200 Problems on Languages, Automata, and Computation

Formal languages and automata have long been fundamental to theoretical computer science, but students often struggle to understand these concepts in the abstract. This book provides a rich source of compelling exercises designed to help students grasp the subject intuitively through practice. The text guides the reader through important topics such as finite automata, regular expressions, pushdown automata, grammars, and Turing machines via a series of problems of increasing difficulty. Problems are organised by topic, many with multiple follow-ups, and each section begins with a short recap of the basic notions necessary to make progress. Complete solutions are given for all exercises, making the book well suited for self-study as well as for use as a course supplement. Developed over the course of the editors' two decades of experience teaching the acclaimed Automata, Formal Languages, and Computation course at the University of Warsaw, it is an ideal resource for students who want to understand the subject deeply and for instructors who want to bring new ideas to their teaching.

FILIP MURLAK is Associate Professor at the University of Warsaw, where he lectures on automata, complexity, logic, and databases. His publications include the books *Relational and XML Data Exchange* (2010) and *Foundations of Data Exchange* (2014).

DAMIAN NIWIŃSKI is Professor at the University of Warsaw, where he lectures on automata theory, languages and computations, complexity, algorithmic game theory, and information theory. His publications include the monograph *Rudiments of Mu-calculus* (with A. Arnold, 2001).

WOJCIECH RYTTER is Professor at the University of Warsaw and a member of the Academia Europaea. He is the author of a large number of publications on automata, formal languages, parallel computing, and text algorithms. He is a co-author of several books, including *Efficient Parallel Algorithms* (1987), *Text Algorithms* (1994) and *125 Problems in Text Algorithms* (2021).

200 Problems on Languages, Automata, and Computation

Edited by

FILIP MURLAK
University of Warsaw

DAMIAN NIWIŃSKI
University of Warsaw

WOJCIECH RYTTER
University of Warsaw

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,

New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316513460
DOI: 10.1017/9781009072632

First published 2023

A catalogue record for this publication is available from the British Library

ISBN 978-1-316-51346-0 Hardback
ISBN 978-1-009-07378-3 Paperback

Contents

Preface *page* vii
List of Notation ix

PART I PROBLEMS

1 Words, Numbers, Graphs 3

2 Regular Languages 7
2.1 Regular Expressions and Finite Automata 7
2.2 Proving Non-Regularity 10
2.3 Closure Properties 13
2.4 Minimal Automata 16
2.5 Variants of Finite Automata 18
2.6 Combinatorics of Finite Automata 20
2.7 Algorithms on Automata 23
2.8 Stringology 25

3 Context-Free Languages 26
3.1 Context-Free Grammars 26
3.2 Context-Free or Not? 30
3.3 Pushdown Automata 34
3.4 Properties of Context-Free Languages 38

4 Theory of Computation 41
4.1 Turing Machines 41
4.2 Computability and Undecidability 45
4.3 Chomsky Hierarchy 48
4.4 Computational Complexity 49

PART II SOLUTIONS

5 Words, Numbers, Graphs 55

6 Regular Languages 65

7 Context-Free Languages 129

8 Theory of Computation 193

Further Reading 250

Index 251

Preface

This book contains problems collected over more than two decades by Damian Niwiński and Wojciech Rytter for the course *Languages, Automata, and Computation* that they alternately taught at the University of Warsaw. Coming from different scientific backgrounds – Wojciech specializing in algorithms and Damian in logic – they shared the idea that the problems should help students develop their skills through the joy of creative thinking. Over the years the collection was gradually expanded and circulated informally in countless versions. Damian and Wojciech always felt that it should be turned into a proper problem book, but this was impossible without written solutions. These were eventually provided by a large group of people, related in various ways to the automata group at the University of Warsaw, under the coordination of Filip Murlak. And thus the long-expected book was born.

A course on formal languages, automata, and computation is present in almost every computer science curriculum around the world, usually in a standardized form based on one of the several classic textbooks. While these textbooks usually complement the expository material with a small set of exercises, sometimes with brief solutions or hints, the present book reverts these proportions: it offers rich problem-solving materials for the course, appropriate both for in-class use and for self-study, accompanied by a minimalistic exposition of the necessary notions.

The book consists of two parts. Part I *Problems* (by D. Niwiński and W. Rytter), contains 200 problems, ranging from easy, marked with ☆, through intermediate (unmarked), to hard and very hard, marked with ★ and ★★, respectively. Some of the harder problems are well-known textbook theorems; we include them because, during the years of teaching the automata course, we have seen that under appropriate guidance students can rediscover a lot of classical material on their own, gaining deeper understanding and a greater sense of accomplishment. Problem statements are interleaved with concise definitions of key notions; for a deeper and broader background we refer to the work listed in the Further Reading section at the end of the book.

Part II, *Solutions* (collectively by 19 authors), contains full solutions to all 200 problems. The focus of the book is on creativity, rather than on practising specific methods, but many solutions build upon previous problems. Some initially simple ideas are further developed in a sequence of problems, the very order of problem statements guiding the students towards powerful methods.

The book has an index. It can help locate the definition of a key notion, like *off-line Turing machine*, and find problems related to a specific topic, like *one-letter alphabet*, or using a specific solution idea, like the *fooling method*.

Notation

$f : A \rightharpoonup B$ — f is a partial function from A to B.
$\#_u(w)$ — the number of (possibly overlapping) occurrences of the word u as a subword of the word w.
$|w|$ — the length of the word w.
$w[i]$ — the ith letter of the word w (counted from 1).
$w[i..j]$ — the infix of the word w from position i to position j, inclusive.
w^{R} — the reverse of the word w, or w written backwards.
$KL^{-1} = \{u : \exists v \in L.\, uv \in K\}$ — the right quotient of K by L.
$L^{-1}K = \{v : \exists u \in L.\, uv \in K\}$ — the left quotient of K by L.
$r \cdot s = \{(x,z) : \exists y.\, (x,y) \in r \wedge (y,z) \in s\}$ — the left composition of the binary relations r and s.
r^* — the reflective–transitive closure of the binary relation r.
$[w]_2$ — the numerical value of the binary sequence w; e.g., $[011]_2 = 3$.
$\mathrm{bin}(n)$ — the binary representation of $n \in \mathbb{N}$, without leading zeros.
\$ — a distinguished letter used as a separator.