CSE 326: Data Structures
Assignment #6
November 22, 2004
due: Monday, December 6

The purpose of this programming assignment is to figure out exactly how to use priority queues in implementing Huffman trees.

1. Implement the PriorityQueue class using heaps. You may assume that the heap never has more than 500 entries. You may assume that the Key field has type `double`, and the Info field is a pointer to a binary tree. DeleteMin should return a record containing both the minimum Key and its associated Info field.

2. Implement the following procedures that do Huffman encoding and decoding:

   (a) BuildHuffman (described in the paragraph starting on page 146 and ending on page 147): this reads a file of character-frequency pairs, and returns a Huffman decoding tree and encoding table built from them. Note that the leaves of your tree must each store a `char`. Unlike Figure 5.8, frequencies do not need to be stored in the tree. (Why?)

   Unlike the book's description, you should assume that the frequencies have type `double`: character frequencies are often given as percentages. Figure 1 gives the character-frequency pairs for you to use.

   These are the only 28 characters you need handle, except that your program should allow upper case and lower case letters to be input interchangeably. Your program must read these character-frequency pairs from a file, because in practice these frequencies might be changed later, or the set of characters might be changed. Your program may assume that there will never be more than 500 character-frequency pairs. There is a file containing the pairs from Figure 1 linked from the course web page so that you don't have to type them in yourself.

   In addition to the decoding tree, BuildHuffman should return an encoding table, which is indexed by characters and whose entries give the encodings of those characters. This table can be constructed by a single traversal of the decoding tree.

   (b) TreeEncode (described in the first two sentences of the first full paragraph on page 146): this takes as inputs an encoding table and a file w of characters, and outputs the file of bits encoding w.

   (c) TreeDecode (described in Algorithm 5.2 on page 147): this takes as inputs a decoding tree and a file b of bits, and outputs the character file decoding b.

As described above, you will use character files (for the decoded string) and bit files (for the encoded string) as inputs and outputs. Character files are easy and standard. Bit files are slightly more complicated. Since the whole point of encoding is to compress characters, in practice it would make no sense if your TreeEncode function outputs the chars '0' and '1', since this would use 8 bits for each bit in the encoding. But for simplicity in this assignment, go ahead and use character files in place of bit files, but understand that this is not what you would really do in practice.

Instructions for your driver will follow by electronic broadcast.

‘ ’	.200
‘.’	.010
‘a’	.073
‘b’	.009
‘c’	.030
‘d’	.044
‘e’	.130
‘f’	.028
‘g’	.016
‘h’	.035
‘i’	.074
‘j’	.002
‘k’	.003
‘l’	.035
‘m’	.025
‘n’	.078
‘o’	.074
‘p’	.027
‘q’	.003
‘r’	.077
‘s’	.063
‘t’	.093
‘u’	.027
‘v’	.013
‘w’	.016
‘x’	.005
‘y’	.019
‘z’	.001

Figure 1: Common English Character Frequencies