Handbook of Automata Theory

Edited by Jean-Éric Pin

Version of 18th May, 2015

Contents

Foundations

1 J.-É. Pin **3**

2 J. Sakarovitch **39**

3 T. Harju and J. Karhumäki **75**

4 M. Droste, D. Kuske **107**

5 S. Lombardy, J. Mairesse **143**

6 T. Wilke **179**

7 C. Löding, W. Thomas **219**

8 C. Löding **249**

9 S. Crespi-Reghizzi, D. Giammarresi, V. Lonati **281**

Complexity issues

10 J. Berstel, L. Boasson, O. Carton, I. Fagnot **317**

11 H. Björklund, J. Högberg, W. Martens **351**

12 H. Gruber, M. Holzer, M. Kutrib **383**

13 H. Gruber, J. Lee, J. Shallit **415**

14 M. Koucky **445**

15 J. Kari, M. Volkov **475**

Algebraic and topological theory of automata

16 H. Straubing, P. Weil **509**

17 J. Almeida, A. Costa **551**

18 T. Colcombet **585**

19 J. Duparc **623**

20 Z. Ésik **655**

21 M. Kunc, A. Okhotin **689**

22 M. Bojańczyk **723**

Automata in Mathematics

23 L. Bartholdi, P. Silva 759

24 L. Bartholdi, P. Silva 787

25 B. Adamczewski, J. Bell 825

26 F. Durand, M. Rigo 857

27 M.P. Béal, J. Berstel, S. Eilers, D. Perrin 895

28 S. Rubin 939

29 W. Gelade, T. Schwentick 971

30 J. Karhumäki, J. Kari 989

Selected applications

31 D. Kuske, A. Muscholl 1029

32 B. Boigelot 1067

33 I. Walukiewicz 1095

34 P. Bouyer 1129

35 A. Carayol, O. Serre 1161

36 K. Etessami **1203**

37 M.-J. Nederhof, G. Satta **1239**

38 J. Esparza, O. Kupferman, M. Vardi, P. Wolper **1269**

39 A. Ambainis, A. Yakaryılmaz **1307**

Foundations

CHAPTER 1

Finite automata (J.-É. Pin)

Abstract

The aim of this chapter is to introduce several fundamental notions and results of automata theory. We first define finite automata and recognizable languages. Then we explore the basic properties of finite automata (deterministic, complete, accessible, standard, etc.) and we describe the standard opertions on languages (Boolean operations, concatenation product, star, quotiente, etc.). Then we define the minimal automaton of a language and its algebraic counterpart, the syntactic monoid and its ordered version. We prove Kleene's theorem using Glushkov's algorithm in one direction and linear equations for the other direction. In the last section we briefly define rational and recognizable subsets of arbitrary monoids.

CHAPTER 2

Automata and rational expressions (J. Sakarovitch)

Abstract

To do

CHAPTER 3

Finite transducers and rational transductions (T. Harju and J. Karhumäki)

Abstract

Finite transducers are generalized finite automata with output. Their history in the modern setting goes back to the end of 1950's when Rabin and Scott published their study on the topic.

One of the basic results of the chapter characterizes rational transductions realized by finite transducers in terms of compositions of morphisms. This result arises from Nivat's theorem, which was extended in the early 1980s to purely morphic representations by Culik, Fich and Salomaa and later improved by several authors.

We also consider an impressive application of transductions due to J. Kari on aperiodic tilings of the plane. Also, transductions are used to solve the isomorphism problem of F-semigroups. Finally, we study decision problems for transducers and their transductions by establishing several rather simply formulated natural undecidable problems. In this context especially problems related to finite substitutions are considered.

Undecidability of the equivalence problem for finite transductions, due to L. Lisovik and O. Ibarra, is stated and sharpened in many steps to a "final" deep form, due to M. Kunc.

CHAPTER 4

Weighted automata (M. Droste, D. Kuske)

Abstract

Weighted automata are classical finite automata in which the transitions carry weights. These weights may model quantitative properties like the amount of resources needed for executing a transition or the probability or reliability of its successful execution. Using weighted automata, we may also count the number of successful paths labeled by a given word.

As an introduction into this field, we present selected classical and recent results concentrating on the expressive power of weighted automata.

CHAPTER 5

Max-Plus automata (S. Lombardy, J. Mairesse)

CHAPTER 6

ω-Automata (T. Wilke)

CHAPTER 7

Automata on finite trees (C. Löding, W. Thomas)

Abstract

This survey starts with a sketch of the basic theory of finite tree automata over finite trees, both regarding their expressiveness and algorithmic properties. The subsequent sections address the topics of ground-tree rewriting systems, tree-walking automata, and automata over unranked trees as they arise in XML-documents. Finally, known results on the classification of regular tree languages are summarized.

CHAPTER 8

Automata on infinite trees (C. Löding)

Abstract

This article presents the basic theory of automata on infinite trees. We introduce the standard model of nondeterministic automata and the generalization to alternating automata. The central part of the theory considers the closure properties of tree languages definable by such automata, the equivalence of the different models, and their algorithmic properties. We show how these results are obtained and how games of infinite duration can be used as a powerful tool to obtain transparent proofs and algorithms. We also present some applications of tree automata in logic, in particular decidability results and characterizations for monadic second-order logic over infinite trees, and algorithms for the modal μ-calculus.

CHAPTER 9

Two dimensional models (S. Crespi-Reghizzi, D. Giammarresi, V. Lonati)

Abstract

The survey presents the basic concepts for picture definition using several formal models: 2D regular expressions, local languages, tile recognizable (REC) languages, and various grammar types. The REC family and its properties are studied in some detail, including closure properties, determinism, and necessary condition for tiling recognizability. For grammars, the models considered are: context-free Kolam grammars and related symbolic grammars, grammars with grids as righthand sides, grammars with isometric rules based on tiles, including the tiling and regional grammars. Some grammar-defined families are compared with the REC family. A discussion of the membership problem for some picture language families and a polynomial time parsing algorithm for the regional grammars conclude the survey, which also includes typical language examples and historical notes.

Complexity issues

CHAPTER 10

Minimization of automata (J. Berstel, L. Boasson, O. Carton, I. Fagnot)

Abstract

This chapter is concerned with the design and analysis of algorithms for minimizing finite automata. Getting a minimal automaton is a fundamental issue in the use and implementation of finite automata tools in frameworks like text processing, image analysis, linguistic computer science, and many other applications.

Although minimization algorithms are rather old, there were recently some new developments which are explained or sketched in this chapter.

There are two main families of minimization algorithms. The first are by a sequence of refinements of a partition of the set of states, the second by a sequence of fusions or merges of states. Hopcroft's and Moore's algorithms belong to the first family. The linear-time minimization of acyclic automata of Revuz belongs to the second family.

We describe, prove and analyze Moore's and Hopcroft's algorithms. One of our studies is the comparison of these algorithms. It appears that they quite different both in behavior and in complexity. In particular, we show that it is not possible to simulate the computations of one of the algorithm by the other. We also state the results known about average complexity of both algorithms.

We describe the minimization algorithm by fusion for acyclic automata, and for a more general class of automata, namely those recognizing regular languages having polynomially bounded number of words.

Finally, we consider briefly incremental algorithms for building minimal automata for finite sets of words. We consider the updating of a minimal automaton when a word is added or removed.

CHAPTER 11

Learning algorithms (H. Björklund, J. Högberg, W. Martens)

Abstract

This chapter is concerned with automatic learning of regular languages. We treat the classical results and algorithms for learning deterministic finite word automata such as learning in the limit, learning through a minimal adequate teacher, and probably approximately correct learning. We discuss extensions to learning tree automata, nondeterministic automata, and weighted automata. Finally, we discuss examples of where these learning approaches are used in practice.

CHAPTER 12

Descriptional complexity of regular languages (H. Gruber, M. Holzer, M. Kutrib)

Abstract

We give a guided tour to questions regarding the descriptional complexity of regular languages. There are many different possibilities for representing regular languages. We will focus our attention on two classical models: finite automata and regular expressions. Each model has its own advantages, so we often want to switch among different representations. Another issue is to modify given representations of a regular language in the sense of performing language operations. Both in the case of switching among representations and of performing language operations, it turns out that the computational cost of the respective task is chiefly governed by the size of the resulting description. We consider in particular the descriptional complexity of determinization, conversion between finite automata and regular expressions and *vice versa*, and of performing basic language operations on finite automata, as well as on regular expressions. To obtain such results, we often crucially rely on suitable lower bound techniques. These are needed for proving that a certain size of description is required. We thus also present the most relevant lower bound techniques for each type of representation. Finally, we give a brief outlook on other recent results regarding the descriptional complexity of various representations of regular languages.

CHAPTER 13

Enumerating Regular Expressions and Their Languages (H. Gruber, J. Lee, J. Shallit)

Abstract

In this chapter we discuss the problem of enumerating distinct regular expressions by size and the regular languages they represent. We discuss various notions of the size of a regular expression that appear in the literature and their advantages and disadvantages. We consider a formal definition of regular expressions using a context-free grammar.

We then show how to enumerate strings generated by an unambiguous context-free grammar using the Chomsky-Schützenberger theorem. This theorem allows one to construct an algebraic equation whose power series expansion provides the enumeration. Classical tools from complex analysis, such as singularity analysis, can then be used to determine the asymptotic behavior of the enumeration.

We use these algebraic and analytic methods to obtain asymptotic estimates on the number of regular expressions of size n. A single regular language can often be described by several regular expressions, and we estimate the number of distinct languages denoted by regular expressions of size n. We also give asymptotic estimates for these quantities. For the first few values, we provide exact enumeration results.

CHAPTER 14

Circuit complexity (M. Koucky)

Abstract

Finite automata and boolean circuits are two different, well studied models of computation. Finite automata are inherently sequential whereas boolean circuits are inherently parallel computational devices. In this chapter we survey what is known about their relationship. We exhibit a close connection between various natural classes of regular languages and well studied circuit classes. This connection is of fundamental interest and connects well studied problems from both of these separate areas.

CHAPTER 15

Cerný's conjecture and the Road coloring problem (J. Kari, M. Volkov)

Algebraic and topological theory of automata

CHAPTER 16

Varieties (H. Straubing, P. Weil)

CHAPTER 17

Profinite Topologies (J. Almeida, A. Costa)

CHAPTER 18

The factorisation Forest Theorem (T. Colcombet)

Abstract

This chapter is devoted to the presentation of the factorisation forest theorem, a deep result due to Simon, which provides advanced Ramsey-like arguments in the context of algebra, automata, and logic. We present several proofs and several variants the result, as well as applications.

CHAPTER 19

Wadge-Wagner hierarchies (J. Duparc)

CHAPTER 20

Equational theories for automata (Z. Ésik)

Abstract

This chapter is devoted to an axiomatic treatment of finite automata that is mainly based on equational axioms. We identify those properties of the rational operations that underlie Kleene's theorem and give an overview of completeness results.

CHAPTER 21

Language equations (M. Kunc, A. Okhotin)

Abstract

Equations with formal languages as unknowns naturally appear whenever sets of words are being used. In particular, they are fundamental for the theory of formal grammars, with systems of equations of the form $X_i = \varphi(X_1, \ldots, X_n)$ representing the inductive nature of the context-free grammars and their natural variants. Some variants of these systems naturally represent finite automata and a basic class of cellular automata. Equations of the general form are notable for their computational completeness, with universal computation encoded in extremely simple examples. The chapter provides a survey of the known results on language equations, classifying them according to the methods of research, and comparing similar properties of different families.

CHAPTER 22

Algebras for trees (M. Bojańczyk)

Abstract

This chapter presents several algebraic approaches to tree languages. The idea is to design a notion for trees that resembles semigroups or monoids for words. The focus is on the connection between the structure of an algebra recognizing a tree language, and the kind of logic needed to define the tree language. Four algebraic approaches are described in this chapter: trees as terms of universal algebra, preclones, forest algebra, and seminearrings. Each approach is illustrated with an application to logic on trees.

Automata in Mathematics

CHAPTER 23

Rational subsets of groups (L. Bartholdi, P. Silva)

Abstract

This chapter is devoted to the study of rational subsets of groups, with particular emphasis on the automata-theoretic approach to finitely generated subgroups of free groups. Indeed, Stallings' construction, associating a finite inverse automaton with every such subgroup, inaugurated a complete rewriting of free group algorithmics, with connections to other fields such as topology or dynamics.

Another important vector in the chapter is the fundamental Benois' Theorem, characterizing rational subsets of free groups. The theorem and its consequences really explain why language theory can be successfully applied to the study of free groups. Rational subsets of (free) groups can play a major role in proving statements (*a priori* unrelated to the notion of rationality) by induction. The chapter also includes related results for more general classes of groups, such as virtually free groups or graph groups.

CHAPTER 24

Groups defined by automata (L. Bartholdi, P. Silva)

Abstract

Finite automata have been used effectively in recent years to define infinite groups. The two main lines of research have as their most representative objects the class of automatic groups (including word-hyperbolic groups as a particular case) and automata groups (singled out among the more general self-similar groups).

The first approach implements in the language of automata some tight constraints on the geometry of the group's Cayley graph, building strange, beautiful bridges between far-off domains. Automata are used to define a normal form for group elements, and to monitor the fundamental group operations.

The second approach features groups acting in a finitely constrained manner on a regular rooted tree. Automata define sequential permutations of the tree, and represent the group elements themselves. The choice of particular classes of automata has often provided groups with exotic behaviour which have revolutioned our perception of infinite finitely generated groups.

CHAPTER 25

Automata in number theory (B. Adamczewski, J. Bell)

Abstract

In this chapter, we survey some of the connections between the theory of automatic sequences/sets and number theory. Several substantial advances have recently been made in this area and we give an overview of some of these new results. This includes discussions about prime numbers, the decimal expansion of algebraic numbers, the search for an analogue of the Skolem-Mahler-Lech theorem in positive characteristic and the description of an algebraic closure of the field $\mathbb{F}_p(t)$.

CHAPTER 26

On Cobham's Theorem (F. Durand, M. Rigo)

Abstract

Let $k \geq 2$ be an integer. A set X of integers is k-recognizable if the language of k-ary representations of the elements in X is accepted by a finite automaton. The celebrated theorem of Cobham from 1969 states that if a set of integers is both k-recognizable and ℓ-recognizable, then it is a finite union of arithmetic progressions. We present several extensions of this result to non-standard numeration systems, we describe the relationships with substitutive and automatic words and list Cobham-type results in various contexts.

CHAPTER 27

Symbolic Dynamics (M.P. Béal, J. Berstel, S. Eilers, D. Perrin)

Abstract

This chapter presents some of the links between automata theory and symbolic dynamics. The emphasis is on two particular points. The first one is the interplay between some particular classes of automata, such as local automata and results on embeddings of shifts of finite type. The second one is the connection between syntactic semigroups and the classification of sofic shifts up to conjugacy.

CHAPTER 28

Automatic structures (S. Rubin)

Abstract

In order to compute with infinite mathematical structures one needs, typically at least, a finitary description/presentation of the structure. Automata and their equivalent logical formalisms are natural choices. These give rise to the automatic structures whose salient feature is that every definable relation is recognised by an automaton. This chapter focuses on foundational questions: What are equivalent ways of defining automatic structures? Under what operations are automatic structures closed? How does one tell if a given structure is automatic? What does it mean for two presentations of the same structure to be different?

CHAPTER 29

Automata and finite model theory (W. Gelade, T. Schwentick)

CHAPTER 30

Finite automata, image manipulation and automatic real functions (J. Karhumäki, J. Kari)

Selected applications

CHAPTER 31

Communicating automata (D. Kuske, A. Muscholl)

CHAPTER 32

Symbolic methods and automata (B. Boigelot)

Abstract

This chapter discusses the use of finite automata as data structures for representing symbolically sets of values. This consists in using an encoding function that maps the elements of a domain onto words, and of representing a set of values by an automaton that accepts their encodings. This method is applied to two domains that are relevant for applications: those of vectors with integer and with real components. For each of these domains, algorithms are proposed for constructing elementary sets, for combining them, and for deciding arithmetic formulas using automata representations. The advantages of using finite-state representations is that these representations are expressive enough for many applications, naturally closed under a large class of operations, and that, with some forms of automata, they admit an easily computable canonical form.

CHAPTER 33

Synthesis with finite automata (I. Walukiewicz)

Abstract

The synthesis problem asks to automatically construct a device from a specification. This chapter presents some classical formulations of the problem when devices are finite state machines that read and output sequences of bits.

The starting point is a standard formulation proposed by Church more than a half a century ago. It asks to construct one input/output device subject to a specification given in monadic second-order logic. A richer formulation of Ramadge and Wonham introduces a notion of a plant and a controller who is supposed to limit behaviours of the plant so that a specification is satisfied. The final part of the chapter is devoted to a distributed synthesis problem as formulated by Pnueli and Rosner. The objective is to construct a distributed system subject to communication and visibility constraints. The chapter describes principal constructions in these three frameworks with an emphasis on decidability and undecidability results.

CHAPTER 34

Timed Automata (P. Bouyer)

Abstract

This chapter is concerned with the model of timed automata. Timed automata are an extension of finite automata, which allows to model systems with real-time constraints. We present this model, its operational semantics, and its language-based semantics.

We then argue why the emptiness problem is a fundamental problem for verification purposes (the initial application domain of timed automata). The core of this chapter is the construction of the region automaton, which is a finite automaton representing the abstract behaviour of the timed automaton. We then show that the timed language recognized by the original timed automaton is empty if and only if the (untimed) language recognized by the region automaton is empty. This shows in particular that the emptiness problem is decidable in timed automata. We also discuss complexity issues.

We then study other language-based problems, like complementation, universality, inclusion and determinism. We unfortunately show several negative results: timed automata are not closed under complementation, and the universality problem is undecidable. This shows that timed automata are difficult to manipulate and should be handled with care for verification purposes.

We conclude with some current directions of research.

CHAPTER 35

Higher-order recursion schemes and their automata models (A. Carayol, O. Serre)

CHAPTER 36

Analysis of probabilistic processes and automata theory (K. Etessami)

Abstract

This chapter surveys some basic algorithms for analyzing Markov chains (MCs) and Markov decision processes (MDPs), and discusses their computational complexity. We focus on discrete-time processes, and we consider both finite-state models as well as countably infinite-state models that are finitely-presented. The analyses we will primarily focus on are hitting (reachability) probabilities and ω-regular model checking, but we will also discuss various reward-based analyses.

Although it may not be evident at first, there are fruitful connections between automata theory and stochastic processes. Firstly, and not surprisingly, ω-automata play a naturally important role for specifying ω-regular properties of sample paths (trajectories) of stochastic processes. Computing the probability of the event that a random sample path satisfies a given ω-regular property constitutes the (linear-time) model checking problem for probabilistic systems.

Secondly, it turns out that there are close relationships between classic infinite-state automata-theoretic models and classic denumerably infinite-state stochastic processes, even though these models were developed independently in separate mathematical communities. Roughly speaking, some classic stochastic processes share their underlying state transition systems with corresponding classic automata-theoretic models. Furthermore, exploiting these connections to automata theory is fruitful for the algorithmic analysis of such stochastic processes, and for their controlled MDP extensions. This holds even when the analyses are much simpler than model checking, such as computing (optimal) hitting probabilities.

A number of important infinite-state stochastic models connected with automata theory can be captured as (restricted fragments of) *recursive Markov chains* and *recursive Markov decision processes*, which are obtained by adding a natural recursion feature to finite-state MCs and MDPs. Key computational problems for analyzing classes of re-

cursive MCs and MDPs can be reduced to computing the *least fixed point* (LFP) solution of corresponding classes of *monotone* systems of nonlinear equations. The complexity of computing the LFP for such equations is a intriguing problem, with connections to several areas of research in theoretical computer science.

CHAPTER 37

Natural language processing (M.-J. Nederhof, G. Satta)

Abstract

This chapter addresses the use of weighted push-down automata for parsing natural languages. One feature that sets natural languages apart from programming languages is ambiguity, which means that the relevant push-down automata are nondeterministic. We discuss methods of tabulation that can be used to simulate nondeterministic automata in polynomial time, and we show how a number of well-known parsing algorithms can be derived through these generic methods.

CHAPTER 38

Verification (J. Esparza, O. Kupferman, M. Vardi, P. Wolper)

CHAPTER 39

Automata and quantum computing (A. Ambainis, A. Yakaryılmaz)

Abstract

Quantum computing is a new model of computation, based on quantum physics. Quantum computers can be exponentially faster than conventional computers for problems such as factoring. Besides full-scale quantum computers, more restricted models such as quantum versions of finite automata have been studied. In this paper, we survey various models of quantum finite automata and their properties.