Report - assignment 8

Camilla K. Birkelund IDATT2101 November 8th, 2022

Assignment description

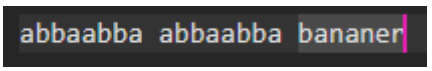
The assignment was to create a program that could compress and decompress a file using the Lempel-Ziv and Huffman algorithms. The requirements for the solution was that the resulting decompressed file would have to be equal to the original file, and the compressed file had to be smaller than the original file.

Implementation

The source code can be found in “LZ77.java”, “Huffman.java” and “client.java”. To run the program, the LZ77.java, Huffman.java and client.java files have to be open. When running client.java, the user gets the option to either compress or decompress a file (or quit the program). The program then asks which file to perform the operations on. The default method to compress is to first use Lempel-Ziv 77, and then Huffman. The result will be written to a filename of the user's choice during the program running. With decompression, the program performs Huffman first, then Lempel-Ziv.

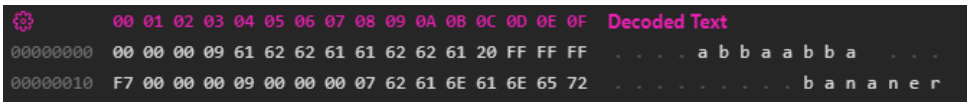
Testing

During the implementation, several tests were performed on a small .txt file to make sure that the algorithms for compression and decompression gave the expected results.

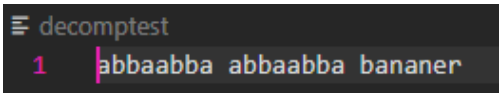


The original file.

Testing the Lempel-Ziv 77 algorithm

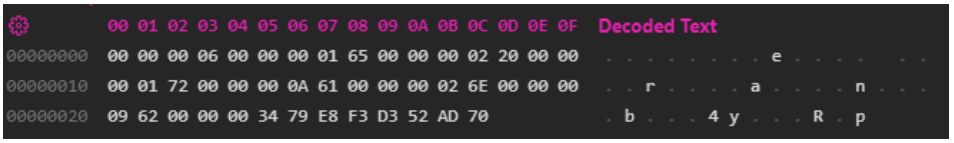


The compressed file.

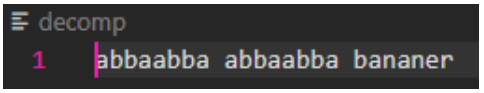


The decompressed file.

Testing the Huffman algorithm

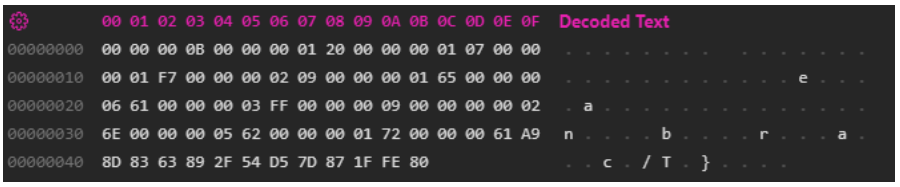


The compressed file

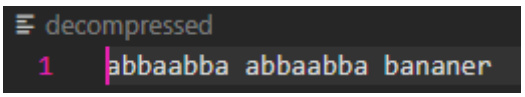


The decompressed file.

Testing the Lempel-Ziv and Huffman together



The compressed file.



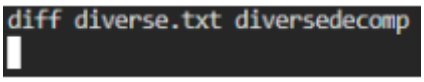
The decompressed file.

Results

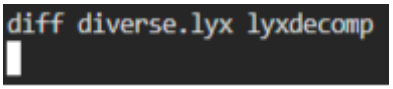
The requirement from the assignment description was to be able to compress and decompress at least two of the attached files, which are much larger than the testfile used to check the algorithms. The files used in this case “diverse.txt”, “diverse.lyx” and “diverse.pdf”.

The first requirement from the program is that the size of the compressed file has to be smaller than the original. The compressed files were smaller than the original files in all three cases. When compressing “diverse.txt”, which is 16764 bytes, the resulting file is 11134 bytes. When compressing “diverse.lyx” the resulting file is 39121 bytes, whereas the original is 181076 bytes. The “diverse.pdf” file was originally 156671 bytes, and the compressed version was 147789. The difference between the compressed file and the original file for “diverse.pdf” might have been bigger if the file was compressed with only Lempel-Ziv, because the Huffman algorithm may have expanded the file size due to the Huffman tree structure.

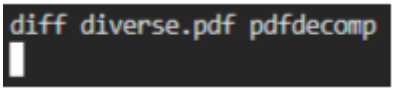
The second requirement was that the decompressed file is equal to the original file. When testing the compressed, then decompressed version of diverse.txt, diverse.lyx and diverse.pdf against the original diverse.txt, diverse.lyx and diverse.pdf files, there are no differences.



The result of checking the differences between the original file and the compressed and decompressed version of diverse.txt.



The result of checking the differences between the original file and the compressed and decompressed version of diverse.lyx.



The result of checking the differences between the original file and the compressed and decompressed version of diverse.pdf