
Nivio Ziviani, Universidade Federal de Minas Gerais, Brazil

Additional Reviewers

Charles Ornelas Almeida
Claudine Santos Badue
Alejandro Bassi
Daniela Ferreira da Mota
Alejandro Hevia
Costas S. Iliopoulos
Juha Kärkkäinen
Pekka Kilpeläinen
Shmuel T. Klein
Stuart Margolis
Edleno Silva de Moura
Fernando das Neves
Artur Alves Pessoa
Mathieu Raffinot
Davood Rafiei
Raul Renteria
Altigran Soares da Silva
Jesús Vegas
Organizing Committee

Ricardo Baeza-Yates, Univ. de Chile, Chile
Edgar Chávez, Univ. Michoacana, Mexico
Gonzalo Navarro, Univ. de Chile, Chile (Chair)
Cuauhtemoc Rivera-Loaiza, Univ. Michoacana, Mexico

Sponsors

CYTED Project VII.19 RIBIDI (Recuperación de Información y Bibliotecas Digitales)
SCCC (Sociedad Chilena de Ciencia de la Computación)
Universidad de Chile, Santiago, Chile
Universidad Michoacana de San Nicolas de Hidalgo, Morelia, Mexico