1.

(a)How big (in bytes) are the compressed text files produced in each case?

(b)How big (in bytes) are the symbol models produced in each case?

Char	Model	Compressed Text
Size(byte)	2,117	690,365

Word	Model	Compressed Text
Size(byte)	1,892,146	36,935

(c)How long does it take to

i. build the symbol model?

ii. encode the input file given the symbol model?

iii.decode the compressed file?

	Word	Char
Build Model	19.09	0.43
Encode	0.06	0.74
Decode	1.17	291.95

2.

How could various aspects of the performance, as identified in response to the previous question, be improved?

To improve the “Build Model Time”, we can apply some efficient method on building the Huffman code tree, for example, insert sort is a better way than to append it to the list. Besides, a recursive function can also do some help.

To improve the encode time, we have to traverse all the texts to make them into codes, I don’t think there’s anything can improve it.

To improve the decode time, as for myself, I rebuild the Huffman code tree with the model which took lot of time. However, if we can parse or get the tree in other ways, we must can improve the performance on time.