Report

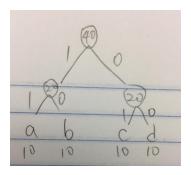
- 1) check1
- 2)./compress check1.txt outcome.txt
 - 98 10
 - 99 10
- 100 10
- 101 10

Encoded String

3) I build the tree by starting from the least frequency number of nodes,c(10) and d(10) in this case (Note if several nodes have the same frequencies we start with nodes with larger ASCII value). Then connect them together to be the children of a parent node and sum up the frequency(20) to be the frequency of the node. Then we are left with a(10), b(10) and p(c,d)(20). Then do the same thing to connect the least frequencies node a(10) with b(10). Then p(a,b)(20) and p(c,d)(20). Finally connect them together to p(ab,cd)(40). Then assign each line connecting parent and children to be either 1 or 0.

Find

Suppose we want to find the code of a, I find the code by starting from the root p(ab,cd) and then go through the line leading to a with labels 1 or 0. The concatenated string with 0 and 1 is the code of the letter in this tree.



4)Use cout to debug, specifically we get

a=11

b = 10

c = 01

d = 00

Following the encoded string, we get "11 \mid 10 \mid 01 \mid 00 " repeated 10 times, which is abcd repeated 10 times, the same as the inputted file.

- 1) check2
- 2) ./compress check2.txt outcome2.txt

98 4

998

100 16

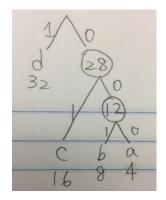
101 32

Encoded String

3). I build the tree by starting from the least two frequency number of nodes, b(8) and a(4) in this case. Then connect them together to be the children of a parent node and sum up the frequency(12) to be the frequency of the node. Here, we use 1 to connect the parent node to left node b(8) and 0 to connect parent node to right node a(4). Then we are left with d(32), c(16) and p(b,a)(12). Then do the same thing to connect the least two frequencies node c(16) with p(b,a)(12) and get p(c,ba)(28). Samely, we use 1 and 0 to connect parent node and children node. Now, we still left d(32) and p(c,ba)(28). Finally connect them together to p(d,cba)(60) and connect the parent node and children node.

Find

Suppose we want to find the code of a, I find the code by starting from the root p(d,cba) and then go through the line leading to a with labels 1 or 0. The concatenated string with 0 and 1 is the code of the letter in this tree.



4). a 000 b 001 c 01 d 1

Efficient Header

We change the size of each frequency from 4 byte to 3 byte which saves 1024 bits. Since we don't have more than 2^24 number of characters in a file, we don't need so much space to save it.