World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:02, 2017

Compressed Suffix Arrays to Self-Indexes Based on Partitioned Elias-Fano

Authors : Guo Wenyu, Qu Youli

Abstract : A practical and simple self-indexing data structure, Partitioned Elias-Fano (PEF) - Compressed Suffix Arrays (CSA), is built in linear time for the CSA based on PEF indexes. Moreover, the PEF-CSA is compared with two classical compressed indexing methods, Ferragina and Manzini implementation (FMI) and Sad-CSA on different type and size files in Pizza & Chili. The PEF-CSA performs better on the existing data in terms of the compression ratio, count, and locates time except for the evenly distributed data such as proteins data. The observations of the experiments are that the distribution of the φ is more important than the alphabet size on the compression ratio. Unevenly distributed data φ makes better compression effect, and the larger the size of the hit counts, the longer the count and locate time.

Keywords : compressed suffix array, self-indexing, partitioned Elias-Fano, PEF-CSA
Conference Title : ICWAIM 2017 : International Conference on Web-Age Information Management
Conference Location : Venice, Italy
Conference Dates : February 16-17, 2017