BIRZEIT UNIVERSITY

Design and Analysis of Algorithms (COM336)
First Semester 2023/2024
Project # 2
Due date (Project#1, Project#2) 18-Dec-2023

Huffman Coding

Huffman coding is a lossless data compression algorithm. The idea is to assign variable-length codes to input characters; lengths of the assigned codes are based on the frequencies of corresponding characters. The most frequent character gets the smallest code and the least frequent character gets the largest code.

The variable-length codes assigned to input characters are Prefix Codes, means the codes (bit sequences) are assigned in such a way that the code assigned to one character is not prefix of code assigned to any other character. This is how Huffman Coding makes sure that there is no ambiguity when decoding the generated bit stream.

In this project, you will be using a priority queue and a binary tree of your design to implement a file compression/uncompression algorithm called "Huffman Coding".

Your program will read any file and compress it using your implementation of the Huffman coding algorithm found in the explanation. The compressed data will be written to a file. That compressed file will be then be read back by your program and uncompressed. The uncompressed file will then be written to a third file. The uncompressed file should of course match the original file.

Summary of Processing

- Read the specified file and count the frequency of all bytes in the file.

- Create the Huffman coding tree based on the frequencies.

- Create and display a table of encodings for each byte from the Huffman coding tree.

- Encode the file and output the encoded/compressed file.

- Display the header that include the huffman code.

- Read the encoded/compressed file you just created, decode it and output the decoded file.

<div dir="rtl">

ملاحظة:

- لن يتم تصحيح أي مشروع لا يتبع الخطوات التي تم شرحها بالمحاضرة للحل.
- يجب ان يكون هنالك واجهة تطبيق مرتبة وتناسب المطاليب ولا يجوز طباعة أي نتيجة الامن خلال واجهة التطبيق

</div>