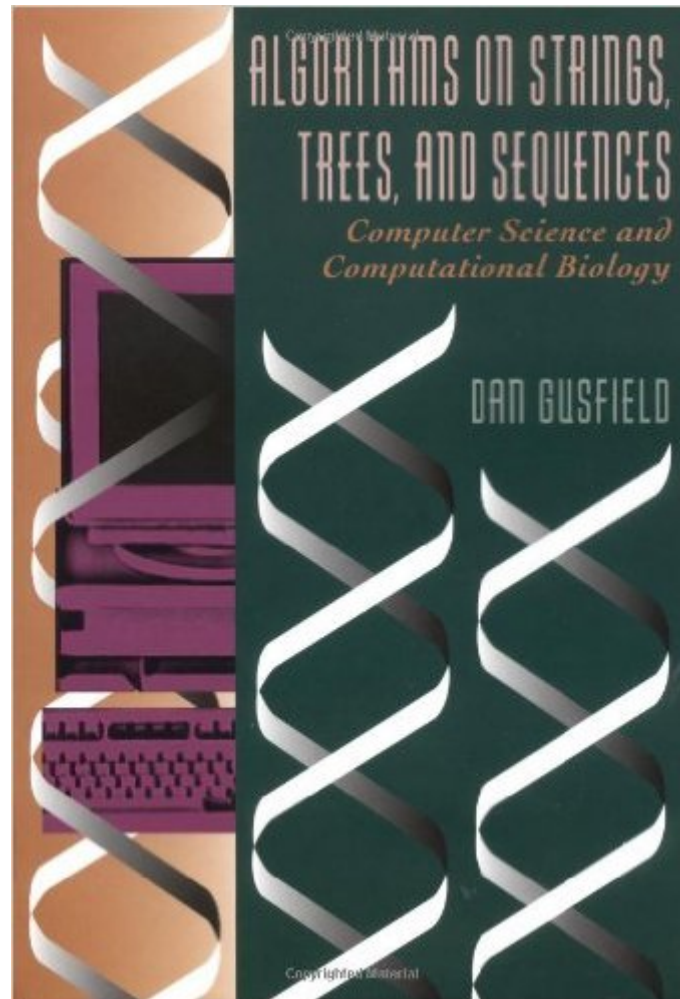# Algorithms On Strings, Trees And Sequences: Computer Science And Computational Biology

Synopsis

Traditionally an area of study in computer science, string algorithms have, in recent years, become an increasingly important part of biology, particularly genetics. This volume is a comprehensive look at computer algorithms for string processing. In addition to pure computer science, Gusfield adds extensive discussions on biological problems that are cast as string problems and on methods developed to solve them. This text emphasizes the fundamental ideas and techniques central to today's applications. New approaches to this complex material simplify methods that up to now have been for the specialist alone. With over 400 exercises to reinforce the material and develop additional topics, the book is suitable as a text for graduate or advanced undergraduate students in computer science, computational biology, or bio-informatics.

Book Information

Hardcover: 556 pages

Publisher: Cambridge University Press; 1 edition (May 28, 1997)

Language: English

ISBN-10: 0521585198

ISBN-13: 978-0521585194

Product Dimensions: 7 x 1.2 x 10 inches

Shipping Weight: 3.4 pounds (View shipping rates and policies)

Average Customer Review: 5.0 out of 5 starsÂ Â See all reviewsÂ (13 customer reviews)

Best Sellers Rank: #122,476 in Books (See Top 100 in Books)   #47 inÂ Books > Science & Math > Mathematics > Pure Mathematics > Discrete Mathematics   #73 inÂ Books > Computers & Technology > Programming > Algorithms   #582 inÂ Books > Textbooks > Computer Science > Programming Languages

Customer Reviews

This textbook gives a rigorous introduction to the algorithms of computational biology from the standpoint of theoretical computer science. It does however give the reader an overview of the practical application of these algorithms to the subject. The author gives a very detailed discussion of the most important results in the field, but the book is very dense: there are 228 definitions, 127 theorems, 490 references, and over 400 exercises that both illustrate the topics in the book and extend them. The author omits any real source code, but does give a URL where code for many of the algorithms can be found. The author restricts his attention to deterministic approaches to string matching and comparison, and thus there is no treatment of hidden Markov models or Monte Carlo

methods. The major algorithms such as the Aho-Corasick, Boyer-Moore, Knuth-Morris-Pratt, Needleman-Winsch, and Smith-Waterman are discussed and brilliantly motivated in the book. The author employs very effective diagrams to illustrate the matching concepts that are detailed in the book. The book does require some time to read but it is worth the effort. Also, the exercises can be challenging but some should he done in order to understand the concepts in the book. The empirical results of the algorithms as sequence databases are also included, with FASTA, BLAST, BLOCKS, BLOSUM, and PROSITE are discussed in detail. The chapter that discusses these is the least mathematical of all the ones in the book and was no doubt included to connect the reader with real-world applications of the techniques in the book. The last quarter of the book is a lot more trendy than the rest, with emphasis placed on algorithms for physical mapping, fragment assembly, and phylogenetic trees.

If you like definition-theorem-proof-example and exercise books, Gusfield's book is the definitive text for string algorithms. The algorithms are abstracted from their biological applications, and the book would make sense without reading a single page of the biological motivations. Gusfield aims his book at readers who are fluent in basic algorithms and data structures (at the level of Cormen, Leisersohn and Rivest's excellent text). The exercises are wonderfully illustrative, being neither trivial nor impossible.All of the major exact string algorithms are covered, including Knuth-Morris-Pratt, Boyer-Moore, Aho-Corasick and the focus of the book, suffix trees for the much harder probem of finding all repeated substrings of a given string in linear time. In addition to exact string matching, there are extensive discussions of inexact matching. Even the discussions of widely known topics like dynamic programming for edit distance are insightful; for instance, we find how to easily cut space requirements from quadratic to linear. There is also a short chapter on semi-numerical matching methods, which are also of use in information retrieval applications. Inexact matching is extended to the threshold all-against-all problem, which finds all substrings of a string that match up to a given edit distance threshold. The theoretical development concludes with the much more difficult problem of aligning multiple sequences with ultrametric trees, with applications to phylogenetic alignment for evolutionary trees (an approach that has also been applied to the evolution of natural languages).Note that there is no discussion of statistical string matching.

Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology

Biological Modeling and Simulation: A Survey of Practical Models, Algorithms, and Numerical

Methods (Computational Molecular Biology) Introduction to Computational Social Science: Principles and Applications (Texts in Computer Science) Genetic Algorithms and Genetic Programming in Computational Finance Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications (International Series on Computational Intelligence) Bio-inspired Algorithms for the Vehicle Routing Problem (Studies in Computational Intelligence) Verification of Computer Codes in Computational Science and Engineering Python: Python Programming For Beginners - The Comprehensive Guide To Python Programming: Computer Programming, Computer Language, Computer Science Python: Python Programming For Beginners - The Comprehensive Guide To Python Programming: Computer Programming, Computer Language, Computer Science (Machine Language) Bioinformatics and Computational Biology in Drug Discovery and Development Computational Biology -: Unix/Linux, Data Processing and Programming Python for Bioinformatics (Chapman & Hall/CRC Mathematical and Computational Biology) RNA-seq Data Analysis: A Practical Approach (Chapman & Hall/CRC Mathematical and Computational Biology) Principles of Computational Cell Biology The Design of Innovation: Lessons from and for Competent Genetic Algorithms (Genetic Algorithms and Evolutionary Computation) Algorithms in C++ Part 5: Graph Algorithms (3rd Edition) (Pt.5) Parallel and Distributed Map Merging and Localization: Algorithms, Tools and Strategies for Robotic Networks (SpringerBriefs in Computer Science) Sigma Delta Modulators: Nonlinear Decoding Algorithms and Stability Analysis (The Springer International Series in Engineering and Computer Science) Algorithms to Live By: The Computer Science of Human Decisions The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation

[Dmca](#)