Computer Network & Security - Assignment #2
Breaking Vignere Cipher

Harsh Agarwal
CS15BTECH11019

Decrypted Text

We the people of india having solemnly resolved to constitute india into a sovereign socialist secular democratic republic and to secure to all its citizens justice social economic and political liberty of thought expression belief faith and worship equality of status and of opportunity and to promote among them all fraternity assuring the dignity of the individual and the unity and integrity of the nation in our constituent assembly this twenty sixth day of november nineteen forty nine do hereby adopt enact and give to ourselves this constitution

Key

yjclh (if numbering starts from 1)
xibkg (if numbering starts from 0)

Procedure

Finding the key length

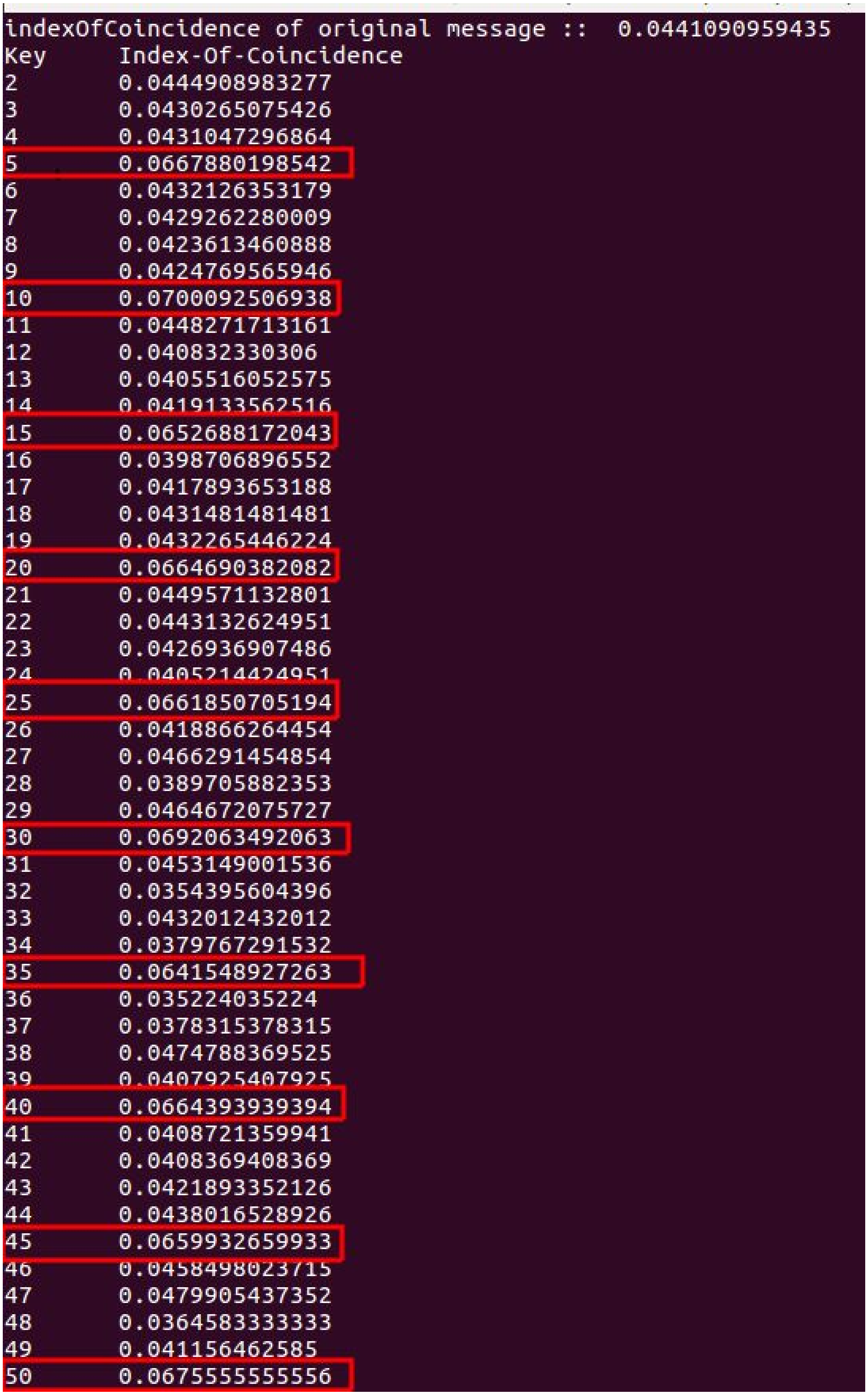
Acc to section 3.3 of the book “Cryptography” by David R. Kohel , the index of coincidence of a text space (e.g. that of all plaintext or ciphertext) is defined to be the probability that two randomly chosen characters are equal.

By iterating over the length of key from 2 - 25 , we found the index of coincidence for the ciphertext for a given key length, according the following formula

$$\sum_{i=1}^{26} \frac{n_i * (n_i - 1)}{N * (N-1)}$$

Where N is the total length of the cipher & n_i is the number of occurrences of the character i.

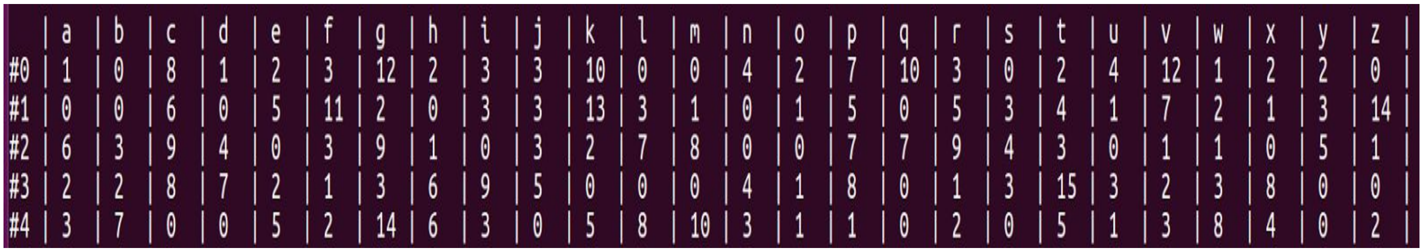
Iterating over 2-50 , we initially extracted the corresponding string in the ciphertext which corresponds to each index on the key. Hence we fill get no_of_groups as the key length.
For each group we find the index of coincidence and finally get the mean for the index of coincidence all the 5 groups
We can observe that in general almost every value is ~ 0.44 , but all the multiples of 5 have value ~ 0.66.
Since index of coincidence of English Distribution in English Language is ~ 0.67 , we can say with certainty that key length is 5.
I have attached a screenshot of the index of coincidences for various key lengths



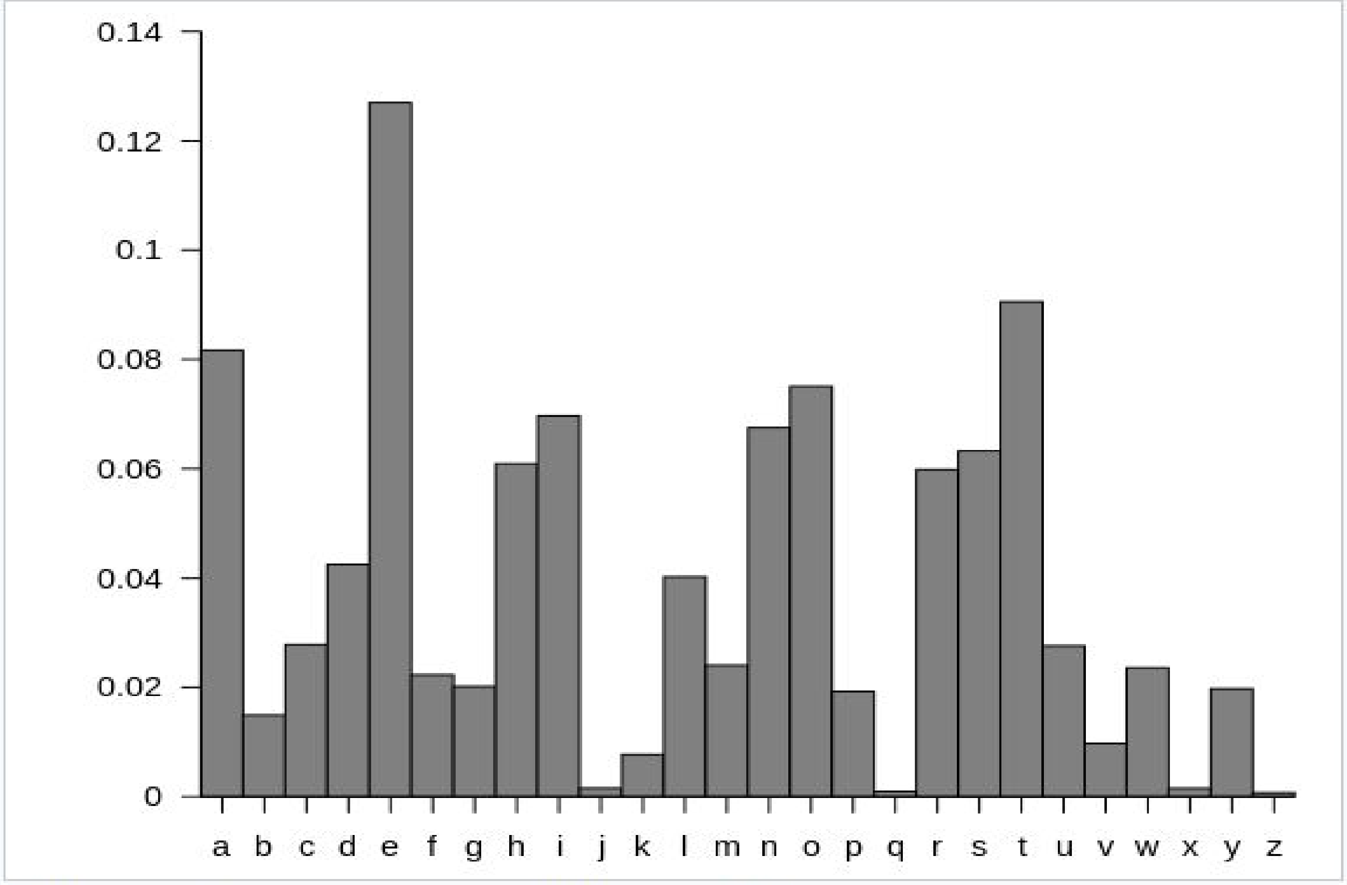
```
indexOfCoincidence of original message ::  0.0441090959435
Key      Index-Of-Coincidence
2        0.0444908983277
3        0.0430265075426
4        0.0431047296864
5        0.0667880198542
6        0.0432126353179
7        0.0429262280009
8        0.0423613460888
9        0.0424769565946
10       0.0700092506938
11       0.0448271713161
12       0.040832330306
13       0.0405516052575
14       0.0419133562516
15       0.0652688172043
16       0.0398706896552
17       0.0417893653188
18       0.0431481481481
19       0.0432265446224
20       0.0664690382082
21       0.0449571132801
22       0.0443132624951
23       0.0426936907486
24       0.0405214424951
25       0.0661850705194
26       0.0418866264454
27       0.0466291454854
28       0.0389705882353
29       0.0464672075727
30       0.0692063492063
31       0.0453149001536
32       0.0354395604396
33       0.0432012432012
34       0.0379767291532
35       0.0641548927263
36       0.035224035224
37       0.0378315378315
38       0.0474788369525
39       0.0407925407925
40       0.0664393939394
41       0.0408721359941
42       0.0408369408369
43       0.0421893352126
44       0.0438016528926
45       0.0659932659933
46       0.0458498023715
47       0.0479905437352
48       0.0364583333333
49       0.041156462585
50       0.0675555555556
```

Guessing Message from a key of length 5

After determining that the key length is 5 , we found the frequencies of character in each of the 5 groups(caesar ciphers) of characters that will be made.



```
   | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
#0 | 1 | 0 | 8 | 1 | 2 | 3 | 12| 2 | 3 | 3 | 10| 0 | 0 | 4 | 2 | 7 | 10| 3 | 0 | 2 | 4 | 12| 1 | 2 | 2 | 0 |
#1 | 0 | 0 | 6 | 0 | 5 | 11| 2 | 0 | 3 | 3 | 13| 3 | 1 | 0 | 1 | 5 | 0 | 5 | 3 | 4 | 1 | 7 | 2 | 1 | 3 | 14|
#2 | 6 | 3 | 9 | 4 | 0 | 3 | 9 | 1 | 0 | 3 | 2 | 7 | 8 | 0 | 0 | 7 | 7 | 9 | 4 | 3 | 0 | 1 | 1 | 0 | 5 | 1 |
#3 | 2 | 2 | 8 | 7 | 2 | 1 | 3 | 6 | 9 | 5 | 0 | 0 | 0 | 4 | 1 | 8 | 0 | 1 | 3 | 15| 3 | 2 | 3 | 8 | 0 | 0 |
#4 | 3 | 7 | 0 | 0 | 5 | 2 | 14| 6 | 3 | 0 | 5 | 8 | 10| 3 | 1 | 1 | 0 | 2 | 0 | 5 | 1 | 3 | 8 | 4 | 0 | 2 |
```



Relative frequencies of letters in text.

Index of distribution of each alphabet is

Alphabet #0 = 0.068862960421

Alphabet #1 = 0.073165030388

Alphabet #2 = 0.0582047685835

Alphabet #3 = 0.0684899485741

Alphabet #4 = 0.0652173913043

Since index of coincidence of English Distribution in English Language is ~ 0.67, the closest alphabet to this is Alphabet #3.

Alphabet #3

In this we observe that 15 is repeated maximum times and there is no other character that is repeated that much. Hence we can fairly say that in Alphabet #3, there is a shifting of 11 and e has been mapped as t.

Alphabet #0

Next we tackle Assignment 1 because after #3 , #0 is the closest to 0.647

In this we observe that all alphabets with larger count i.e.
C -> 8 , K -> 10 , Q -> 10 , V -> 12 , G -> 12
are spaced as vowels are in the English alphabet. Acc to Relative Frequencies of characters above , vowels have more frequency as compared to consonants. Hence we can assume that most of the characters with larger count are in fact vowels.
Then we try different possible combinations and eliminate the ones which seem implausible.

Hypothesis :: g in ciphertext is actually a in plaintext
Contradiction :: In this scenario c in ciphertext is assigned to w in plaintext , with a frequency count as 8. Since relative frequency of w is small as compared to other alphabets, this hypothesis seems incorrect

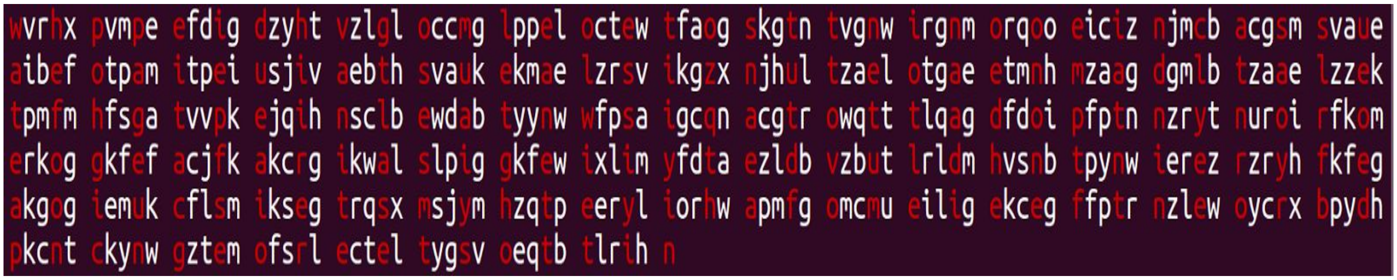
Hypothesis :: g in ciphertext is actually i in plaintext
Contradiction :: In this scenario v in ciphertext is assigned to x in plaintext , with a frequency count as 12. Since relative frequency of x is small as compared to other alphabets, this hypothesis seems incorrect

Hypothesis :: g in ciphertext is actually o in plaintext
Contradiction :: In this scenario q in ciphertext is assigned to y in plaintext , with a frequency count as 10. Since relative frequency of y is small as compared to other alphabets, this hypothesis seems incorrect

Hypothesis :: g in ciphertext is actually u in plaintext
Contradiction :: In this scenario k in ciphertext is assigned to y in plaintext , with a frequency count as 10. Since relative frequency of w is small as compared to other alphabets, this hypothesis seems incorrect.

Hypothesis :: g in ciphertext is actually e in plaintext
Verification :: In this scenario , all the vowels in plaintext are assigned the most occuring characters in ciphertext. Hence, this hypothesis seems correct.

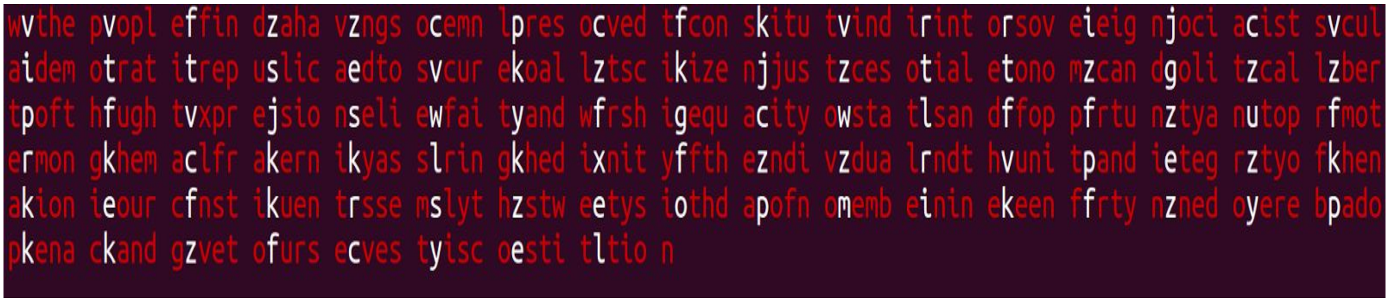
After decoding #0 & #3 we get the following ciphertext



```
wvrhx pvmpe efdig dzyht vzlgl occmg lppel octew tfaog skgtn tvgnw irgnm orqoo eiciz njmcb acgsm svaue
aibef otpam itpei usjiv aebth svauk ekmae lzrsv ikgzx njhul tzael otgae etmnh mzaag dgmlb tzaae lzzek
tpmfm hfsga tvvpk ejqih nsclb ewdab tyynw wfpsa igcqn acgtr owqtt tlqag dfdoi pfptn nzryt nuroi rfkom
erkog gkfef acjfk akcrg ikwal slpig gkfew ixlim yfdta ezldb vzbut lrldm hvsnb tpynw ierez rzryh fkfeg
akgog iemuk cflsm ikseg trqsx msjym hzqtp eeryl iorhw apmfg omcmu eilig ekceg ffptr nzlew oycrx bpydh
pkcnt ckynw gztem ofsrl ectel tygsv oeqtb tlrih n
```

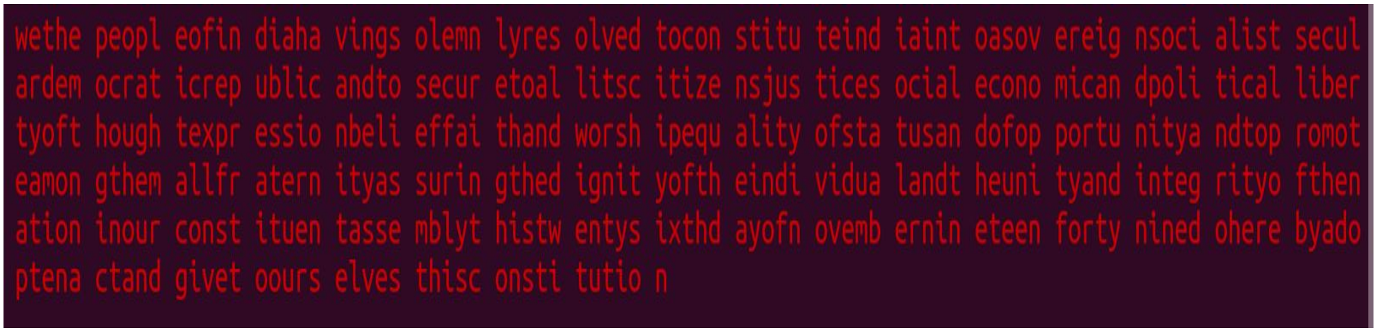
The red signify the characters that have been shifted to their proper position.
If we observe the last few characters , we see that there is i_n.
On seeing this we can assume that the word involves **tion** because tion always comes at the end of a word and no other substring involving i_n seems suitable at this position.

Going with this hypothesis, we get the following ciphertext



```
wvthe pvopl effin dzaha vzngs ocemn lpres ocved tfcon skitu tvind irint orsov eieig njoci acist svcul
aidem otrat itrep uslic aedto svcur ekoal lztsc ikize njjus tzces otial etono mzcan dgoli tzcal lzber
tpoft hfugh tvxpr ejsio nseli ewfai tyand wfrsh igequ acity owsta tlsan dffop pfrtu nztya nutop rfmot
ermon gkhem aclfr akern ikyas slrin gkhed ixnit yffth ezndi vzdua lrndt hvuni tpand ieteg rztyo fkhen
akion ieour cfnst ikuen trsse mslyt hzstw eetys iothd apofn omemb einin ekeen ffrty nzned oyere bpado
pkena ckand gzvet ofurs ecves tyisc oesti tltio n
```

Now we can safely state that in alphabet #1 , v in ciphertext is translated to e in plaintext.



```
wethe peopl eofin diaha vings olemn lyres olved tocon stitu teind iaint oasov ereig nsoci alist secul
ardem ocrat icrep ublic andto secur etoal litsc itize nsjus tices ocial econo mican dpoli tical liber
tyoft hough texpr essio nbeli effai thand worsh ipequ ality ofsta tusan dofop portu nitya ndtop romot
eamon gthem allfr atern ityas surin gthed ignit yofth eindi vidua landt heuni tyand integ rityo fthen
ation inour const ituen tasse mblyt histw entys ixthd ayofn ovemb ernin eteen forty nined ohere byado
ptena ctand givet oours elves thisc onsti tutio n
```

Hence we break the Vigenere cipher in this way.

References

http://iml.univ-mrs.fr/~kohel/tch/M2-CryptoSymetrique/crypto.pdf
https://www.youtube.com/watch?v=P3UA53VTQsg&t=3319s
http://nob.cs.ucdavis.edu/classes/ecs155-2013-04/extras/vigenere.html
https://www.nayuki.io/page/automatic-caesar-cipher-breaker-javascript