Homework 1 - CSCI 181

1. (10 pts) The following ciphertext is encrypted using a monoalphabetic substitution cipher. Use the frequency analysis technique described in the lecture to decrypt this. Show the frequency of letters in the ciphertext and show your work. You can use different tools (with citing the source) to find the frequency of letters, or write your own code. But you are not allowed to brute force the ciphertext for finding the plaintext. It is important to show your work. If you only submit the final plaintext no points will be given. Explain your thought process as you try to decrypt this. This ciphertext is included in file **hw1q1cipher.txt**.

   ZIW VQKD LAFLIOFW YOSZWKWR ZIKGAUI ZIW SWQCWL GY ZIW ZKWWL EQLZOFU RQHHSWR LIQRGVL GF ZIW UKGAFR TWSGV TOKRL EIOKHWR QFR LQFU OF ZIW TKQFEIWL QRROFU ZG ZIW HWQEWYAS QDTOQFEW GY ZIW YGKWLZ Q SOUIZ TKWWMW KALZSWR ZIW SWQCWL EKWQZOFU Q UWFZSW VIOLHWKOFU LGAFR ZIQZ YOSSWR ZIW QOK

2. (10 pts) Write a program in C/C++ or Python3 that does encryption and decryption for Vigenere cipher. This should include two functions, one for encryption and one for decryption. The inputs to the encryption function is the plaintext and the keyword. The input to the decryption function is the ciphertext and the keyword. Submit your code. Test it to make sure it works properly. The input ciphertext and plaintext should not have space characters in them. A ciphertext file is included (**vigenere-cipher.txt**) to show you the format of the ciphertext. This is the example in the lecture notes.

3. (30 pts) The ciphertext provided in the file **hw1q3-cipher.txt** was encrypted using the Vigenere cipher. Determine the length of the keyword. First use the Kasiski method to make a conjecture about the key length by finding several trigraphs and factoring the distances between them. You can use this website to find the common trigraphs in the ciphertext and then find the distances. Assuming k is the length of the keyword, now write a program that finds the frequency of letters A... Z for every letters in positions 0, k, $2k$, etc. in the ciphertext. This is exactly what we did in the lecture during the cryptanalysis. The output looks like this for example: $[10, 0, 0, 1, 1, 3, 7, 0, 0, 5, 7, 3, 2, 2, 0, 0, 1, 0, 4, 1, 2, 3, 10, 0, 1, 6]$, which means that letter A appeared 10 times, letter B appeared 0 times and ... . These numbers are different from the numbers you will get. You should have k output vectors of length 26. The second shows frequency of [A,...,Z] of the letters appearing in positions 1, $k+1$, $2k+1$, etc. The third shows the frequency of the letters appearing in positions 2, $k+2$, etc. Now write down those histograms in your answer sheet and underline every number that is 4 or more. Use the fact that the distances between A, E, T and A in the alphabet are 4, 15, and 7 to decide how much each shift was. Write down all possible shifts for each histogram. Determine the keyword. You can test your work by finding the keyword and using it to decrypt the ciphertext using the code in question 2.

   It is really important to show all your work. Show the vectors you have found, show all possible shifts, show what is the key, and finally the plaintext. Your program needs to do exactly what is asked in the question.