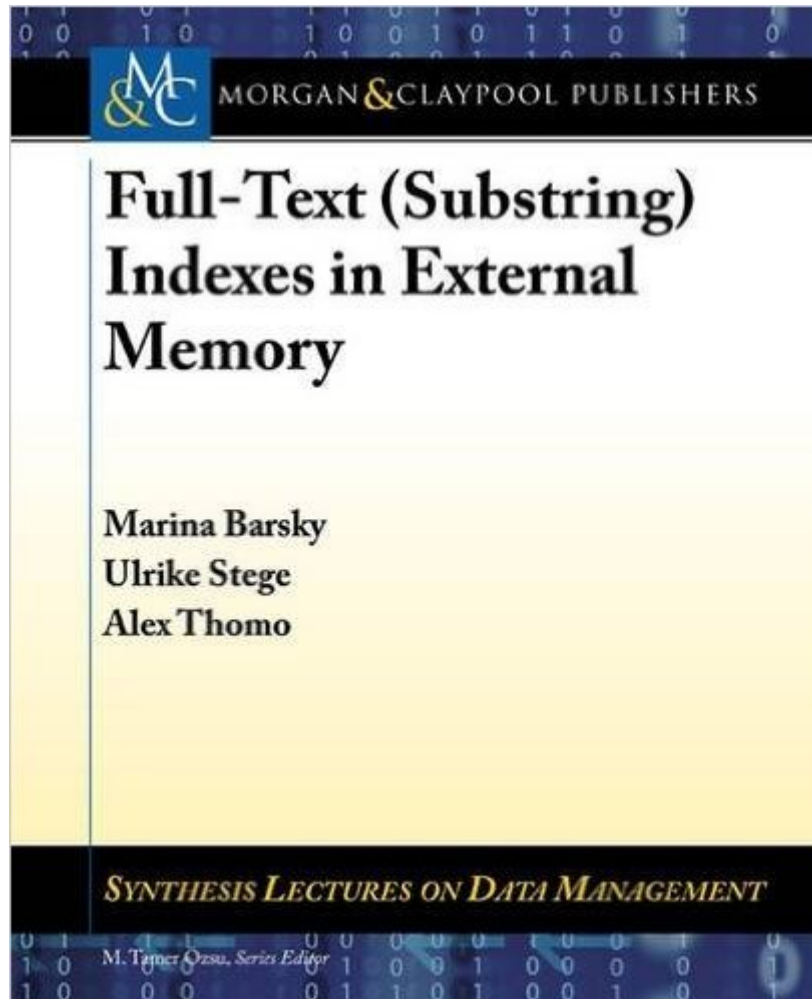The book was found

Full-Text (Substring) Indexes In External Memory (Synthesis Lectures On Data Management)

MORGAN&CLAYPOOL PUBLISHERS

Full-Text (Substring) Indexes in External Memory

Marina Barsky
Ulrike Stege
Alex Thomo

SYNTHESIS LECTURES ON DATA MANAGEMENT

M. Tamer Özsu, Series Editor

Synopsis

Nowadays, textual databases are among the most rapidly growing collections of data. Some of these collections contain a new type of data that differs from classical numerical or textual data. These are long sequences of symbols, not divided into well-separated small tokens (words). The most prominent among such collections are databases of biological sequences, which are experiencing today an unprecedented growth rate. Starting in 2008, the "1000 Genomes Project" has been launched with the ultimate goal of collecting sequences of additional 1,500 Human genomes, 500 each of European, African, and East Asian origin. This will produce an extensive catalog of Human genetic variations. The size of just the raw sequences in this catalog would be about 5 terabytes. Querying strings without well-separated tokens poses a different set of challenges, typically addressed by building full-text indexes, which provide effective structures to index all the substrings of the given strings. Since full-text indexes occupy more space than the raw data, it is often necessary to use disk space for their construction. However, until recently, the construction of full-text indexes in secondary storage was considered impractical due to excessive I/O costs. Despite this, algorithms developed in the last decade demonstrated that efficient external construction of full-text indexes is indeed possible. This book is about large-scale construction and usage of full-text indexes. We focus mainly on suffix trees, and show efficient algorithms that can convert suffix trees to other kinds of full-text indexes and vice versa. There are four parts in this book. They are a mix of string searching theory with the reality of external memory constraints. The first part introduces general concepts of full-text indexes and shows the relationships between them. The second part presents the first series of external-memory construction algorithms that can handle the construction of full-text indexes for moderately large strings in the order of few gigabytes. The third part presents algorithms that scale for very large strings. The final part examines queries that can be facilitated by disk-resident full-text indexes. Table of Contents: Structures for Indexing Substrings / External Construction of Suffix Trees / Scaling Up: When the Input Exceeds the Main Memory / Queries for Disk-based Indexes / Conclusions and Open Problems

Book Information

ISBN-13: 978-1608457953

Product Dimensions: 7.5 x 0.2 x 9.2 inches

Shipping Weight: 7.8 ounces (View shipping rates and policies)

Average Customer Review: Be the first to review this item

Best Sellers Rank: #2,152,003 in Books (See Top 100 in Books)   #36 in Books > Computers & Technology > Programming > Algorithms > Memory Management   #6091 in Books > Computers & Technology > Databases & Big Data   #459941 in Books > Reference

Full-Text (Substring) Indexes in External Memory (Synthesis Lectures on Data Management) Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business. Leveraging the Power of Data Analytics, Data ... (Hacking Freedom and Data Driven) (Volume 2) Transactional Memory, 2nd Edition (Synthesis Lectures on Computer Architecture) An Introduction to Text-to-Speech Synthesis (Text, Speech and Language Technology) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Operating System Security (Synthesis Lectures on Information Security, Privacy, and Trust) Spoken Dialogue Systems (Synthesis Lectures on Human Language Technologies) Natural Language Processing for Social Media (Synthesis Lectures on Human Language Technologies) Quantum Memory Power: Learn to Improve Your Memory with the World Memory Champion! Lighting Design on Broadway: Designers and Their Credits, 1915-1990 (Bibliographies and Indexes in the Performing Arts) Spon's External Works and Landscape Price Book 2014 Maximizing the Value of Consulting: A Guide for Internal and External Consultants The Complete Journal of Discourses - Deluxe LDS Reference Edition - with Comprehensive TOPICAL Guide, Multiple Indexes, Speaker Biographies, & Over 12,500 Links Big Data Appliances for In-Memory Computing: A Real-World Research Guide for Corporations to Tame and Wrangle Their Data Lectures on Calvinism, The Stone Lectures of 1898 The Feynman Lectures on Physics, Vol. II: The New Millennium Edition: Mainly Electromagnetism and Matter (Feynman Lectures on Physics (Paperback)) (Volume 2) Asset Management: A Systematic Approach to Factor Investing (Financial Management Association Survey and Synthesis) Holy Bible Text Edition NLT: New Living Translation (Text Edition: Full Size) The Project Management Memory Jogger (Second Edition) (Memory Jogger Series)