Caesar Cryptananlysis and Decryption

Natprawee Pattayawij

27 November 2020

Contents

1 Introduction

Caesar cipher is one of the simplest encryption techniques. To encrypt, it shifts each alphabet in a text by a value k. To illustrate, with k=5, all ”a” are changed to ”f”, all ”b” are changed to ”g”, and so on [1]. Therefore, to decrypt the ciphered text the value k is needed to know how to shift back the ciphered text to the plain text.

However, some ciphered text might not attach the value k, for it might be too easy to decrypt. Thus, cryptanalysis has to be considered. Cryptanalysis of the Caesar cipher in English is not too hard; the average frequencies of alphabets is used. ”e” letter is the most frequent alphabet in English [2], so, to calculate the value k, the most frequent letter in the ciphered text is observed, and the difference between its index and the index of ”e” letter (that is 4) is considered as the value k. This method is applicable in the text that is long enough, as the frequency of each letter hardly changes when adding a few letters.

With cryptanalysis and decryption methods of the Caesar cipher, This program aims to decipher the Caesar ciphered text without knowing the value k[1].

2 Cryptanalysis

Firstly, to observe the most frequent letters in the text needs to encode the text into ASCII codes. Then, .count() is used in for loop from the ASCII code of ”a”

[1] Only lower-case letters are applicable.

to the ASCII code of ”z” to count the frequency of each letter (Only lower-case letters are counted.), and all frequencies are appended in list. Next, .index() and .max() are used to observe which ASCII code has the most frequency and what is the index of it. Since the most frequent letter is considered as ”e” letter index, so the difference of the index of it minus by the index of ”e” (that is 4) is the value k. Therefore, the value k could be used to decrypt in Section 3 Decryption.

3 Decryption

To decrypt the ciphered text, there are 2 inputs needed, the ciphered text and the value k. The ciphered text is input from the user, and the value k is obtained from Section 2 Cryptanalysis. To decrypt the text, first of all, the ciphered text is encoded into a list of ASCII codes. Then, to shift all letters back by the k position, the applicable ASCII codes (only lower-case) minus 19 because firstly the ASCII of ”a” mod by 26 is 19. Next, the result minus k, apply mod to shift them back and then plus 97 to set them back to the interval of lower-case letters' ASCII codes. Finally, the deciphered plain text is the output.

4 Suggestion

To make it do less calculation, the ASCII encoding part in both sections can be used once. Moreover, it can be improved to be applicable for upper-case letters as well.

References

[1] Dennis Luciano and Gordon Prichett. Cryptology: From caesar ciphers to public-key cryptosystems. *The College Mathematics Journal*, 18(1):2–17, Jan 1987.

[2] Marsha Lynn Moreno. Frequency analysis in light of language innovation. *Math*, 187, 2005.