Compact Text Indexing for Advanced Pattern Matching Problems: Parameterized, Order-Isomorphic, 2D, etc.

Sharma V. Thankachan
Department of Computer Science, University of Central Florida, Orlando, FL, USA

Abstract

In the past two decades, we have witnessed the design of various compact data structures for pattern matching over an indexed text [22]. Popular indexes like the FM-index [6], compressed suffix arrays/trees [15, 26], the recent r-index [8, 23], etc., capture the key functionalities of classic suffix arrays/trees [20, 28] in compact space. Mostly, they rely on the Burrows-Wheeler Transform (BWT) and its associated operations [2]. However, compactly encoding some advanced suffix tree (ST) variants, like parameterized ST [1, 19, 21], order-isomorphic/preserving ST [4], two-dimensional ST [14, 16], etc. [24, 27]- collectively known as suffix trees with missing suffix links [3], has been challenging. The previous techniques are not easily extendable because these variants do not hold some structural properties of the standard ST that enable compression. However, some limited progress has been made in these directions recently [11, 7, 5, 25, 10, 18, 17, 12, 13, 9]. This talk will briefly survey them and highlight some interesting open problems.

2012 ACM Subject Classification Theory of computation → Pattern matching

Keywords and phrases Text Indexing, Suffix Trees, String Matching

Digital Object Identifier 10.4230/LIPIcs.CPM.2022.3

Category Invited Talk

Funding Supported in part by the US NSF grants CCF-1527435, CCF-2112643 and CCF-2137057.

Acknowledgements I want to thank my collaborators Rahul Shah and Arnab Ganguly.

References

Compact Text Indexing for Advanced Pattern Matching Problems


