

1997 Paper 8 Question 6

Advanced Algorithms

A posting on a newsgroup announces the invention of a new compression algorithm, and claims that the method will guarantee to compress at least 10% of all possible input files by at least 10% of their original size, but that it might (unfortunately) cause some of the other 90% of possible inputs to expand by a factor of up to 90.

Discuss how believable and reasonable the claim is.

[8 marks]

Various compression utilities running on personal computers typically reduce text files to 1/3 of their original size. Estimate the proportion of all possible files (including binary ones) whose original length is less than 64 kbytes that could be compressed to this extent by an ideal compression algorithm. What proportion of the same set of files consists of just alphanumeric characters, blanks and newlines?

[12 marks]