

2003 Paper 6 Question 1

Data Structures and Algorithms

- (a) In the Burrows–Wheeler Block Compression algorithm it is necessary to sort all the suffixes of a vector of possibly tens of millions of bytes.
- (i) Explain why Shell sort using a simple character by character string comparison function is unlikely to be satisfactory. [2 marks]
 - (ii) Describe in detail the data structures and algorithms you would use to sort the suffixes and explain why your method is an improvement. You may assume that you have plenty of RAM available. [8 marks]
- (b) In the Burrows–Wheeler Block Compression algorithm a sequence of small positive integers are transmitted using Huffman encoding.
- (i) Describe how the Huffman code is constructed. [4 marks]
 - (ii) Outline how it can be represented compactly in the compressed file. [6 marks]