CENG 520 Homework 1

202171651 - Deniz Adalı ATLIHAN

1 Overview

All the source codes, outputs and reports are published in the github link below. `https://github.com/dnzatlihan/ceng520`

For this assignment, English alphabet is selected and three different data sources are used for experiments. For compression, Brotli algorithm is used. External compression tool (`https://www.multiutil.com/text-to-brotli-compress/`) is used for generating Brotli files. These files renamed with same name but with .brotli extension.

There are three test files are used. In loremipsum5paragraph.txt file, 5 paragraph Lorem Ipsum test is used, in pfWallAll.txt file, all lyrics and narrator notes of The Wall album of Pink Floyd is used, and in ayreonTheoryOfEverything-Lyrics.txt file, all lyrics and narrator notes of Theory of Everything album of Ayreon is used. Literature review mentioned in question is given with separate file named with **Hw1Q3-TCP_Error_Correction.pdf**

2 Question 1

In ceng520/homework1/Hw1Q1Q2.java file, frequency latter generator function (private static Map¡Character, Double¿ createFrequencyTable(File file, Set¡Character¿ alphabet) throws IOException) is developed. The function developed with attributes given in the question 1 applied three data files for English alphabet. Frequency tables of these files are given in the output.txt file.

3 Question 2

Entropy of Brotli compressed files are calculated with all characters for Brotli text. As shown in output.txt file, entropy of compressed sources are higher than raw files. While entropi of raw files increasing, entropi of compressed files are increasing too.

4 Question 3

Literature review mentioned in question is given with separate file named with **Hw1Q3-TCP_Error_Correction.pdf**