

- [3] H. Comon, M. Dauchet, R. Gilleron, F. J. D. Lugiez, S. Tison, and M. Tommasi. Tree automata techniques and applications. Available at: <http://www.grappa.univ-lille3.fr/tata>, 1997. release October, 1st 2002.
- [4] J. Doner. Tree acceptors and some of their applications. *Journal of Computer and System Sciences*, 4:406–451, 1970.
- [5] Y. Gurevich and L. Harrington. Trees, automata, and games. In *Proceedings of the fourteenth annual ACM Symposium on Theory of Computing, San Francisco, California, May 5–7, 1982*, pages 60–65, New York, NY, USA, 1982. ACM Press.
- [6] Y. S. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-theoretic Logics*. Springer Verlag, 1985.
- [7] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.
- [8] D. Perrin and J.-E. Pin. Infinite words. Available at <http://www.liafa.jussieu.fr/~jep/Resumes/InfiniteWords.html>, to appear.
- [9] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. of Amer. Math. Soc.*, 141:1–35, 1969.
- [10] M. O. Rabin. Decidable theories. In J. Barwise, editor, *Handbook of Mathematical Logic*. North-Holland, Amsterdam, 1977.
- [11] J. Rogers. *A Descriptive Approach to Language Theoretic Complexity*. CSLI Publications, 1998.
- [12] C. Stirling. *Modal and Temporal Properties of Processes*. Springer, 2001.
- [13] J. W. Thatcher and J. B. Wright. Generalized finite automata with an application to a decision problem of second order logic. *Mathematical Systems Theory*, 2:57–82, 1968.
- [14] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*. Elsevier, 1990.

Review⁴ of
Automatic sequences: Theory, Applications, Generalizations
Authors: Jean-Paul Allouche and Jeffrey Shallit
Cambridge University Press

Author of Review: Jean Berstel, Institut Gaspard Monge, Université de Marne-la-Vallée

⁴©2004 Jean Berstel

1 Overview

Automatic sequences are sequences of symbols recognized by finite automata. These are the central topic of the book.

Sequences of numbers or words occur in mathematics and in theoretical computer sciences, and also in theoretical physics, and in computational biology. This scattered research on more or less the same object explains widespread publication support, many variations in terminology, and a great variety of results.

Automatic sequences are simple to generate or to recognize, and moreover — and this is the main motivation of their study by number theorists — there are many interesting connections with number theory.

A famous example is the so-called Thue–Morse sequence t that reads, written on the digits 0 and 1 as

$$t = 0110100110010110 \dots$$

It can be defined in various ways. The simplest is by induction: $t_0 = 0$, and $t_{2n} = t_n$, $t_{2n+1} = 1 - t_n$. Another definition is to consider the binary expansion $\text{bin}(n)$ of the integer n . Then $t_n = 1$ iff $\text{bin}(n)$ contains an odd number of 1's, and $t_n = 0$ otherwise. This definition is “automatic” because one can easily conceive a two state automaton that recognizes those binary expansions that have an odd number of 1's. Another way to construct this infinite word t is to observe that it is a fixed point of the morphism μ that maps 0 to 01 and 1 to 10. Thus t can also be constructed as limit of a sequence of finite words, the so-called Thue–Morse blocks X_i defined by $X_0 = 0$ and $X_{i+1} = X_i \bar{X}_i$ where \bar{w} exchanges 0 and 1 in w . There are several other magic ways to obtain t . One basic question of interest to number theorists is to determine the status of the real number with binary expansion $0.0110100110010110 \dots$.

It has been proved that this number is transcendental. As a matter of fact, there is a very close, amazing relationship between automatic sequences and transcendental numbers that is treated in detail in chapters 12 and 13 of the book. Basically, these numbers are either rational or transcendental although no general result of this kind has been proved. Another question is about the distribution of digits, and more generally of blocks, in this sequence. This is closely related to symbolic dynamics (this was the initial goal of Morse), but also to the search of so-called “normal numbers”. These questions also are of interest in theoretical physics. Finally, it is well-known that the Thue–Morse sequence is cube-free (no three adjacent equal blocks) and even overlap-free (no overlapping equal blocks). This was the the domain of interest of Axel Thue. So Thue–Morse sequences are related to combinatorics on words, and to what biologists call “tandem repetitions”.

The objective of the present book is to present automatic sequences and related fields under three point of views: combinatorics on words, especially in relation with automata theory, number theory, especially in relation with algebraic and transcendental numbers, functions, series, and applications to theoretical physics. It is the first time that such a presentation is given in a unified framework, using consistent notation and definition.

Recently, two books appeared which smell similar. First, Lothaire’s second volume entitled “Combinatorics on Words” (Cambridge University Press 2002), and second Pytheas Fogg’s “Substitutions in Dynamics, Arithmetics and Combinatorics” (Lecture Notes in Mathematics 1794, Springer-Verlag). In fact, Lothaire’s book does not cover any relation to number theory with the exception of number systems, and Pytheas Fogg deals more with ergodic theory, spectral analysis and geometric representation of sequences. None of these books contains for instance a proof of

Cobham's theorem. The nature of the book under review is much close to a textbook, and can be used so.

2 Contents

The book consists of 17 chapters and an Appendix. Although it is not structured explicitly in this way, it can be roughly viewed as composed of three parts: Chapters 1–4 are preliminaries, even if they may contain deep results. Chapters 5–13 are on automatic sequences, their generalizations and their applications, and chapters 14–17 are on generalizations and applications.

Chapter 1, **Stringology** introduces finite and infinite words, languages as sets of words, morphisms as a way to transform words. It also contains basic results on equations, such as Fine and Wilf's theorem.

The second part of this chapter contains the proof that the Thue-Morse word is overlap-free, and also detailed results on the overlap-free words that are fixed point of a morphism, and the characterization of overlap-free binary morphisms, that is morphisms that map overlap-free words on overlap-free words. This chapter is quite combinatoric in nature.

Chapter 2, **Number Theory and Algebra** is presented as a smorgasbord of some basic results from algebra and number theory. It presents a proof that π is irrational, and also Liouville's example of an irrational number. Some results on transcendental numbers are stated here without proof. Continued fractions and approximations are considered, with a proof that $\{n\theta\}$ is uniformly distributed for irrational θ . It also contains the three-distance theorem.

A second part deals with definition on algebraic structures, fields, polynomials, formal power series and formal Laurent series, p -adic numbers. Some estimates are grouped at the end of this chapter.

Chapter 3, **Numeration Systems** is on representation of numbers, mainly of integers, in several bases, first the standard integral base, then negative bases, and then Fibonacci representation and representation in complex bases. There are developments on the sum-of-digits function, and on pattern sequences, and on sums of sums of digits.

Fibonacci's number system is presented, and replaced in the context of Ostrowski's numeration system. There is a careful presentation of complex bases, and of perfect complex number systems (those with base $-a \pm i$ for some integer $a \geq 1$).

Chapter 4, **Finite Automata and Other Models of Computation** introduces finite automata, deterministic and nondeterministic, proves Kleene's theorem and Myhill-Nerode's theorem, a pumping lemma. It also considers finite automata with output on states: the result of a path is the output associated to the state (this corresponds to automata with several types of final states).

Transducers are introduced next, and closure of regular languages under transducer mapping is proved, and applications to sets of numbers recognized by finite automata are given.

Next, there is a short section on context-free grammars and languages, pda's and algebraic formal power series and on Turing machines.

Chapter 5, **Automatic Sequences** introduces the fundamental concept of the book. Automatic sequences are defined by automata with output on states. Many examples are given. Two-sided automatic sequences are reduced to one-sided sequences. The chapter contains some basic properties, also closure properties, and examples of non-automatic sequences.

Chapter 6, **Uniform Morphisms and Automatic Sequences** presents the second definition of automatic sequences, by uniform morphisms (Cobham's first theorem). The tower of Hanoi

sequence is an excellent example treated in detail. Paper folding is considered. The kernel is introduced next (that is right regular equivalence on automata). Other closure properties are given, using either the kernel or transducers.

Chapter 7, **Morphic Sequences** presents the general theory of morphic sequences, moving thus from the uniform to the general case. A typical example is the Fibonacci word. A description of morphic words is given, and several closure results are proved. Closure under morphism and under transduction are the two main results.

Chapter 8, **Frequency of Letters** covers usual frequency and logarithmic frequency of letters in sequences. The Perron–Frobenius theory is developed in detail, with full proofs. It is used to show that frequencies of letters in morphic words, when they exist, are algebraic numbers and they are rational for automatic sequences.

Chapter 9, **Characteristic Words** presents a part of the theory of Sturmian words, concentrating on the number theoretic aspects of characteristic words. This is an excellent choice since it allows a streamlined, not very difficult presentation of the relation to continued fractions expansion, and to Ostrowski number systems.

Chapter 10, **Subwords** is on subword complexity, that is the number of words of length n appearing in an infinite sequence, also the gap between two consecutive apparitions of a finite factor, or the first apparition of a factor in a word. The chapter contains Morse and Hedlund’s gap theorem, bound for the subword complexity of morphic words, the characterization of Sturmian words and a theorem of Mignosi relating power-freeness of characteristic sequences to bounded partial quotients in continued fractions. Recurrence, uniform recurrence are also studied.

Chapter 11, **Cobham’s Theorem** is devoted to Cobham’s theorem on the base dependence of automatic sequences. This is one of the main theorems of the theory, and a detailed and careful proof is given.

Chapter 12, **Formal Power Series** contains the third equivalent definition of automatic sequences, namely by algebraic series over the polynomials over a finite field (Christol’s theorem). This chapter studies algebraicity and transcendence of formal power series when the ground field is finite. Using Christol’s result, one can reformulate Cobham’s theorem and state that certain algebraic series must in fact be rational. Christol’s theorem can also be used directly to prove that some series are transcendental.

Chapter 13, **Automatic Real Numbers** is on real numbers associated to infinite sequences, when the symbols in the sequence are considered as digits. Typically, a (automatic) sequence $a_0a_1 \cdots a_n \cdots$ with $0 \leq a_i < b$ is associated with the real number (“automatic” real number) $\sum_{i \geq 0} a_i b^{-i}$. The set of automatic real numbers is shown to be a vector space, but not closed under multiplication. The chapter continues with transcendence results for some automatic real numbers, and for morphic real numbers.

Chapter 14, **Multidimensional Automatic Sequences** extends results on automatic sequences to greater dimensions. It contains again numerous examples, and a discussion of automatic sequences in base $-1 + i$. It ends with the Pascal triangle modulo a non-prime number.

Chapter 15, **Automaticity** measures how much a set or a sequence is regular by approximation. More precisely, given a language L , denote by $A_L(n)$ the minimal number of states of finite automata recognizing a language that have the same words as L up to length n . It appears that $A_L(n) \geq (n+3)/2$ for infinitely many n if L is non-regular. Similarly $A_s^k(n)$ is the minimal number of states of finite automata that recognize k -automatic sequences that have the same n initial symbols that s . It is shown that there exist 2-automatic sequences s such that $A_s^k(n) = \Omega(k\sqrt{n})$ for odd $k > 1$.

Chapter 16, ***k*-Regular Sequences** presents these sequences that are very closely related to rational formal power series in non-commutative variables. Thus, finiteness of automatic sequence transposes to finite dimension for *k*-regular sequences. The presentation does not rely on the theory of non-commutative power series. A great number of closure properties, and of examples, are given.

Chapter 17, **Physics** is a short chapter on relations between physical concepts and infinite sequences. Two cases are treated, namely the Ising model, especially in relation to the Rudin–Shapiro sequence, and the one-dimensional Schrödinger operator, for which some information about the spectrum is presented.

3 Opinion

This book is the very first book that gathers together results from various fields of mathematics and computer science that appeared in a broad variety of journals. It presents the results in a unified manner, making use of a consistent notation and introduces some good unifying terminology, and including most of the material needed for a comprehensive exposition.

The level of presentation is advanced undergraduate to beginning graduate. Although all proofs are given in full details, with the exception of some explicitly mentioned results, some familiarity with mathematical reasoning is required. Familiarity with basics in formal automata is useful, and a first course in number theory is recommended (by me). There are some places where results from number theory are quoted and used which are not in a first course.

This book is useful also as a handbook for self study by the high number of exercises (460) and as basis for research (85 open problems). It is extremely valuable as a reference text for researches and contains a bibliography of more than 1600 citations to the literature, all commented or placed in context in the text. The stuff is presented with great typographical care. It is extremely difficult to find some mistakes. May be, the fact that the paper [Cassaigne and Karhumäki 1995a] appeared in Europ. J. Combinatorics in 1997.

This book can also be used as a basis for a course on formal languages and number theory at the advanced undergraduate or graduate level. I strongly recommend this excellent book to anybody interested in interaction between theoretical computer science and mathematics.

4 Additional Information Available

There is a website associated to the book at:

<http://www.math.uwaterloo.ca/~shallit/asas.html>

(There is a title before the ‘shallit’ .)