INTRODUCTION TO AUTOMATA THEORY, LANGUAGES, AND COMPUTATION

JOHN E. HOPCROFT
JEFFREY D. ULLMAN

INTRODUCTION TO AUTOMATA THEORY, LANGUAGES, AND COMPUTATION

JOHN E. HOPCROFT
Cornell University

JEFFREY D. ULLMAN
Princeton University

ADDISON-WESLEY PUBLISHING COMPANY
Reading, Massachusetts · Menlo Park, California
London · Amsterdam · Don Mills, Ontario · Sydney

This book is in the
ADDISON-WESLEY SERIES IN COMPUTER SCIENCE

Michael A. Harrison,
Consulting Editor

Library of Congress Cataloging in Publication Data

Hopcroft, John E., 1939-
Introduction to automata theory, languages, and computation.

Bibliography: p.
Includes index.
1. Machine theory. 2. Formal languages.
3. Computational complexity. I. Ullman,
Jeffrey D., 1942- joint author. II. Title.
QA267.H56 629.8'312 78-67950
ISBN 0-201-02988-X

CHAPTER

1

PRELIMINARIES

In this chapter we survey the principal mathematical ideas necessary for understanding the material in this book. These concepts include graphs, trees, sets, relations, strings, abstract languages, and mathematical induction. We also provide a brief introduction to, and motivation for, the entire work. The reader with a background in the mathematical subjects mentioned can skip to Section 1.6 for motivational remarks.

1.1 STRINGS, ALPHABETS, AND LANGUAGES

A "symbol" is an abstract entity that we shall not define formally, just as "point" and "line" are not defined in geometry. Letters and digits are examples of frequently used symbols. A *string* (or *word*) is a finite sequence of symbols juxtaposed. For example, *a*, *b*, and *c* are symbols and *abcb* is a string. The *length* of a string *w*, denoted $|w|$, is the number of symbols composing the string. For example, *abcb* has length 4. The empty string, denoted by ϵ, is the string consisting of zero symbols. Thus $|\epsilon| = 0$.

A *prefix* of a string is any number of leading symbols of that string, and a suffix is any number of trailing symbols. For example, string *abc* has prefixes ϵ, *a*, *ab*, and *abc*; its suffixes are ϵ, *c*, *bc*, and *abc*. A prefix or suffix of a string, other than the string itself, is called a *proper* prefix or suffix.

The *concatenation* of two strings is the string formed by writing the first, followed by the second, with no intervening space. For example, the concatena-