Abstracts

An introduction of Burrows-Wheeler transform and its variants

Wing-Kai Hon

National Tsing Hua University, Taiwan

ABSTRACT

Burrows-Wheeler transform (BWT) was proposed in 1994 as a means to perform lossless compression on text strings. Soon after that, Ferragina and Manzini (FOCS'00; JACM'05) discovered the string matching power of BWT, and this result, together with independent work by Grossi and Vitter (STOC'00; SICOMP'07) and by Sadakane (ISAAC'00; JALG'03), subsequently started the field of compressed text indexing. Nowadays, BWT gains much popularity in bioinformatics areas, for a couple of famous applications such as BWA, Bowtie, and SOAP2, use BWT as the core in their design. In this talk, we will give an introduction of BWT and some of its variants, and show how they work, and how they can be used to derive interesting results, both in theory and in practice. No background knowledge is required.

Bio

Wing-Kai Hon is a Professor in the Computer Science Department at National Tsing Hua University. He received his doctorate from the University of Hong Kong. His research interests are string matching, indexing, data compression, and combinatorial optimization. Webpage: http://www.cs.nthu.edu.tw/ wkhon

Combinatorial algorithms for grammar-based text compression

Shunsuke Inenaga

Kyushu University, Japan

ABSTRACT

Grammar compression of a string is a class of data compression that describes the string with a context-free grammar that generates only the string. It is known that grammar compression is a powerful compression especially for highly repetitive strings. While computing the smallest grammar for a given string is NP-hard, a number of approximation algorithms and practical compression algorithms have been proposed and have widely been used (e.g., the LZ78 / LZW family).

In this tutorial talk, we first give an overview on compressed-string processing, where that task is to process grammar-compressed strings without explicitly decompressing the data and hence in compressed space. This line of research includes fundamental work for text-indexing, text-mining, and pattern discovery, on grammar-compressed strings. Most, if not all, of theses algorithms make use of combinatorial properties on strings and grammars. We then also review recent work on grammar compression algorithms themselves. The talk is targeted to an audience with general interests to algorithms / data structures / data compression / formal languages / combinatorics on words.

Bio

Shunsuke Inenaga received M.S. and Ph.D. degrees from Kyushu University in 2002 and 2003, respectively. From 2003 to 2005, he worked as a pos-doc researcher for Japan Science Technology Agency (JST), the University of Helsinki, Kyoto University, and Japan Society for the Promotion of Science (JSPS). From 2005 to 2011, he was a tenure track research fellow at Kyushu University. Since 2011, he has been an associate Professor at Kyushu University. Since 2019, he has concurrently been a researcher for JST PRESTO (Sakigake).

Graph models and algorithms in genome assembly

Yu Lin

Australian National University, Australia

ABSTRACT

Sequencing data is being collected in enormous amounts by biologists, environmental scientists and medical researchers. Sequencing data usually consists of millions or billions of short DNA sequences, called reads, that are randomly drawn from genomes. The problem of genome assembly is to assemble these reads into a single genome by using the overlapping ends to link them. In this tutorial, I will introduce existing graph models and algorithms for genome assembly and show how to apply and generalize these models to assemble both short accurate reads (from the second-generation sequencing) and long error-prone reads (from the third-generation sequencing).

Bio

Yu Lin is a lecturer and group leader at the Research School of Computer Science, the Australian National University. Prior to this, he was a postdoctoral scholar, hosted by Prof Pavel Pevzner at the Department of Computer Science and Engineering, University of California, San Diego. He received his PhD in Computer Science under Prof Bernard Moret from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. His research focuses on computational biology, and he has been working on algorithms for genome assembly, the analysis of genome rearrangements, and phylogenetic reconstruction.