Preface

String pattern matching is an important component of many areas of science and information processing. It occurs naturally as part of data processing, text editing, symbol manipulation, term-rewriting, lexical analysis, code generation, spelling correction, bibliographic search, text retrieval, and natural language processing. String pattern matching techniques can be also applied to the recognition of patterns such as shapes, pictures, scenes, and so on. In biology, string pattern matching problems arise in the analysis of protein sequences and nucleic acids and in the investigation of molecular phylogeny. String pattern matching is the most time-consuming part of many programs, and the substitution of a poor matching method by a good one often leads to a substantial increase in speed. Therefore, a fast methodology should be selected. The aim of this volume is to introduce the basic concepts and characteristics of string pattern matching strategies and to provide numerous references for further reading.
The pattern matcher is a program that takes as input the text string x and produces as output the locations in x at which patterns, or keywords, appear as substrings. The simplest patterns are single keywords that match themselves. A somewhat broader class of patterns would be sets of keywords. There are two important variants of pattern-matching problems. One is approximate string matching problems, in which one must find all substrings in a text that are close to a pattern according to some measure of closeness. Another is multidimensional matching problems for finding patterns in higher-dimensional structures such as trees and graphs. In recent years some types of pattern matching algorithms have been implemented on hardware based on the finite state automata and signature files in order to improve processing efficiency. In this book, string pattern matching strategies are classified into the following five chapters.

• Single keyword matching
• Matching sets of keywords
• Approximate string matching
• Multidimensional matching
• Hardware matching

As an introduction to each chapter, my survey article describes the basic concepts of classification mentioned above. Fifteen papers have been selected to further illustrate these concepts. Also, I have made considerable efforts to find a large number of corresponding references and to organize them. Most of the string matching techniques are treated in detail, with mathematical analyses and suggestions for practical applications, in the books and articles cited throughout the book. The references [Aho, 80], [Aho, 90], [Apostolico et al., 85], [Gonnet et al., 85], and [Sankoff et al., 83] are good surveys for general string pattern matching techniques. The six books [Aho et al., 74], [Frakes et al., 92], [Knuth, 73], [Mehlhorn, 84], [Sedgewick, 86], and [Standish, 80] are useful for the corresponding basic data structures and algorithms.

ix