Algorithms – Huffman Compression Assignment

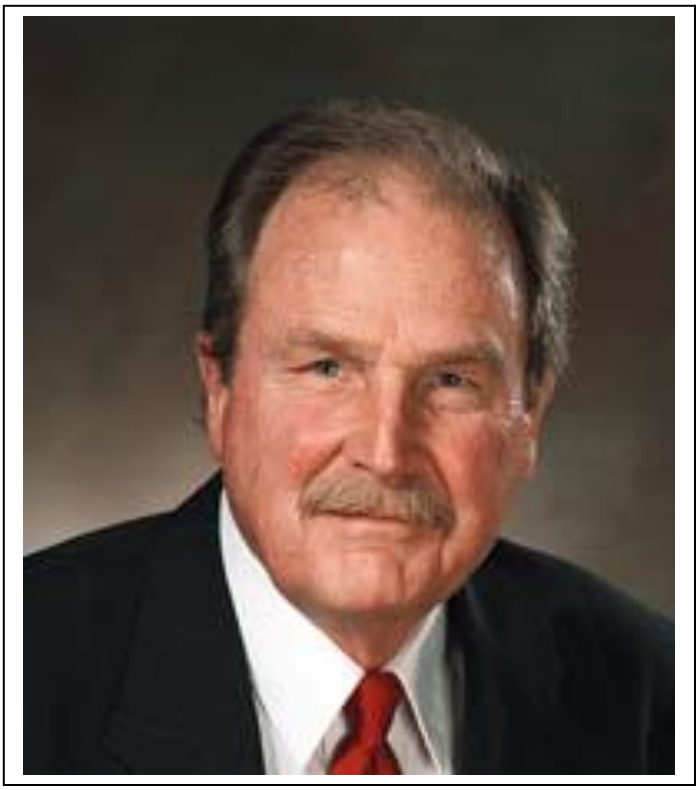
Name: Louay Hesham Saber

ID: 3303

Name: Rowan Tarek Hassan Tahseen

ID: 4447

Brief Look at History:



David Huffman

The story of the invention of Huffman codes is a great story that demonstrates that students can do better than professors. **David Huffman** (1925-1999) was a student in an electrical engineering course in 1951. His professor, **Robert Fano**, offered students a choice of taking a final exam or writing a term paper. Huffman did not want to take the final so he started working on the term paper. The topic of the paper was to find the most efficient (optimal) code. What Professor Fano did not tell his students was the fact that it was an open problem and that he was working on the problem himself. Huffman spent a lot of time on the problem and was ready to give up when the solution suddenly came to him. The code he discovered was optimal, that is, it had the lowest possible average message length. The method that Fano had developed for this problem did not always produce an optimal code. Therefore, Huffman did better than his professor. Later Huffman said that likely he would not have even attempted the problem if he had known that his professor was struggling with it.

Algorithm Description:

Compression

- Read File to get the frequency of each character – O(n)
- Build a Min-Heap of the frequencies – O(m log m)
- DFS travers to assign 0 to the left node and 1 to right node – O(2m)
- Change each 8 bits to 1 byte – O(n)
- Write in file – O(n)

Total = O(n) + O(m log m) , when n is size and m is the unique values

Decompression

- Read Decompressed file – O(n)
- Match values from dictonary – O(n)

Total = O(n)

Header Format:

SizeOfFile;Byte:Code#Byte:Code#Byte:Code#

Example: 32963136;32:00#98:010000#

Bonus:

1. Compression and Decompression of Binary Files
2. Compression and Decompression of Folders

Code Tested on:

- .txt
- .pdf
- audio