Lecture Notes in Computer Science 9309

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7407

Costas S. Iliopoulos · Simon J. Puglisi
Emine Yilmaz (Eds.)

String Processing and Information Retrieval

22nd International Symposium, SPIRE 2015
London, UK, September 1–4, 2015
Proceedings

Springer

Editors
Costas S. Iliopoulos
King's College London
London
UK

Emine Yilmaz
University College London
London
UK

Simon J. Puglisi
University of Helsinki
Helsinki
Finland

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-23825-8 ISBN 978-3-319-23826-5 (eBook)
DOI 10.1007/978-3-319-23826-5

Library of Congress Control Number: 2015947399

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

From humble beginnings as a regional meeting focused on string algorithms (under the auspices of WSP: South American Workshop on String Processing), the International Symposium on String Processing and Information Retrieval (SPIRE) has, in the last two decades, developed into a vibrant conference at the broad nexus of algorithms and data structures for sequences and graphs, data compression, databases, data mining, and information retrieval.

This volume contains the papers presented at SPIRE 2015, the 22nd International Symposium on String Processing and Information Retrieval, held from August 31 to September 2, 2015 in London, UK, in the Great Hall of King's College London's Strand Campus. There were 90 submissions. Each submission was reviewed by at least 3, and on average 3.1, Program Committee members. The committee decided to accept 34 papers. The program also included 3 invited talks.

The main conference, which spanned the three days during 1–3 September, featured keynote talks by Aristides Gionis (Aalto Univeristy, Finland), Mounia Lalmas (Yahoo! Labs London, UK), and Rajeev Raman (University of Leicester, UK), together with presentations by authors of the 33 accepted papers. The 10th Workshop on Compression, Text, and Algorithms (WCTA 2015) was then held on September 4, the day immediately after the main conference, as has become a recent tradition. WCTA was coordinated this year by Travis Gagie and Tatiana Starikovskaya, and featured a keynote talk from Richard Durbin of the Wellcome Trust Sanger Institute, UK.

We take this opportunity to thank King's College London for its generous sponsorship of SPIRE this year. Our deep thanks also go to all the members of this year's Program Committee and additional reviewers, for the prompt, thorough reviewing and vibrant discussion that made our job as chairs easy. We thank the SPIRE Steering Committee, for giving us the opportunity to host this wonderful community of researchers in London, and finally, the Local Organizing Committee (led by Solon Pissis), for their efforts to ensure that the whole week ran smoothly, and that a relaxed and inspiring time was had by all.

July 2015

Costas S. Iliopoulos
Simon J. Puglisi
Emine Yilmaz

Organization

Program Committee

Sengor Altingovde	
Amihood Amir	Bar-Ilan University, Israel
Leif Azzopardi	University of Glasgow, UK
Golnaz Badkobeh	University of Sheffield, UK
Hideo Bannai	Kyushu University, Japan
Philip Bille	Technical University of Denmark, Denmark
Christina Boucher	Colorado State University, CO, USA
Ben Carterette	University of Delaware, DE, USA
Charles Clarke	University of Waterloo, Canada
Gianluca Demartini	University of Sheffield, UK
Johannes Fischer	TU Dortmund, Germany
Travis Gagie	University of Helsinki, Finland
Paweł Gawrychowski	University of Warsaw, Poland
Simon Gog	Karlsruhe Institute of Technology, Germany
Danny Hermelin	Ben-Gurion University, Israel
Djoerd Hiemstra	University of Twente, The Netherlands
Katja Hofmann	Microsoft Research Cambridge, UK
Costas S. Iliopoulos	King's College London, UK
Jaap Kamps	University of Amsterdam, The Netherlands
Evangelos Kanoulas	University of Amsterdam, The Netherlands
Gabriella Kazai	Lumi, Semion Ltd, UK
Juha Kärkkäinen	University of Helsinki, Finland
Susana Ladra	University of A Coruña, Spain
Gad Landau	University of Haifa, Israel
Zsuzsanna Lipták	University of Verona, Italy
Gonzalo Navarro	University of Chile, Chile
Kunsoo Park	Seoul National University, South Korea
Nadia Pisanti	University of Pisa, Italy
Solon Pissis	King's College London, UK
Simon J. Puglisi	University of Helsinki, Finland
Jakub Radoszewski	University of Warsaw, Poland
Falk Scholer	RMIT, Australia
Marinella Sciortino	University of Palermo, Italy
Jouni Sirén	Wellcome Trust Sanger Institute, UK
Tatiana Starikovskaya	University of Bristol, UK
Torsten Suel	NYU Poly, NY, USA
Yasuo Tabei	Japan Science and Technology Agency, Japan

Rossano Venturini	University of Pisa, Italy
Grace Yang	Georgetown University, DC, USA
Emine Yilmaz	University College London, UK

Additional Reviewers

Abouelhoda, Mohamed
Amit, Mika
Belazzougui, Djamal
Bilò, Davide
Bingmann, Timo
Brown, C. Titus
Christiansen, Anders Roy
Cunial, Fabio
Della Vedova, Gianluca
Doerr, Benjamin
Eisenberg, Estrella
Epifanio, Chiara
Farruggia, Andrea
Fici, Gabriele
Ganguly, Debasis
Gasieniec, Leszek
Giaquinta, Emanuele
Grossi, Roberto
Harrison, Thomas
I, Tomohiro
Ilie, Lucian
Inenaga, Shunsuke
Karimi, Sarvnaz
Kayaaslan, Enver
Kempa, Dominik
Keogh, Eamonn
Korkin, Dmitry
Kurpicz, Florian
Köppl, Dominik
Larsson, N. Jesper
Levy, Avivit
Lewenstein, Moshe
Mantaci, Sabrina
Marino, Andrea
Markov, Ilya
Metke, Alejandro
Micale, Giovanni
Na, Joong Chae
Nadalin, Francesca
Ozcan, Rifat
Piatkowski, Marcin
Pulvirenti, Alfredo
Rahman, M. Sohel
Raymond, Rob
Reynier, Pierre-Alain
Rodriguez, Juan
Rosone, Giovanna
Rozenberg, Liat
Sacomoto, Gustavo
Salmela, Leena
Seco, Diego
Shangsong, Liang
Song, Xuemeng
Straszak, Damian
Sugimoto, Shiho
Tischler, German
Tomescu, Alexandru I.
Turpin, Andrew
Valenzuela, Daniel
Välimäki, Niko
Walen, Tomasz
Weimann, Oren

Invited Talks

Computational Problems in Mining Urban Data

Aristides Gionis

Department of Computer Science, Aalto University

With the fast growth of smart devices and sensor networks, large amounts of data are collected recording location, activity, and mobility of people living in urban environments. Additionally, data generated on location-aware social media provide rich information about places where people spend their time (shopping malls, café s, parks, etc). The availability of this type of data provides novel opportunities for developing methods for extracting interesting patterns, detecting trends, modelling people's behaviour, and eventually building intelligent systems that improve the interaction of citizens with their cities and help them to utilize better the available resources. In this talk we will review recent work in the area of mining urban data. We formulate and discuss computational problems motivated by applications in detecting events, mining trajectories, finding similar neighbourhoods, and recommending locations.

A Journey into Evaluation: From Retrieval Effectiveness to User Engagement

Mounia Lalmas

Yahoo! Labs London

Building retrieval systems that return results to users that satisfy their information need is one thing; Information Retrieval has a long history in evaluating how effective retrieval systems are. Building a retrieval system that not only returns good results to users, but does so in a way that users will want to use that system again is something more challenging; a positive search experience has been shown to lead to users engaging long-term with the retrieval system. In this talk, I will review state-of-the-art evaluation approaches for search, with respect to retrieval effectiveness but also user satisfaction. I will then focus on those approaches aiming at evaluating user engagement, and describe current works in this area within and outside the search realm. The talk will end with the proposal of a framework incorporating effectiveness evaluation into user engagement in search. An important component of this framework is to consider both within- and across-search session measurement.

Encodings = (Data Structures) - (Data)

Rajeev Raman

Department of Computer Science, University of Leicester

Driven by the increasing need to analyze and search for complex patterns in very large data sets, the area of compressed and succinct data structures has grown rapidly in the last 10-15 years. Such data structures have very low memory requirements, allowing them to fit into the main memory of a computer, which in turn avoids expensive computation on hard disks.

This talk will focus on a sub-topic that has become popular recently: encoding "the data structure" itself. Some data structuring problems involve supporting queries on data, but the queries that need to be supported do not allow the original data to be deduced from the queries. This presents opportunities to obtain space savings even when the data is incompressible, by pre-processing the data, extracting only the information needed to answer the queries, and then deleting the data. The minimum information needed to answer the queries is called the effective entropy of the problem: precisely determining the effective entropy can involve interesting combinatorics.

Contents