University of Dublin
TRINITY COLLEGE

A 12*n* Byte Approach to Speeding up Qsufsort

BY CHIA LAU

B.A (Mod.) Computer Science
Final Year Project, April 2020
Supervisor: Jeremy Jones

School of Computer Science and Statistics
O'Reilly Institute, Trinity College, Dublin 2, Ireland

Abstract

Suffix arrays are a space-efficient data structure used for string processing. Perhaps the most important application of suffix arrays is in the computation of the Burrows-Wheeler Transform which has uses in data compression and more recently bio-informatics.

This paper suggests changes to *qsufsort*, a suffix array construction algorithm, with the aim of speeding it up. The modifications proposed maintain qsufsort's worst case time complexity of $O(n \log n)$. Speed-up is achieved by trading space for running time and exploiting radix sorting for faster integer sorting. This speed-up is demonstrated experimentally on real world data.