

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, Lancaster, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Zurich, Switzerland

John C. Mitchell

Stanford University, Stanford, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

C. Pandu Rangan

Indian Institute of Technology Madras, Chennai, India

Bernhard Steffen

TU Dortmund University, Dortmund, Germany

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA


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


Carlos Martín-Vide · Alexander Okhotin
Dana Shapira (Eds.)


Language and Automata Theory and Applications

13th International Conference, LATA 2019
St. Petersburg, Russia, March 26–29, 2019
Proceedings

Editors

Carlos Martín-Vide 
Rovira i Virgili University
Tarragona, Spain

Dana Shapira 
Ariel University
Ariel, Israel

Alexander Okhotin 
Saint Petersburg State University
St. Petersburg, Russia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-13434-1 ISBN 978-3-030-13435-8 (eBook)
<https://doi.org/10.1007/978-3-030-13435-8>

Library of Congress Control Number: 2019931952

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

© Springer Nature Switzerland AG 2019

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

These proceedings contain the papers that were presented at the 13th International Conference on Language and Automata Theory and Applications (LATA 2019), held in Saint Petersburg, Russia, during March 26–29, 2019.

The scope of LATA is rather broad, including: algebraic language theory, algorithms for semi-structured data mining, algorithms on automata and words, automata and logic, automata for system analysis and program verification, automata networks, automatic structures, codes, combinatorics on words, computational complexity, concurrency and Petri nets, data and image compression, descriptive complexity, foundations of finite state technology, foundations of XML, grammars (Chomsky hierarchy, contextual, unification, categorial, etc.), grammatical inference and algorithmic learning, graphs and graph transformation, language varieties and semigroups, language-based cryptography, mathematical and logical foundations of programming methodologies, parallel and regulated rewriting, parsing, patterns, power series, string processing algorithms, symbolic dynamics, term rewriting, transducers, trees, tree languages and tree automata, and weighted automata.

LATA 2019 received 98 submissions. Every paper was reviewed by three Programme Committee members. There were also some external experts consulted. After a thorough and lively discussion phase, the committee decided to accept 31 papers (which represents a competitive acceptance rate of about 32%). The conference program included five invited talks as well.

The excellent facilities provided by the EasyChair conference management system allowed us to deal with the submissions successfully and handle the preparation of these proceedings in time.

We would like to thank all invited speakers and authors for their contributions, the Program Committee and the external reviewers for their cooperation, and Springer for its very professional publishing work.

January 2019

Carlos Martín-Vide
Alexander Okhotin
Dana Shapira

Organization

Program Committee

Krishnendu Chatterjee	Institute of Science and Technology, Austria
Bruno Courcelle	University of Bordeaux, France
Manfred Droste	University of Leipzig, Germany
Travis Gagie	Diego Portales University, Chile
Peter Habermehl	Paris Diderot University, France
Tero Harju	University of Turku, Finland
Markus Holzer	University of Giessen, Germany
Radu Iosif	Verimag, France
Kazuo Iwama	Kyoto University, Japan
Juhani Karhumäki	University of Turku, Finland
Lila Kari	University of Waterloo, Canada
Juha Kärkkäinen	University of Helsinki, Finland
Bakhadyr Khoussainov	The University of Auckland, New Zealand
Sergey Kitaev	University of Strathclyde, UK
Shmuel Tomi Klein	Bar-Ilan University, Israel
Olga Kouchnarenko	University of Franche-Comté, France
Thierry Lecroq	University of Rouen, France
Markus Lohrey	University of Siegen, Germany
Sebastian Maneth	University of Bremen, Germany
Carlos Martín-Vide (Chair)	Rovira i Virgili University, Spain
Giancarlo Mauri	University of Milano-Bicocca, Italy
Filippo Mignosi	University of L'Aquila, Italy
Victor Mitran	Polytechnic University of Madrid, Spain
Joachim Niehren	Inria Lille, France
Alexander Okhotin	Saint Petersburg State University, Russian Federation
Dominique Perrin	University of Paris-Est, France
Matteo Pradella	Polytechnic University of Milan, Italy
Jean-François Raskin	Université Libre de Bruxelles, Belgium
Marco Roveri	Bruno Kessler Foundation, Italy
Karen Rudie	Queen's University, Canada
Wojciech Rytter	University of Warsaw, Poland
Kai Salomaa	Queen's University, Canada
Sven Schewe	University of Liverpool, UK
Helmut Seidl	Technical University of Munich, Germany
Ayumi Shinohara	Tohoku University, Japan
Hans Ulrich Simon	Ruhr-University of Bochum, Germany
William F. Smyth	McMaster University, Canada
Frank Stephan	National University of Singapore, Singapore

Martin Sulzmann	Karlsruhe University of Applied Sciences, Germany
Jorma Tarhio	Aalto University, Finland
Stefano Tonetta	Bruno Kessler Foundation, Italy
Rob van Glabbeek	Data61, CSIRO, Australia
Margus Veanes	Microsoft Research, USA
Mahesh Viswanathan	University of Illinois, Urbana-Champaign, USA
Mikhail Volkov	Ural Federal University, Russian Federation
Fang Yu	National Chengchi University, Taiwan
Hans Zantema	Eindhoven University of Technology, New Zealand

Additional Reviewers

Abdeddaïm, Saïd	Giorgetti, Alain	Nakamura, Katsuhiko
Baeza-Yates, Ricardo	Godin, Thibault	Nehaniv, Chrystofer
Balogh, Jozsef	Groote, Jan Friso	Nugues, Pierre
Beier, Simon	Groschwitz, Jonas	Paul, Erik
Bernardinello, Luca	Gutiérrez, Martín	Penelle, Vincent
Bersani, Marcello M.	Guyeux, Christophe	Persiano, Giuseppe
Bollig, Benedikt	Heam, Pierre-Cyrille	Plandowski, Wojciech
Bride, Hadrien	Hellouin de Menibus, Benjamin	Popa, Alexandru
Cadilhac, Michaël	Hendrian, Diptarama	Praveen, M.
Capelli, Florent	Hugot, Vincent	Prunescu, Mihai
Chiari, Michele	Jeandel, Emmanuel	Pyatkin, Artem
Choudhury, Salimur	Kari, Jarkko	Quaas, Karin
Ciobanu, Laura	Kociumaka, Tomasz	Radoszewski, Jakub
Courcelle, Bruno	Kopczynski, Eryk	Rossi, Matteo
Crespi Reghizzi, Stefano	Kuske, Dietrich	Sakho, Momar
Crosetti, Nicolas	Lemay, Aurelien	Sangnier, Arnaud
De La Higuera, Colin	Lisitsa, Alexei	Schmidt-Schauss, Manfred
Delecroix, Vincent	Lonati, Violetta	Starikovskaya, Tatiana
Dennunzio, Alberto	Luttik, Bas	Talbot, Jean-Marc
Dose, Titus	Maletti, Andreas	Vaszil, György
Dücker, Stefan	Mandrioli, Dino	Wendlandt, Matthias
Ferretti, Claudio	Mens, Irini	Willems, Tim
Fici, Gabriele	Mhaskar, Neerja	Yakovets, Nikolay
Fülöp, Zoltan	Morawska, Barbara	Yoshinaka, Ryo
Geffert, Viliam	Morzenti, Angelo	Zetsche, Georg
Gilroy, SORCHA		

Abstracts of Invited Papers

Searching and Indexing Compressed Text

Paweł Gawrychowski

Institute of Computer Science, University of Wrocław, Wrocław, Poland
gawry@cs.uni.wroc.pl

Abstract. Two basic problems considered in algorithms on strings are pattern matching and text indexing. An instance of the pattern matching problem comprises two strings, usually called the pattern and the text, and the goal is to locate an occurrence of the former in the latter. In the more general text indexing problem we wish to preprocess a long text for multiple queries with (probably short) patterns. Efficient solutions to both problems have been already designed in the 1970s, and by now we know several linear-time pattern matching algorithms as well as linear-space indexes answering queries in linear (or almost linear) time that are simple enough to be taught in a basic algorithms course. However, with the ever-increasing amount of data being generated and stored, it is not clear if the seemingly optimal linear complexity is actually good enough. This is because linear in the length of the pattern or the text might be larger than the size of its description, for example when we are working with strings over a small alphabet and are able to store multiple characters in a single machine word. The difference might be even more dramatic if we store the data in a compressed form. For some compression schemes, such as the Lempel-Ziv family of compression algorithms, it may as well be the case that the length of the original string is exponential in the size of its compressed representation. In such a case, we would like to design a solution running in time and space proportional to the size of the compressed representation, or at least close to it. This brings the question of developing compressed pattern matching algorithms and designing compressed text indexes.

I will survey the landscape of searching and indexing compressed text, focusing on the Lempel-Ziv family of compression algorithms and the related grammar-based compression. For compressed pattern matching, I will mostly assume that only the text is compressed, but will also briefly describe the recent progress on the more general case of fully compressed pattern matching, where both the text and the pattern are compressed, and on the approximate version of the problem, in which we seek fragments of the text within small Hamming or edit distance to the pattern. For compressed text indexing, I will discuss the known trade-offs between the size of the structure and the query time and highlight the remaining open questions.

Pattern Discovery in Biological Sequences

Esko Ukkonen 

Department of Computer Science, University of Helsinki,
P. O. Box 68 (Gustaf Hällströmin katu 2b), 00014 Helsinki, Finland
`esko.ukkonen@helsinki.fi`

Abstract. Sequence motifs are patterns of symbols that have recurrent occurrences in biological sequences such as DNA and are presumed to have biological function. Modeling and identification of regulatory DNA motifs such as the binding sites of so-called transcription factors (TFs) is in the core of the attempts to understand gene regulation and functioning of the genome as a whole. Transcription factors are proteins that may bind to DNA, close to transcription start site of a gene. Binding activates or inhibits the transcription machinery (expression) of the associated gene. As the regulated gene may itself be a transcription factor, such pairwise regulatory relation between genes induces a genome-wide network model for gene regulation.

The possible binding sites of a transcription factor T are short DNA segments (DNA words). Different sites of T are close variants of an underlying consensus word specific to T . As for most transcription factors an accurate biophysical modeling of this variation is currently infeasible, simplified combinatorial and probabilistic models of binding motifs are used. The parameters of the models are learned from training DNA sequences that contain plenty of instances of the motif but their exact location within the sequences may not be known a priori.

Applying concepts of formal languages and automata, motifs are modeled with words in generalized alphabets and with other regular-expression-like structures representing the language of possible words of the motif. Such motifs can be extracted from training data using string processing algorithms that find repetitions in sequences.

On probabilistic side, sequence motifs can be modeled with inhomogeneous Markov chains of order 0 or higher and also with more general Markov models. Markov chains of order 0, usually called Position Weight Matrices (PWMs) and visualized with so-called sequence logos, is the motif class commonly used in motif databases. PWM assumes that the motif positions are independent. For some TFs this is too weak, and then Markov models of order higher than 0 capable of representing dependencies between two or more positions suit better. Given training DNA sequences that contain occurrences of motif instances proportionally to the target distribution, machine learning methods can estimate the distribution by learning a probabilistic model that fits best the data.

The talk surveys representations and corresponding discovery algorithms for transcription factor binding motifs. We will discuss suffix-tree based methods for discovery of combinatorial models as well as expectation maximization (EM) algorithm based learning of probabilistic models. We consider basic motifs for single factors (monomers) as well as composite motifs for pairs of factors (dimers) and for chains of factors. Such chains model regulatory modules

that are built of clusters of several factors making together a regulatory complex. Regulatory modules can be discovered from alignments of genomes of related species. Alignment-based method is possible as regulatory modules are conserved in evolution.

References

1. Bailey, T.L., Elkan, C.: The value of prior knowledge in discovering motifs with MEME. In: Proceedings of Third International Conference on Intelligent Systems for Molecular Biology, pp. 21–29. AAAI Press (1995)
2. Jolma, A., Yan, J., Whittington, T., Toivonen, J., et al.: DNA-binding specificities of human transcription factors. *Cell*, **152**(1–2), 327–339 (2013)
3. Khan, A., et al.: JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018)
4. Marsan, L., Sagot, M.: Algorithms for extracting structured motifs using a suffix tree with application to promoter and regulatory site consensus identification. *J. Comput. Biol.* **7**, 345–360 (2000)
5. Omidi, S., Zavolan, M., Pachkov, M., Breda, J., Berger, S., van Nimwegen, E.: Automated incorporation of pairwise dependency in transcription factor binding site prediction using dinucleotide weight tensors. *PLoS Comput. Biol.* **13**(7), e1005176 (2017)
6. Palin, K., Taipale, J., Ukkonen, E.: Locating potential enhancer elements by comparative genomics using the EEL software. *Nat. Protoc.* **1**(1), 368 (2006)
7. Pavesi, G., Mauri, G., Pesole, G.: An algorithm for finding signals of unknown length in DNA sequences. *Bioinf.* **17**(Suppl. 1), S207–S214 (2001)
8. Ruan, S., Stormo, G.D.: Inherent limitations of probabilistic models for protein-DNA binding specificity. *PLoS Comput. Biol.* **13**(7), e1005638 (2017)
9. Stormo, G.D., Schneider, T.D., Gold, L., Ehrenfeucht, A.: Use of the ‘Perceptron’ algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.* **10**, 2997–3011 (1982)
10. Siebert, M., Söding, J.: Bayesian markov models consistently outperform pwms at predicting motifs in nucleotide sequences. *Nucleic Acids Res.* **44**(13), 6055–6069 (2016)
11. Toivonen, J., Kivioja, T., Jolma, A., Yin, Y., Taipale, J., Ukkonen, E.: Modular discovery of monomeric and dimeric transcription factor binding motifs for large data sets. *Nucleic Acids Res.* **46**(8), e44 (2018)

Contents

Invited Papers

Modern Aspects of Complexity Within Formal Languages	3
<i>Henning Fernau</i>	
Observation and Interaction: Invited Paper	31
<i>Edward A. Lee</i>	
From Words to Graphs, and Back	43
<i>Vadim Lozin</i>	

Automata

An Oracle Hierarchy for Small One-Way Finite Automata	57
<i>M. Anabtawi, S. Hassan, C. Kapoutsis, and M. Zakzok</i>	
Orbits of Abelian Automaton Groups	70
<i>Tim Becker and Klaus Sutner</i>	
Bounded Automata Groups are co-ETOL	82
<i>Alex Bishop and Murray Elder</i>	
Decidability of Sensitivity and Equicontinuity for Linear Higher-Order Cellular Automata	95
<i>Alberto Dennunzio, Enrico Formenti, Luca Manzoni, Luciano Margara, and Antonio E. Porreca</i>	
On Varieties of Ordered Automata	108
<i>Ondřej Klíma and Libor Polák</i>	
Automata over Infinite Sequences of Reals	121
<i>Klaus Meer and Ameen Naif</i>	
Nonuniform Families of Polynomial-Size Quantum Finite Automata and Quantum Logarithmic-Space Computation with Polynomial-Size Advice	134
<i>Tomoyuki Yamakami</i>	
Equivalence Checking of Prefix-Free Transducers and Deterministic Two-Tape Automata	146
<i>Vladimir A. Zakharov</i>	

Efficient Symmetry Breaking for SAT-Based Minimum DFA Inference 159
*Ilya Zakirzyanov, Antonio Morgado, Alexey Ignatiev,
 Vladimir Ulyantsev, and Joao Marques-Silva*

Complexity

Closure and Nonclosure Properties of the Compressible
 and Rankable Sets. 177
*Jackson Abascal, Lane A. Hemaspaandra, Shir Maimon,
 and Daniel Rubery*

The Range of State Complexities of Languages Resulting
 from the Cut Operation 190
Markus Holzer and Michal Hospodár

State Complexity of Pseudocatenation 203
Lila Kari and Timothy Ng

Complexity of Regex Crosswords 215
Stephen Fenner and Daniel Padé

Grammars

Generalized Predictive Shift-Reduce Parsing for Hyperedge
 Replacement Graph Grammars 233
Berthold Hoffmann and Mark Minas

Transformation of Petri Nets into Context-Dependent Fusion Grammars 246
Hans-Jörg Kreowski, Sabine Kuske, and Aaron Lye

Generalized Register Context-Free Grammars 259
Ryoma Senda, Yoshiaki Takata, and Hiroyuki Seki

Languages

Logic and Rational Languages of Scattered and Countable
 Series-Parallel Posets. 275
Amazigh Amrane and Nicolas Bedon

Toroidal Codes and Conjugate Pictures 288
Marcella Anselmo, Maria Madonia, and Carla Selmi

Geometrical Closure of Binary $V_{3/2}$ Languages. 302
Jean-Philippe Dubernard, Giovanna Guaiana, and Ludovic Mignot

Deterministic Biautomata and Subclasses of Deterministic
 Linear Languages 315
Galina Jirásková and Ondřej Klíma

Learning Unions of k -Testable Languages 328
Alexis Linard, Colin de la Higuera, and Frits Vaandrager

Graphs, Trees and Rewriting

Regular Matching and Inclusion on Compressed Tree Patterns
 with Context Variables 343
Iovka Boneva, Joachim Niehren, and Momar Sakho

Rule-Based Unification in Combined Theories and the Finite
 Variant Property 356
*Ajay K. Eeralla, Serdar Erbatur, Andrew M. Marshall,
 and Christophe Ringeissen*

Extensions of the Caucal Hierarchy? 368
Paweł Parys

Tight Bounds on the Minimum Size of a Dynamic Monopoly 381
Ahad N. Zehmakan

Words and Codes

Recurrence in Multidimensional Words 397
Émilie Charlier, Svetlana Puzynina, and Élise Vandomme

A Note with Computer Exploration on the Triangle Conjecture. 409
Christophe Cordero

Efficient Representation and Counting of Antipower Factors in Words 421
*Tomasz Kociumaka, Jakub Radoszewski, Wojciech Rytter,
 Juliusz Straszyński, Tomasz Waleń, and Wiktor Zuba*

On the Maximum Number of Distinct Palindromic Sub-arrays 434
Kalpana Mahalingam and Palak Pandoh

Syntactic View of Sigma-Tau Generation of Permutations 447
Wojciech Rytter and Wiktor Zuba

Palindromic Subsequences in Finite Words. 460
Clemens Müllner and Andrew Ryzhikov

Author Index 469