CprE 315

Programming Assignment #1

Due: 9/28/2012
Total 100 points.

In Unix-based systems, there are utilities like zip/unzip for file compression/decompression. In this assignment, you are asked to implement similar compression and decompression programs based on Huffman coding.

The compression program "compress" has the following command line format:

compress <input-file> <output-file>

Compress takes any file <input-file> as input. It applies the Huffman coding algorithm to encode *each byte* by several bits. The Huffman code together with the resulting bit stream is then written to an output file named <output-file>.

The decompression program "decompress" has the following command line format:

decompress <input-file> <output-file>

Its input file <input-file> is any output file from "compress". The bit stream in the input file is decoded based on the associated Huffman code. The decompressed version of the bit stream is then written to an output file named <output-file>.

For example, consider the following sequence of commands:

compress f.doc f.doc.compressed
decompress f.doc.compressed f-new.doc

The file f.doc and f-new.doc should be identical. (In Unix-based systems, the command "cmp" can be used to make sure two files are identical.)

In this assignment, the way to store the Huffman code in the compressed file is not specified. You can derive your own way. (As a result, "decompress" of one student may not be able to handle the compressed file generated by "compress" of another student.) In order to minimize the file size of the compressed file, you should try to minimize the storage required for the Huffman code in the compressed file.

You must use either C or C++ for the implementation. Four sample input files "sample1.txt", "sample2.docx", "sample3.exe" and "sample4.jpg" are posted. Please generate a report "report.docx" which contains the following table (with text in red replaced by the actual file sizes in bytes). Please email report.docx and the source codes of both "compress" and "decompress" to the TA.

	Original	Compressed	Decompressed
sample1.txt	File size of sample1.txt	File size of sample1.txt.compressed	File size of sample1-new.txt
sample2.docx	File size of sample2.docx	File size of sample2.docx.compressed	File size of sample2-new.docx
sample3.exe	File size of sample3.exe	File size of sample3.exe.compressed	File size of sample3-new.exe
sample4.jpg	File size of sample4.jpg	File size of sample4.jpg.compressed	File size of sample4-new.jpg