Author: Phong Nguyen - phn10
Author: Joel Rand - jsr99

## I. Introduction

Our project is Huffman Data Compressing Program. It receives a text file, uses Huffman coding algorithm to compress the data, and outputs a binary file. The output binary file is expected to have smaller memory than the input text file.

## II. Data Structures and Algorithm

1. Tree: used to create Huffman tree
2. Hash: used to stored pairs of character and its binary representation
3. Quicksort: used to sort the occurrence of each character.

## III. Source Files

The program is comprised of 6 java programs. All the binary files are contained in the **bin** folder. All the source java programs are contained in the **src** folder:

1. **BinaryStdOut.java**: this program output a text file in bits unit. This program was written by professor Robert Sedgewick and professor Kevin Wayne from Princeton University. Retrieved from: http://algs4.cs.princeton.edu/55compression/BinaryStdOut.java.html
2. **HuffmanTree.java**: the HuffmanTree class has two functions. The first one named buidTree: it receives an array of nodes and recursively build the Huffman Tree for us. The second one is named enCode: it receives the top node in in the Huffman Tree and encoding every nodes in the tree using Huffman coding.
3. **CharacterList.java**: the CharacterList.java represents each nodes in the Huffman Tree. Basically, it is a node object instance.
4. **FileReader.java**: have two functions. The first one is to parse a text file into a java string. The second one is to output a text file from Java string. However, we realize the output function doesn't optimize the compressing feature, so we not gonna use this function.
5. **Main.java**: the main function of the program.
6. **SortList.java**: has a sort function using quicksort algorithm. The sort function will sort the occurrence of each characters in the input file, from most occurrence to least occurrence.

## IV. Input and Output:

The program receives **input.txt** and output **output.txt.**

**Input.txt (size: 70 bytes)**

hello world this is a file using to test the huffman coding algorithm

**output.txt (size: 30 bytes)**

»¬]2×ŸCàû¢ã)ÁŠúßûæ4«œo÷‚éå¥\00

The **output.txt** is compressed and has half size of the **input.txt.** All the characters in output.txt are encoded and in binary representation.

## V. How you can test the program?

### A. In Linux (and you have **/bin/bash**)

Go to computer **terminal,** navigate to the code directory:
You can see two bash scripts: **install.sh** and **run.sh**
To compile the program, type

**./install.sh**

It gonna create all the **.class** files in the **bin** folder
To run the test, run

**./run.sh**

It gonna create the file **output.txt**

### B. In Windows

Use any IDE you have (I often use DrJava and Eclipse) and open up the file, click on **compile** and **run**

In either cases, after running the program, you gonna see the **output.txt** file
You can try to change the length of **input.txt** and see how length and the diversity of characters in **input.txt** changes **output.txt**