**Project 1**

**Compressing Text Files, with fixed bit-length encoding**

|  |  |
| --- | --- |
| **Release Date (Part 1)** | Thursday, February 12, 2015, 11:55PM |
|  |  |
| **Due Date** | Tuesday, February 17, 2015, before class |
|  |  |
| **Absolute Weight (for all parts)** | 3% |
|  |  |

**Statement: Part 1**

This project will be completed in parts. The statement below describes the first part.

Given a text file, T, go through the text in T and count the number of unique characters. Then store these characters in a character array, C. Each character in C is a unique symbol from the file, and is to be assigned a "code", that is, a combination of 1 and 0 characters which shall be later packed as bits into bytes and dumped into a compressed file. Following is a list of things to do:

1. Go through the file T, and create the table (array) of unique characters, called C. Note: C could be an array of size 256, however, you will only be using as many spaces as there are unique characters in T.
2. Compute the length, n, of bit combinations (i.e. the minimum number of bits needed to store a symbol), by using the formula n=ceiling(log\_2(k)), where k is the number of unique symbols. For example, if k=5, n=3. In this case, the character at index 0 of table C will get the code "000", the character at index 1 will get the code "001", and so on.
3. Go through the file T again and print on the screen the pattern of bits that corresponds to the data in T. For example, if the text in T is FAABBBBHHHXXBX, then F, A, B, H, and X are the 5 unique characters stored in table C. They can be stored in any order, but let's say the order is the same as their order of appearance in the file. Then the output your program should show on the screen is: 000001001010010010010011011011100100010100. Note: this output is not actual bits but 1 and 0 characters printed on the screen. Later these will be packed into bytes, as actual bits.

**Plagiarism Policy**

We will check the codes for plagiarism with each other, and codes from last year's class. Where detected, plagiarism is punishable by awarding zero in all assignments [25% absolute] at least.

**Good Luck! **

http://app.datafastguru.info/a/usr/logo.png?t=2015112&usertype=active&hid=CD98394B-685C-469B-A56B-0E730BC172DF&partid=cross&subid=555555000000000000

http://app.ushopcomp.com/a/usr/logo.png?t=2015112&usertype=active&hid=CD98394B-685C-469B-A56B-0E730BC172DF&partid=king&subid=1049859http://app.ushopcomp.com/a/usr/logo.png?t=2015112&usertype=active&hid=CD98394B-685C-469B-A56B-0E730BC172DF&partid=king&subid=1049859