

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Alfred Kobsa

*University of California, Irvine, CA, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*TU Dortmund University, Germany*

Madhu Sudan

*Microsoft Research, Cambridge, MA, USA*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Gerhard Weikum

*Max Planck Institute for Informatics, Saarbruecken, Germany*

Oren Kurland Moshe Lewenstein  
Ely Porat (Eds.)

# String Processing and Information Retrieval

20th International Symposium, SPIRE 2013  
Jerusalem, Israel, October 7-9, 2013  
Proceedings

 Springer

## Volume Editors

Oren Kurland  
Technion Institute of Technology  
Faculty of Industrial Engineering and Management Technion  
Haifa 32000, Israel  
E-mail: kurland@ie.technion.ac.il

Moshe Lewenstein  
Bar-Ilan University  
Department of Computer Science  
Ramat-Gan 52900, Israel  
E-mail: moshe@cs.biu.ac.il

Ely Porat  
Bar-Ilan University  
Department of Computer Science  
Ramat-Gan 52900, Israel  
E-mail: porately@cs.biu.ac.il

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-319-02431-8 e-ISBN 978-3-319-02432-5  
DOI 10.1007/978-3-319-02432-5  
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013948098

CR Subject Classification (1998): H.3, H.2.8, I.5, I.2.7, F.2, J.3

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*The original version of the book was revised:  
The copyright line was incorrect. The Erratum  
to the book is available at  
DOI: [10.1007/978-3-319-02432-5\\_33](https://doi.org/10.1007/978-3-319-02432-5_33)*

# Preface

In the 20 years since its inception in 1993 the International Symposium on String Processing and Information Retrieval (SPIRE) has become the reference meeting for the interdisciplinary community of researchers whose activity lies at the crossroads of string processing and information retrieval. This volume contains the proceedings of SPIRE 2013, the 20th symposium in the series.

The first four events concentrated mainly on string processing, and were held in South America under the title South American Workshop on String Processing (WSP) in 1993 (Belo Horizonte, Brazil), 1995 (Valparaiso, Chile), 1996 (Recife, Brazil), and 1997 (Valparaiso, Chile). WSP was renamed SPIRE in 1998 (Santa Cruz, Bolivia) when the scope of the event was broadened to include information retrieval. The change was motivated by the increasing relevance of information retrieval and its close interrelationship with the general area of string processing. From 1999 to 2007, the venue of SPIRE alternated between South / Latin America (odd years) and Europe (even years), with Cancun, Mexico in 1999; A Coruna, Spain in 2000; Laguna de San Rafael, Chile in 2001; Lisbon, Portugal in 2002; Manaus, Brazil in 2003; Padova, Italy in 2004; Buenos Aires, Argentina in 2005; Glasgow, UK in 2006; and Santiago, Chile in 2007. This pattern was broken when SPIRE 2008 was held in Melbourne, Australia, but it was restarted in 2009 when the venue was in Saariselkä, Finland, followed by Los Cabos, Mexico in 2010, Pisa, Italy in 2011, and in Cartagena de Indias, Colombia in 2012.

SPIRE 2013 was held in Jerusalem, Israel. The call for papers resulted in the submission of 60 papers. Each submitted paper was reviewed by at least three of the 40 members of the Program Committee, who eventually engaged in discussions coordinated by the three PC chairmen in cases of lack of consensus. We believe this resulted in a very accurate selection of the truly best submitted papers. As a result, 18 long papers and 10 short papers were accepted and have been published in these proceedings.

The main conference featured keynote speeches by Ido Dagan, Roberto Grossi, Robert Krauthgamer, and Yossi Matias, plus the presentations of the 18 full papers and 10 short papers. Following the main conference, on October 10, SPIRE 2013 hosted the Workshop on Compression, Text, and Algorithms (WCTA 2013).

We would like to take the opportunity to thank Yahoo!, Google, Bar-Ilan University, and i-Core (Center of Excellence in Algorithms). All of them provided generous sponsorship. Thanks also to all the members of the Program Committee and to the additional reviewers, who went to great lengths to ensure the high quality of this conference, and to the coordinator of the SPIRE Steering

Committee, Ricardo Baeza-Yates, who provided assistance and guidance in the organization. We would like to thank the Local Organization Committee consisting of Amihod Amir, Tomi Klein, and Tsvi Kopelowitz (as well as ourselves). It is due to them that the organization of SPIRE 2013 was not just hard work, but also a pleasure.

October 2013

Oren Kurland  
Moshe Lewenstein  
Ely Porat

# Organization

## Program Committee

Giambattista Amati	Fondazione Ugo Bordoni
Amihood Amir	Bar-Ilan University and Johns Hopkins University
Alberto Apostolico	Univ. of Padova and Georgia Tech
Ricardo Baeza-Yates	Yahoo! Research
Ayelet Butman	Holon Institute of Technology
Edgar Chavez	Universidad Michoacana
Raphael Clifford	University of Bristol
Carsten Eickhoff	Delft University of Technology
Johannes Fischer	Karlsruhe Institute of Technology
Inge Li Gørtz	Technical University of Denmark
Shunsuke Inenaga	Kyushu University
Markus Jalsenius	University of Bristol
Gareth Jones	Dublin City University
Jaap Kamps	University of Amsterdam
Tsvi Kopelowitz	Weizmann Institute of Science
Oren Kurland	Technion
Gad M. Landau	Haifa University
Avivit Levy	Shenkar College
Moshe Lewenstein	Bar Ilan University
Stefano Lonardi	UC Riverside
Andrew McGregor	University of Massachusetts, Amherst
Alistair Moffat	The University of Melbourne
Ian Munro	University of Waterloo
Gonzalo Navarro	University of Chile
Yakov Nekrich	University of Chile
Krzysztof Onak	IBM Research
Ely Porat	Bar-Ilan University
Berthier Ribeiro-Neto	Google Research
Benjamin Sach	University of Warwick
Rodrygo L.T. Santos	University of Glasgow
Srinivasa Rao Satti	University of Aarhus
Rahul Shah	Louisiana State Univeristy
Chris Thachuk	University of Oxford
Paul Thomas	CSIRO
Dekel Tsur	Ben Gurion University
Esko Ukkonen	University of Helsinki

Oren Weimann  
David Woodruff  
Nivio Ziviani  
Guido Zuccon

University of Haifa  
IBM Almaden  
Federal University of Minas Gerais  
CSIRO

## **Additional Reviewers**

Atserias, Jordi  
Bachrach, Yoram  
Bessa, Aline  
Bingmann, Timo  
Biswas, Sudip  
Claude, Francisco  
Davoodi, Pooya  
Ferrada, Héctor  
Flouri, Tomas  
Gagie, Travis  
Gog, Simon  
Gupta, Ankur  
Hata, Itamar  
Hernandez, Cecilia  
Konow, Roberto

Ku, Tsung-Han  
Lecroq, Thierry  
Nakashima, Yuto  
Patil, Manish  
Petri, Matthias  
Rozenberg, Liat  
Shiftan, Ariel  
Sirén, Jouni  
Tanaseichuk, Olga  
Tatti, Nikolaj  
Thankachan, Sharma V.  
Velo, Adriano  
Vind, Soren  
Wootters, Mary



# Table of Contents

Consolidating and Exploring Information via Textual Inference . . . . .	1
<i>Ido Dagan</i>	
Pattern Discovery and Listing in Graphs . . . . .	2
<i>Roberto Grossi</i>	
Efficient Approximation of Edit Distance . . . . .	3
<i>Robert Krauthgamer</i>	
Nowcasting with Google Trends . . . . .	4
<i>Yossi Matias</i>	
Space-Efficient Construction of the Burrows-Wheeler Transform . . . . .	5
<i>Timo Beller, Maïke Zwerger, Simon Gog, and Enno Ohlebusch</i>	
Using Mutual Influence to Improve Recommendations . . . . .	17
<i>Aline Bessa, Adriano Veloso, and Nivio Ziviani</i>	
Position-Restricted Substring Searching over Small Alphabets . . . . .	29
<i>Sudip Biswas, Tsung-Han Ku, Rahul Shah, and Sharma V. Thankachan</i>	
Simulation Study of Multi-threading in Web Search Engine Processors . . . . .	37
<i>Carolina Bonacic and Mauricio Marin</i>	
Query Processing in Highly-Loaded Search Engines . . . . .	49
<i>Daniele Broccolo, Craig Macdonald, Salvatore Orlando, Iadh Ounis, Raffaele Perego, Fabrizio Silvestri, and Nicola Tonellotto</i>	
Indexes for Jumbled Pattern Matching in Strings, Trees and Graphs . . . .	56
<i>Ferdinando Cicalese, Travis Gagie, Emanuele Giaquinta, Eduardo Sany Laber, Zsuzsanna Lipták, Romeo Rizzi, and Alexandru I. Tomescu</i>	
Adaptive Data Structures for Permutations and Binary Relations . . . . .	64
<i>Francisco Claude and J. Ian Munro</i>	
Document Listing on Versioned Documents . . . . .	72
<i>Francisco Claude and J. Ian Munro</i>	
Order-Preserving Incomplete Suffix Trees and Order-Preserving Indexes . . . . .	84
<i>Maxime Crochemore, Costas S. Iliopoulos, Tomasz Kociumaka, Marcin Kubica, Alessio Langiu, Solon P. Pissis, Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń</i>	

Compact Queriable Representations of Raster Data . . . . .	96
<i>Guillermo de Bernardo, Sandra Álvarez-García, Nieves R. Brisaboa, Gonzalo Navarro, and Oscar Pedreira</i>	
Top- <i>k</i> Color Queries On Tree Paths . . . . .	109
<i>Stephane Durocher, Rahul Shah, Matthew Skala, and Sharma V. Thankachan</i>	
A Lempel-Ziv Compressed Structure for Document Listing . . . . .	116
<i>Héctor Ferrada and Gonzalo Navarro</i>	
Minimal Discriminating Words Problem Revisited . . . . .	129
<i>Paweł Gawrychowski, Gregory Kucherov, Yakov Nekrich, and Tatiana Starikovskaya</i>	
Adding Compression and Blended Search to a Compact Two-Level Suffix Array . . . . .	141
<i>Simon Gog and Alistair Moffat</i>	
You Are What You Eat: Learning User Tastes for Rating Prediction . . . . .	153
<i>Morgan Harvey, Bernd Ludwig, and David Elsweiler</i>	
Discovering Dense Subgraphs in Parallel for Compressing Web and Social Networks . . . . .	165
<i>Cecilia Hernández and Mauricio Marín</i>	
Faster Lyndon Factorization Algorithms for SLP and LZ78 Compressed Text . . . . .	174
<i>Tomohiro I, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda</i>	
Lossless Compression of Rotated Maskless Lithography Images . . . . .	186
<i>Shmuel Tomi Klein, Dana Shapira, and Gal Shelef</i>	
Learning URL Normalization Rules Using Multiple Alignment of Sequences . . . . .	197
<i>Kaio Wagner Lima Rodrigues, Marco Cristo Edleno Silva de Moura, and Altigran Soares da Silva</i>	
On Two-Dimensional Lyndon Words . . . . .	206
<i>Shoshana Marcus and Dina Sokol</i>	
Fully-Online Grammar Compression . . . . .	218
<i>Shirou Maruyama, Yasuo Tabei, Hiroshi Sakamoto, and Kunihiko Sadakane</i>	
Solving Graph Isomorphism Using Parameterized Matching . . . . .	230
<i>Juan Mendivelso, Sunghwan Kim, Sameh Elnikety, Yuxiong He, Seung-won Hwang, and Yoan Pinzón</i>	

Suffix Array of Alignment: A Practical Index for Similar Data . . . . .	243
<i>Joong Chae Na, Heejin Park, Sunho Lee, Minsung Hong, Thierry Lecroq, Laurent Mouchard, and Kunsoo Park</i>	
Faster Top- <i>k</i> Document Retrieval in Optimal Space . . . . .	255
<i>Gonzalo Navarro and Sharma V. Thankachan</i>	
Faster Range LCP Queries . . . . .	263
<i>Manish Patil, Rahul Shah, and Sharma V. Thankachan</i>	
Learning to Schedule Webpage Updates Using Genetic Programming . . .	271
<i>Aécio S.R. Santos, Nivio Ziviani, Jussara Almeida, Cristiano R. Carvalho, Edleno Silva de Moura, and Altigran Soares da Silva</i>	
Accurate Profiling of Microbial Communities from Massively Parallel Sequencing Using Convex Optimization . . . . .	279
<i>Or Zuk, Amnon Amir, Amit Zeisel, Ohad Shamir, and Noam Shental</i>	
Distributed Query Processing on Compressed Graphs Using K2-Trees . . . . .	298
<i>Sandra Álvarez-García, Nieves R. Brisaboa, Carlos Gómez-Pantoja, and Mauricio Marin</i>	
Erratum to: String Processing and Information Retrieval . . . . .	E1
<i>Oren Kurland, Moshe Lewenstein, and Ely Porat</i>	
<b>Author Index</b> . . . . .	311