CS494 Take Home Quiz #6 – The Beale Treasure

" I have deposited in the county of Bedford, (Virginia) about four miles from Buford, in an excavation or vault, six feet below the surface of the ground, the following articles belonging jointly to the parties whose names are given in number three herewith. The first deposit consisted of ten hundred and fourteen pounds of gold and thirty eight hundred and twelve pounds of silver deposited Nov. Eighteen Nineteen. The second was made Dec. Eighteen Twenty one and consisted of nineteen hundred and eighty eight of silver, also jewels obtained in St. Louis in exchange to save transportation and valued at thirteen thousand dollars.

The above is securely packed in iron pots with iron covers the vault is roughly lined with stone and the vessels rest on solid stone and are covered with others. Paper number one describes the exact locality of the vault so that no difficulty will be had in finding it."

About 100 years ago, a fellow by the name of Beale supposedly buried two wagons-full of silver-coin filled pots in Bedford County, near Roanoke. There are local rumors about the treasure being buried near Bedford Lake.

He wrote three encoded letters telling what was buried, where it was buried, and who it belonged to. He entrusted these three letters to a friend and went west. He was never heard from again.
Several years later, someone examined the letters and was able to break the code used in the second letter. The code used the text from the Declaration of Independence. A number in the letter indicated which word in the document was to be used. The first letter of that word replaced the number. For example, if the first three words of the document were "We hold these truths", the number 3 in the letter would represent the letter t (the first letter in “these”).

One of the remaining letters supposedly contains directions on how to find the treasure. To date, no one has solved the code. It is believed that both of the remaining letters are encoded using either the same document in a different way, or another very public document.

(The previous two paragraphs have been excerpted from from
http://einstein.et.tudelft.nl/~arlet/puzzles/sol.cgi/cryptology/Beale)

The type of encryption used in the Beale documents is known as a *book cipher*. A book cipher uses a *key document* for encoding and decoding. For part II of the Beale ciphers, the key document is the Declaration of Independence. Each word in the declaration in is numbered beginning with the number 1. We can then encrypt a document using the following algorithm.

For each letter in the document that we are encrypting
  Find a word in the key document that starts with this letter
  Write the word’s number to the encrypted file
  Mark the word so that it will not be used again

This process will transform a document to a sequence of numbers. Each number represents a single letter.

For example, from the first words in the Declaration of Independence we have:

When(1) in(2) the(3) course(4) of(5) human(6) events(7) it(8) becomes(9) necessary(10) for(11) one(12) people(13) to(14) dissolve(15) the(16) political(17) bands(18) which(19) have(20) connected(21) them(22) with(23) another(24)

We can encode the message:
"Watch it" as 19 24 22 4 6 2 3

Notice that there is no code for spaces so word boundaries are lost.

We can decode the document by replacing each number with the first letter of the corresponding word in the key document. Word 19 is "which" which gives us "W." Word 24 is "another" which gives us "a," and so on.

We can make the book cipher a little harder to crack by adding a *word offset*. With a document offset we start numbering the words in the key document with a number other than 1. Here is our key document with a word offset of 11.

When(11) in(12) the(13) course(14) of(15) human(16) events(17) it(18) becomes(19) necessary(20) for(21) one(22) people(23) to(24) dissolve(25) the(26) political(27) bands(28) which(29) have(30) connected(31) them(32) with(33) another(34)

With a word offset of 11, we can encode the message:
"Watch it" as 29 34 32 14 16 12 13

We can further confuse our adversaries by adding a *character offset*. The character offset allows us to specify a letter other than the first letter as the one used for encoding and decoding. With a word offset of 11 and a character offset of 1 we can encode the message:
"Hi Ho" as 11 33 26 21

If the character offset is larger than the number of letters in the word we *wrap-around* and start back at the beginning of the word. Thus "the" with a character offset of 3 would produce "t," 4 would produce "h," and 5 would produce "e."

The Assignment

Write a program that can encode and decode documents using a book cipher. You will write this assignment without using any explicit loops at all. This is not necessarily a good idea, but I want you to try it for this program. You will rely on the standard template library and its algorithms to provide iteration.

Test you program using the file containing the Declaration of Independence (doi.txt) and the Beale text part II (beale2.txt). Only part II can be deciphered using the Declaration.

Your program must:

1.) Read and store the words in the key document
2.) Decode a file using the key (read numbers from a file, write characters to a file)
3.) Encode a document (read strings from a file, write numbers to a file)

Decoding a Document

Read the number in the inputFile and write the text to the outputFile. A parameter called wordOffset provides the word to start counting from. A parameter called charOffset provides the letter to use for each word (e.g., 0 for 1st letter, 1 for //2nd etc.) If it is not possible to decode a letter, write a question mark.

Encoding a Document
Read the text in the inputFile and write the sequence of integers to the outputFile. The wordOffset provides the word to start counting from. The charOffset provides the letter offset for each word (0 is the first letter).

Use a random number generator to pick a place to start looking in the key document. Make sure that you never reuse the same number in the output file. If it is not possible to generate a code for a particular letter, write 0.

Use your classes and the Declaration of Independence to decode the Beale ciphers.

The Beale ciphers consist of three parts.

Part II http://lasi.lynchburg.edu/ribler_r/public/cs493/beale2.txt can be decoded using the Declaration of Independence. A file containing the Declaration of Independence in a text format without punctuation is available at http://lasi.lynchburg.edu/ribler_r/public/cs493/doi.txt You will want to test your code using Part II and the Declaration. This will create an almost correct translation (there will be a few incorrect/missing characters.)

Parts II and III have never been successfully decoded! Break these codes and you can become rich and famous!

The relevant text files are available at http://lasi.lynchburg.edu/ribler_r/public/cs493/