Journal of Automata, Languages and Combinatorics **6** (2001) 4, 537–554


# AUTOMATIC COMPLEXITY OF STRINGS [1]

JEFFREY SHALLIT[2] and MING-WEI WANG
*Department of Computer Science, University of Waterloo*
*Waterloo, N2L 3G1 Canada*
*e-mail:* {shallit@graceland, m2wang@math}.uwaterloo.ca

ABSTRACT

We define a new measure of complexity for finite strings, called *automatic complexity* and denoted $A(x)$. Although $A(x)$ is analogous to Kolmogorov-Chaitin complexity, it has the advantage of being computable. We give upper and lower bounds for $A(x)$, and estimate it for some specific strings.

*Keywords:* deterministic finite automata, Kolmogorov complexity, linear diophantine equation, Thue-Morse word.

## 1. Introduction

We are interested in a computable measure of complexity for finite strings $x$ over a finite alphabet, typically $\{0,1\}$. Any such measure should reflect, in some sense, how "complicated" the string $x$ is.

Of course, any such discussion must start with Kolmogorov-Chaitin complexity [11] $C(x)$, which (roughly speaking) measures the complexity of a string $x$ as the size of the shortest pair

$$(T, y) = (\text{Turing machine description, input})$$

such that $T$ on input $y$ outputs $x$. Not only does $C(x)$ measure the complexity of $x$, but also the pair $(T, y)$ can be viewed as the optimal way to compress the string $x$.

However it has three major deficiencies (the first two are equivalent):

1. It is uncomputable! It is known that "$C(x) < n$" is computably enumerable, but "$C(x) \geq n$" is not computably enumerable.
2. There is no effective procedure for finding a compression pair $(T, y)$.
3. $K$ depends somewhat on the particular model of universal Turing machine chosen, and is defined in a machine-independent way only up to an additive constant.

---

[1]Full version of a submission presented at the Second International Workshop on *Descriptional Complexity of Automata, Grammars and Related Structures* held in London, Ontario, Canada, July 27 – 29, 2000.

[2]Supported in part by a grant from NSERC