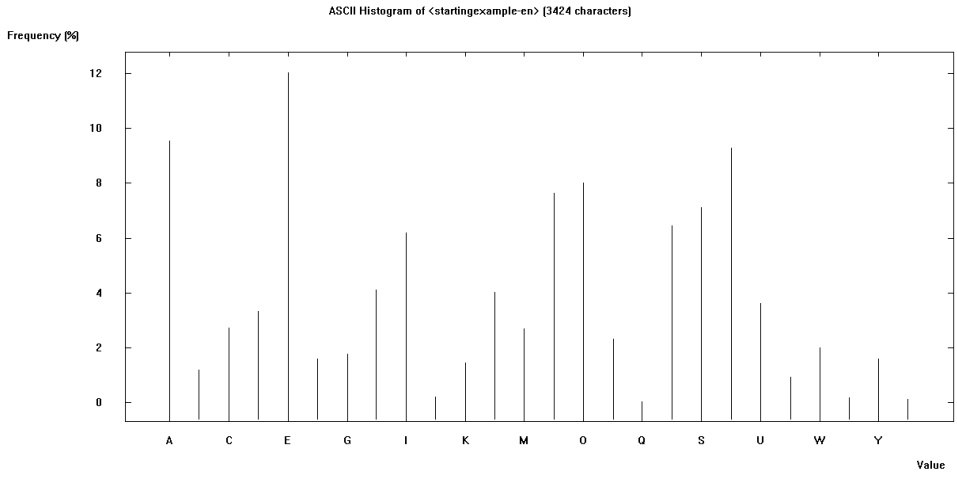
Η κρυπτογραφία υπάρχει από αρχαιοτάτων χρόνων καθώς η ανάγκη της προστασίας των δεδομένων συνοδεύει την ανθρωπότητα από την αρχαιότητα. Σήμερα αυτή η ανάγκη είναι επιτακτική και πορεύεται μαζί με την εξέλιξη της τεχνολογίας και κατ' επέκταση την αναγκαιότητα για ασφαλή επικοινωνία και προστασία των δεδομένων. Για αυτό το σκοπό χρησιμοποιούνται αλγόριθμοι κρυπτογράφησης. Από τις πιο απλές μορφές κρυπτογράφησης που είναι η μονοαλφαβητικές με τεχνικές αντικατάστασης – Caesar Cipher μέχρι τους γραμμικούς αλγόριθμους – Affine Cipher αλλά και πολυαλφαβητικής κρυπτογράφησης και αλγόριθμους Vigenere μέχρι τους σύγχρονους σημερινούς αλγόριθμος όπως AES η κρυπτογράφηση είναι απαραίτητο στοιχείο στην επικοινωνία – κωδικοποίηση δεδομένων. Αν και αυτοί αλγόριθμοι δεν χρησιμοποιούνται πια σήμερα στην κρυπτογράφηση αποσκοπούν στην εκπαίδευση την γενικότερη γνωριμία με την επιστήμη της κρυπτογράφησης.

Κρυπτογραφία - Κρυπτανάλυση

Β) 1)

- Η επιλογή κειμένου έγινε με βάση το κριτήριο ενός γενικού κειμένου, κατ' αυτό τον τρόπο επιλέχθηκε από τη σελίδα: https://theguardian.com/world/ από την επικαιρότητα. Ο τίτλος του κειμένου :***Trump praises North Korea's 'progress' as expert warns against optimism***
- Με τη βοήθεια του εργαλείου Cryptool – επιλέγοντας από το μενού Analysis-> Tools -> Histogram και N-gram παίρνουμε τα εξής αποτελέσματα:

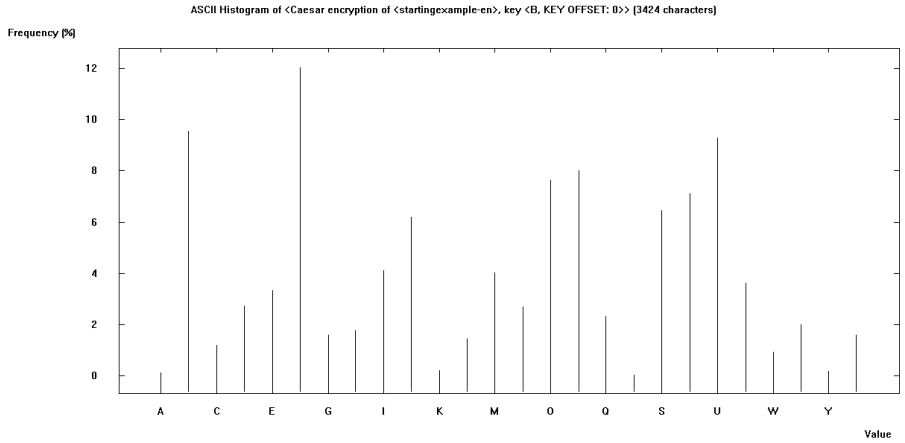


Εικόνα 1 – Histogram

Το ιστόγραμμα δηλαδή παρουσιάζει τις συχνότητες των γραμμάτων του κειμένου οι οποίες πλησιάζουν τις μετρήσεις που παρουσιάζονται στο English Letter Frequency (συχνότητα εμφάνισης γραμμάτων αγγλικού αλφαβήτου).
Παρατηρούμε ότι πρώτο σε συχνότητα είναι το γράμμα Έ ακολουθεί το Α το Τ όπου την μικρότερη εμφάνιση την έχει το γράμμα Z και το γράμμα J.
Αναλύοντας περαιτέρω στην επιλογή N-gram απεικονίζονται με λεπτομέρεια οι συχνότητες εμφάνισης των γραμμάτων του παραπάνω κειμένου. Έτσι το γράμμα Ε εμφανίζεται με συχνότητα 12.0327 το γράμμα Α με συχνότητα 9.5502 το Τ με συχνότητα 9.2874 και ούτω καθεξής μέχρι που φτάνουμε στο τέλος της λίστα συχνοτήτων και βλέπουμε ότι το γράμμα Z έχει συχνότητα εμφάνισης μόλις 0.0292.

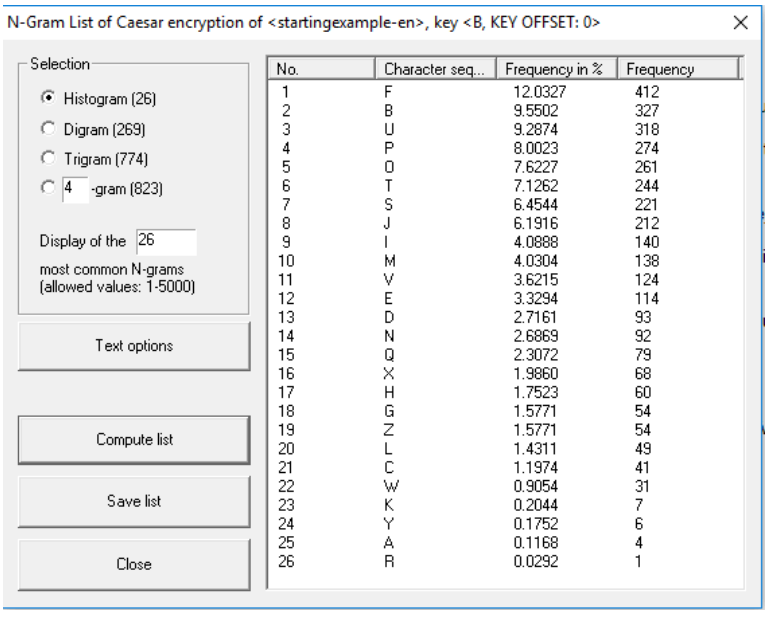
- Επιλέγω Encrypt/ Decrypt ύστερα κρυπτογράφηση με τον αλγόριθμο του Καίσαρα και παρατηρούμε πως μετά την κρυπτογράφηση το κείμενο αλλάζει σε κρυπτογραφημένο κάνοντας την αντίστοιχη ολίσθηση γραμμάτων.

Ακολουθούμε τα ίδια βήματα για την ανάλυση ιστόγραμμα και συχνοτήτων εμφάνισης γραμμάτων παρατηρούμε πως εμφανίζονται τα γράμματα με τη μεγαλύτερη συχνότητα στο κείμενο.



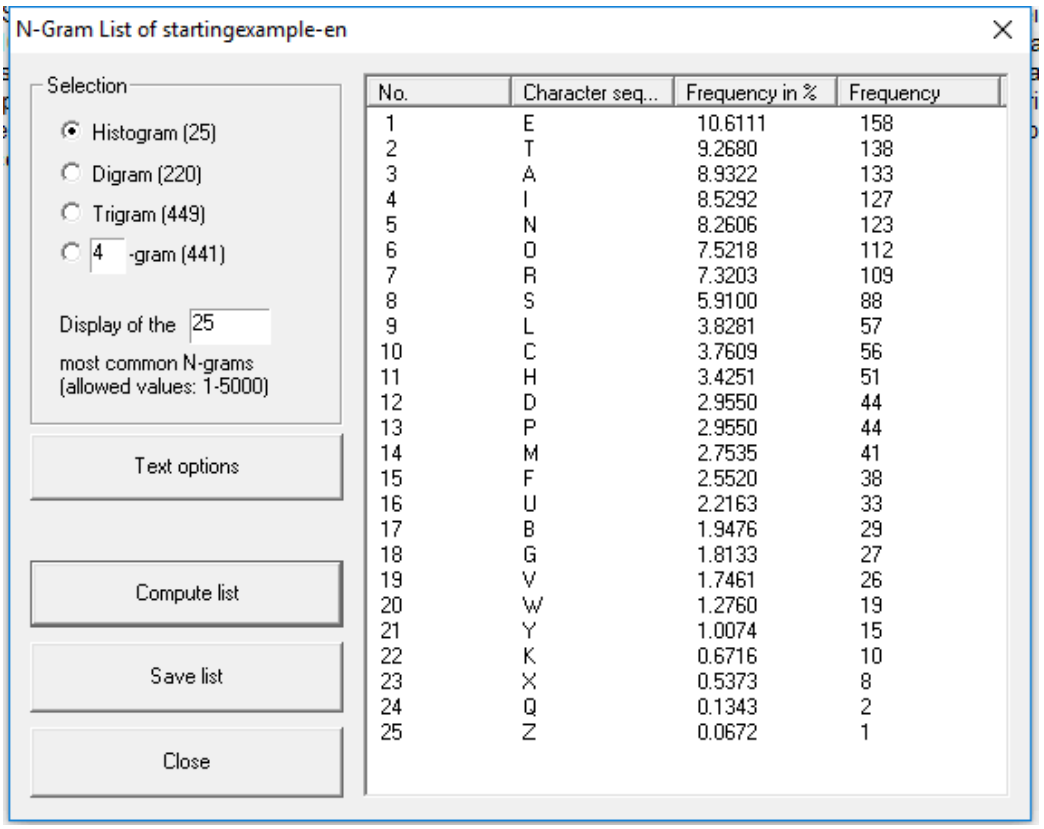
Εικόνα 2

Παρακάτω απεικονίζονται οι συχνότητες των γραμμάτων στο κρυπτογραφημένο κείμενο. Όπου πλέον τη θέση του γράμματος Ε από το πρωτότυπο κείμενο έχει πάρει το F στη δεύτερη θέση με συχνότητα εννέα.5502 είναι πλέον το B και στην τελευταία θέση βρίσκεται το γράμμα με συχνότητα 0.0292 το γράμμα δηλαδή που αντικατέστησε το γράμμα Z.



Εικόνα 3 – N-Gram

Β) 2)
Το δεύτερο κείμενο προς ανάλυση είναι ένα κείμενο αμιγώς τεχνικό με τίτλο :
A computional framework for conceptual blending
Από την ιστοσελίδα:
https://www.sciencedirect.com/science/article/pii/S000437021730142X
Στον κείμενο αυτό η συχνότητα εμφάνισης του γράμματος Έ που είναι και το πιο συχνό γράμμα της αγγλικής αλφαβήτου βρίσκεται στο 10.6111% ενώ το T 9.2680 το γράμμα Z που είναι και τελευταίο έχει συχνότητα εμφάνισης 0.0672.

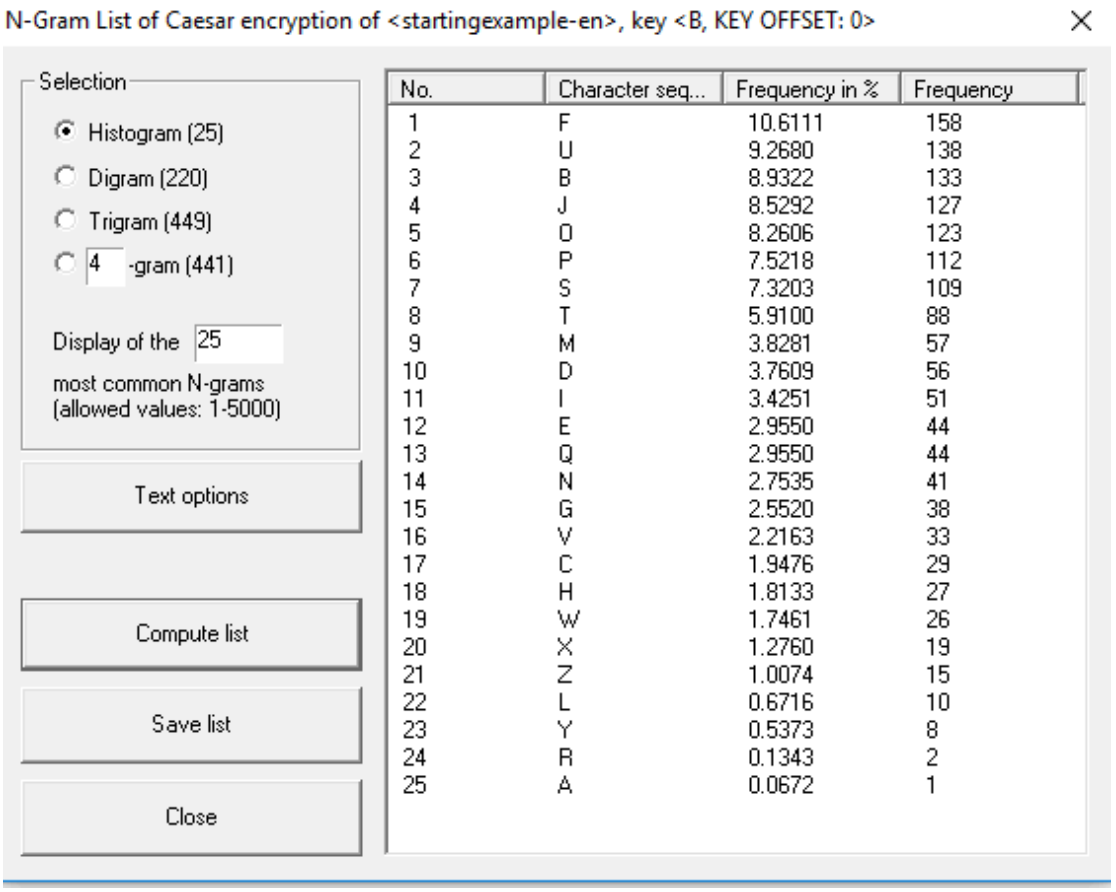


No.	Character seq...	Frequency in %	Frequency
1	E	10.6111	158
2	T	9.2680	138
3	A	8.9322	133
4	I	8.5292	127
5	N	8.2606	123
6	O	7.5218	112
7	R	7.3203	109
8	S	5.9100	88
9	L	3.8281	57
10	C	3.7609	56
11	H	3.4251	51
12	D	2.9550	44
13	P	2.9550	44
14	M	2.7535	41
15	F	2.5520	38
16	U	2.2163	33
17	B	1.9476	29
18	G	1.8133	27
19	V	1.7461	26
20	W	1.2760	19
21	Y	1.0074	15
22	K	0.6716	10
23	X	0.5373	8
24	Q	0.1343	2
25	Z	0.0672	1

Εικόνα 4

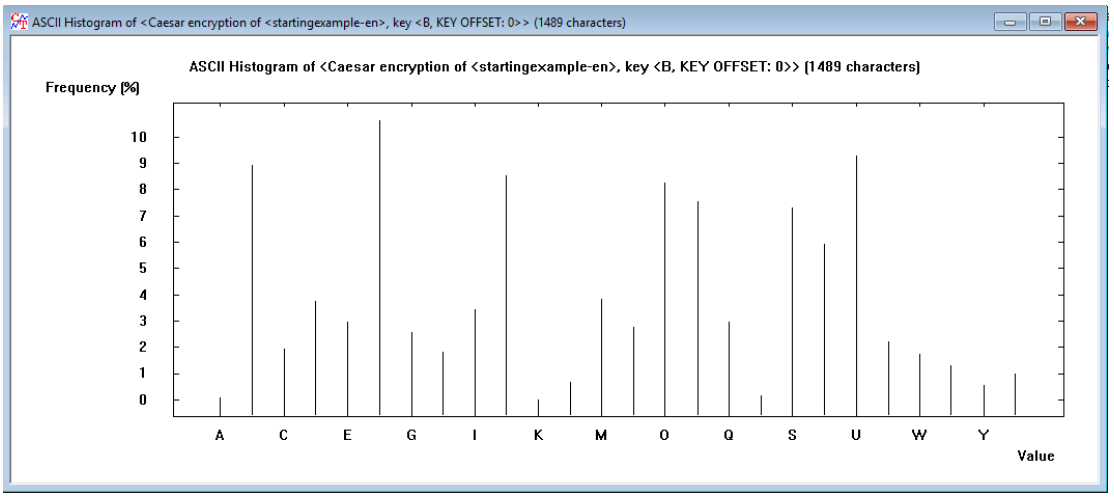
Μετά την κρυπτογράφηση του κείμενου η συχνότητα εμφάνισης γραμμάτων αλλάζει σύμφωνα με το κρυπτογραφημένο κείμενο και έτσι εμφανίζονται:
Συχνότητα στο γράμμα :

F – 10.6111
U – 9.2680
.....
R – 0.1343
A – 0.0672



Εικόνα 5

Το ιστόγραμμα που ακολουθεί δείχνει τα γράμματα με τη μεγαλύτερη συχνότητα και την μικρότερη συχνότητα σε αναλυτική διάταξη μετά τη χαρτογράφηση.



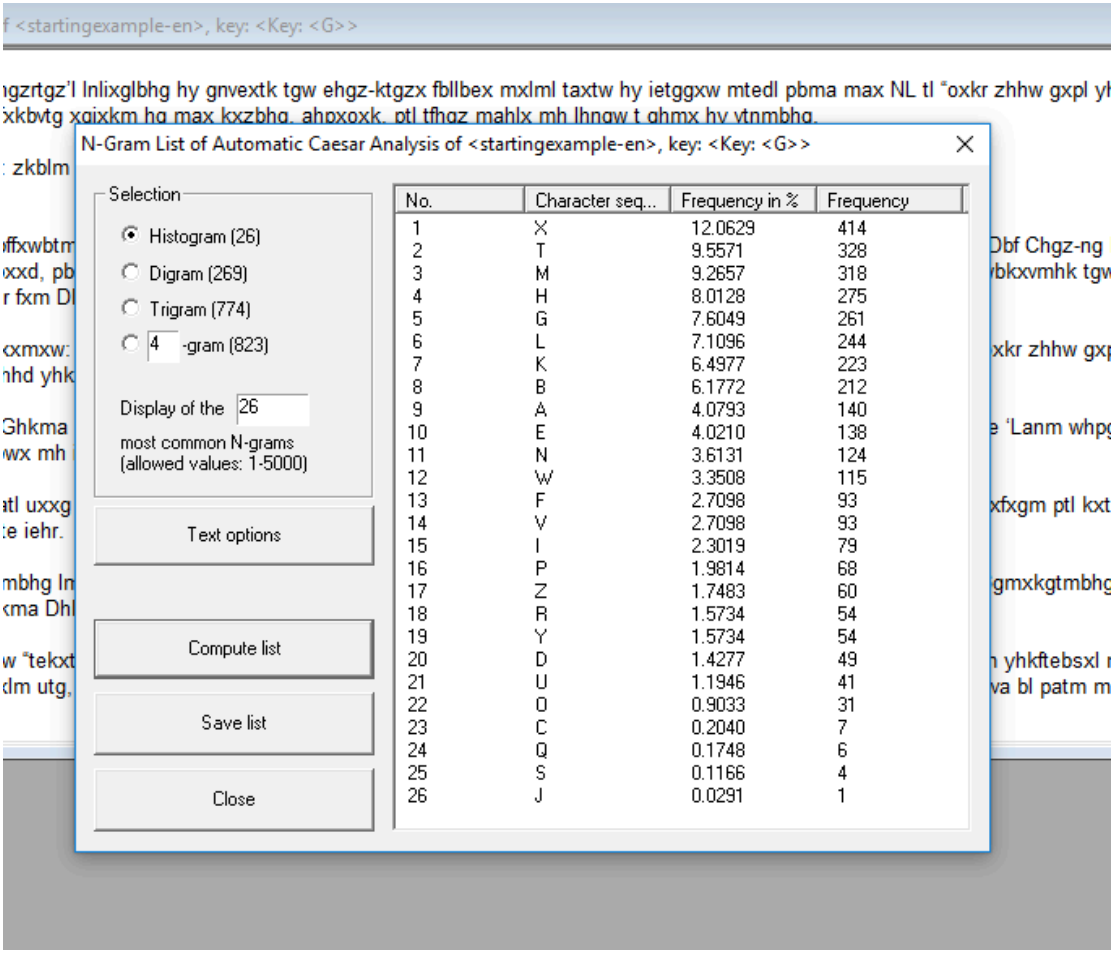
Εικόνα 6

Στο πρώτο κείμενο ουσιαστικά οι συχνότητες πλησιάζουν πάρα πολύ τις συχνότητες της αγγλικής αλφαβήτου – English Letter Frequency, στο δεύτερο κείμενο υπάρχει απόκλιση από τις συχνότητες . Εμφανίζεται δηλαδή διάφορα μεγάλη, στο δεύτερο κείμενο αφού το γράμμα Έ εμφανίζεται με συχνότητα 10.6911 και ούτω καθεξής καταγράφοντας διαφορές σε αρκετά γράμματα.

Γ)

- Το πρώτο κείμενο κρυπτογραφήθηκε με τον αλγόριθμο του Καίσαρα (Εικόνα 7) , όπου που σαν κλειδί χρησιμοποιήθηκε το γράμμα G της αγγλικής αλφαβήτου και στη συνέχεια με την επιλογή N-Gram από το CrypTool οι συχνότητες εμφανίζονται ιδίες με αυτές του αρχικού κειμένου μόνο που αυτή τη φορά αλλαγμένες ως προς τα γράμματα στα οποία αντιστοιχούν. Για παράδειγμα με συχνότητα του 12.0629 δεν έχουμε το Ε αλλά το γράμμα Χ. Μπορούμε εύκολα να συμπεράνουμε πως είναι αλγόριθμος μονοαλφαβητικής αντικατάστασης. Εάν εστιάζαμε στα διγράμματα θα μπορούσαμε εύκολα να συμπεράνουμε πως είναι αλγόριθμος ολίσθησης άρα Καίσαρα.

- Το δεύτερο κείμενο εμφανίζει συχνότητες κάπως "πειραγμένες" καθώς για παράδειγμα το γράμμα D έχει συχνότητα 14.4860 και ούτω καθεξής συμπεραίνουμε πως πιθανότατα το γράμμα D έχει αντικαταστήσει το γράμμα Έ της αλφαβήτου αλλά συμπεραίνουμε ότι έχουμε ολίσθηση επομένως και Καίσαρα.



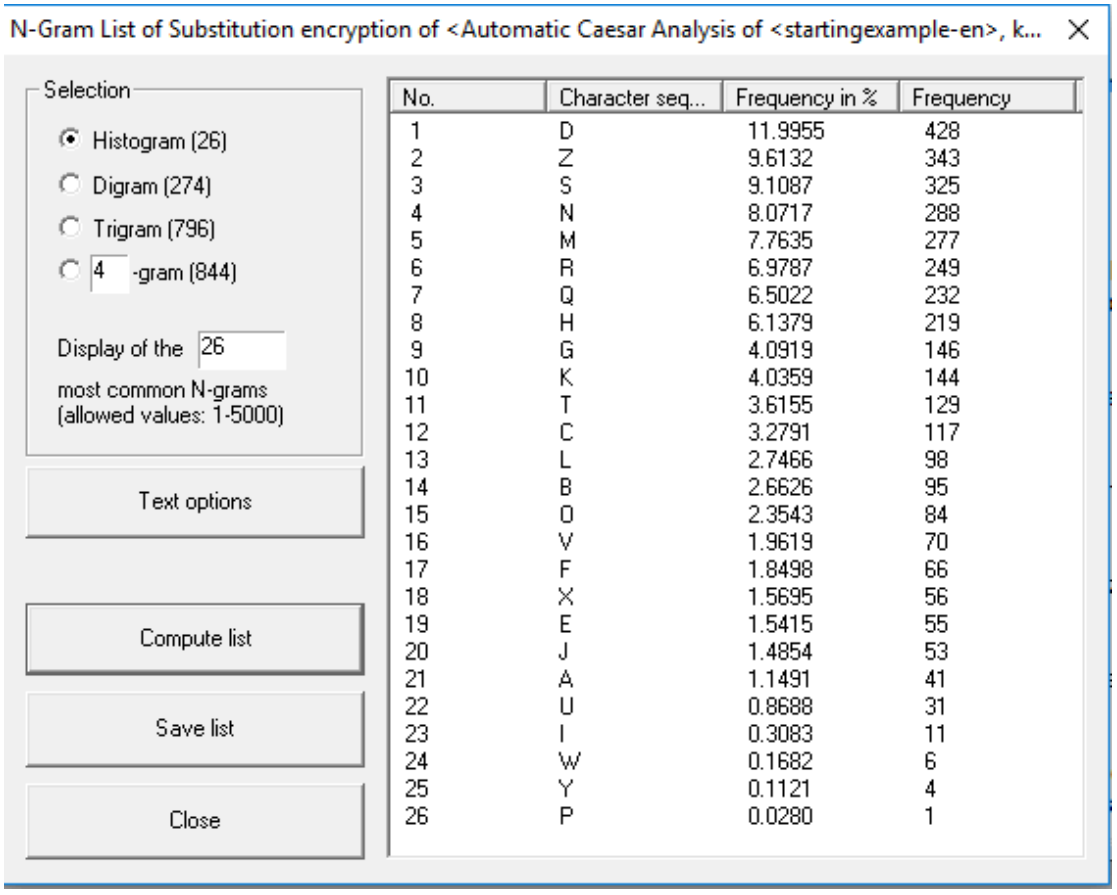
Εικόνα 7

- Για την κρυπτογράφηση απλής αντικατάστασης (*Εικόνα 8*) χρησιμοποιήθηκε το κλειδί Z στο πρώτο κείμενο.

Και πήραμε ως αποτέλεσμα από την ανάλυση του N-Gram της συχνότητες όπως ακολουθούν - Εικόνα 8. Όπου με συχνότητα 11.9955 εμφανίζεται πλέον το D και ακολουθούν όλα τα προγράμματα επομένως επειδή έχουμε συχνότητες κοντά σε αυτές της αγγλικού αλφαβήτου συμπεραίνουμε πως έχουμε αλγόριθμο απλής αντικατάστασης. Επειδή το γράμμα Z εμφανίζεται με συχνότητα 9.6132 άρα με συχνότητα πολύ μεγαλύτερη απ' αυτής της αγγλικής αλφαβήτου και εμφανίζεται συχνά τότε συμπεραίνουμε ότι είναι ένα τα πιθανά κλειδιά. Επομένως έχουμε σίγουρα αλγόριθμο απλής αντικατάστασης και το Z είναι το κλειδί.

- Στο δεύτερο κείμενο οι συχνότητες εμφάνισης των γραμμάτων εμφανίζονται κοντινές με εκείνες του πρωτότυπου κείμενο – πριν δηλαδή την κρυπτογράφηση αλλά εδώ εμφανίζεται πρώτο το C με συχνότητα 14.4860 δεύτερο το γράμμα αλλά να συγνώμη τα 9.6184 και ούτω καθεξής επειδή το κείμενο είναι αμιγώς εξειδικευμένο οι συχνότητες των γραμμάτων εμφανίζονται κάπως αλλαγμένες σε σχέση με αυτό της αγγλικής αλφαβήτου – λόγω της εξειδίκευσης του.

Παρόλα αυτά μπορούμε εύκολα να συμπεράνουμε ποσειδώνας αλγόριθμος απλής αντικατάστασης καθώς πιθανότατα το γράμμα C να έχει αντικαταστήσει το γράμμα E , το γράμμα R να έχει αντικαταστήσει το γράμμα A και ούτω καθεξής.

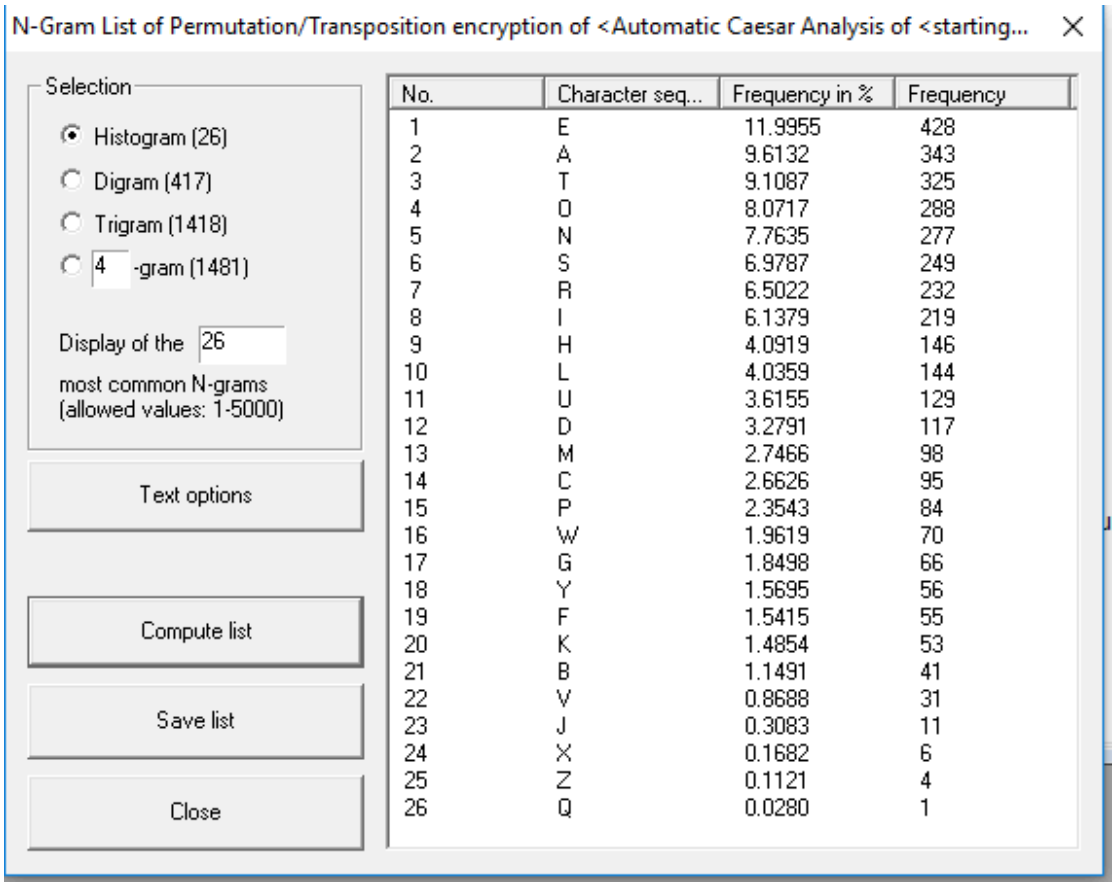


Εικόνα 8

- Για τον αλγόριθμο αντιμετάθεσης - πρώτο κείμενο (*Εικόνα 9*) χρησιμοποιήθηκε κλειδί k=3{3,2,1} και τα αποτελέσματα που παίρνουμε είναι πολύ κοντινά σε αυτά της αγγλικής αλφαβήτου, οι συχνότητες εμφάνισης των γραμμάτων πλησιάζουν εκείνα της εμφάνισης της αγγλικής αλφαβήτου.

- Το Ε εμφανίζει συχνότητα 11.9955 ενώ το γράμμα Α εμφανίζεται με συχνότητα 9.6132. Και ούτω καθεξής συμπεραίνουμε εύκολα πως είναι ένας αλγόριθμος κατηγορίας Μονοαλφαβητικής αντικατάστασης – αντιμεταθέσεις και για να εντοπίσουμε για το ποιος αλγόριθμος είναι ακριβώς απαιτούνται παραπάνω βήματα.

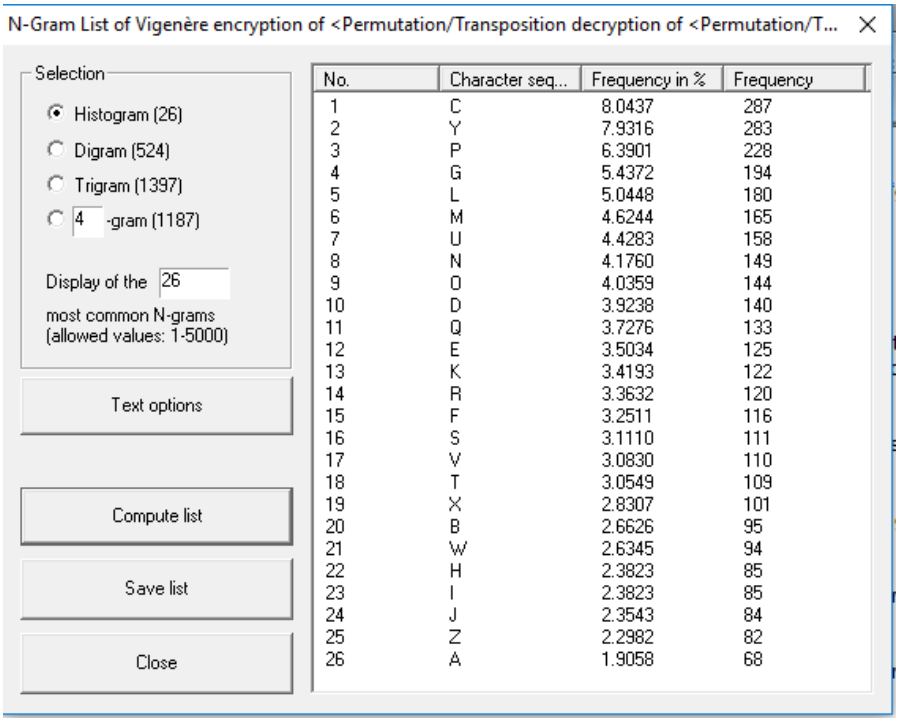
- Για το δεύτερο κείμενο μετά την κρυπτογράφηση βλέπουμε ότι η συχνότητές είναι κοντά σε αυτές της αγγλικής αλφαβήτου ναι πρώτο το Ε να καταγράφει συχνότητα 13.6808 , το γράμμα Τ να καταγράφει συχνότητα 9.3961 και ούτω καθεξής.



Εικόνα 9

- Για τον αλγόριθμο Vigenere - πρώτο κείμενο (*Εικόνα 10*) χρησιμοποιήθηκε κλειδί : LUCKY. Από την ανάλυση N-Gram συμπεραίνουμε πως οι συχνότητες εμφανίζονται "σπασμένες" πράγμα που σημαίνει πιθανόν ότι το κάθε γράμμα κωδικοποιείται με ένα άλλο τυχαίο γράμμα δοσμένου ενός κλειδιού. Αλλά συμπεραίνουμε πως πιθανότατα χρησιμοποιείται ο αλγόριθμος κρυπτογράφησης Vigenere.

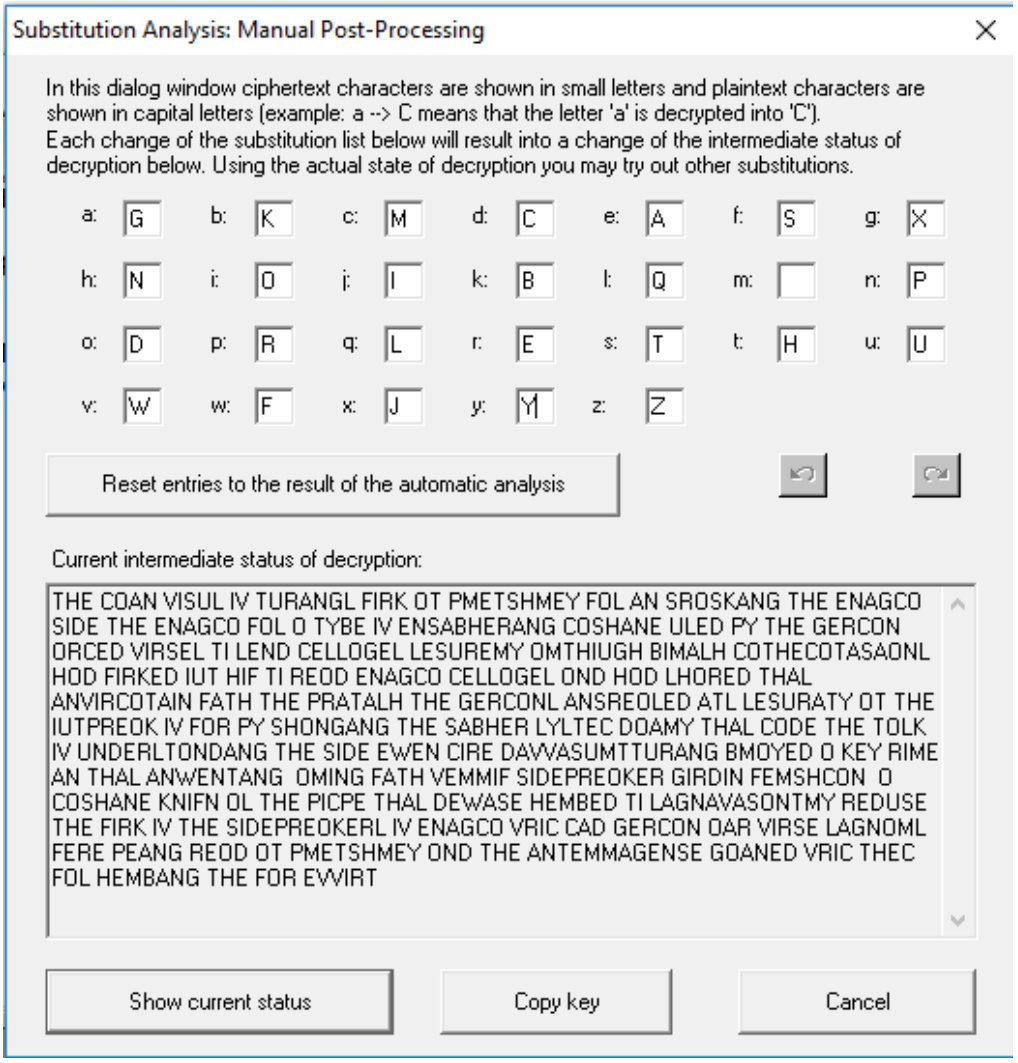
- Για το δεύτερο κείμενο οι συχνότητες επίσης εμφανίζονται «σπασμένες» πως σημαίνει ότι πιθανότατα έχουμε αλγόριθμο Vigenere.



Εικόνα 10

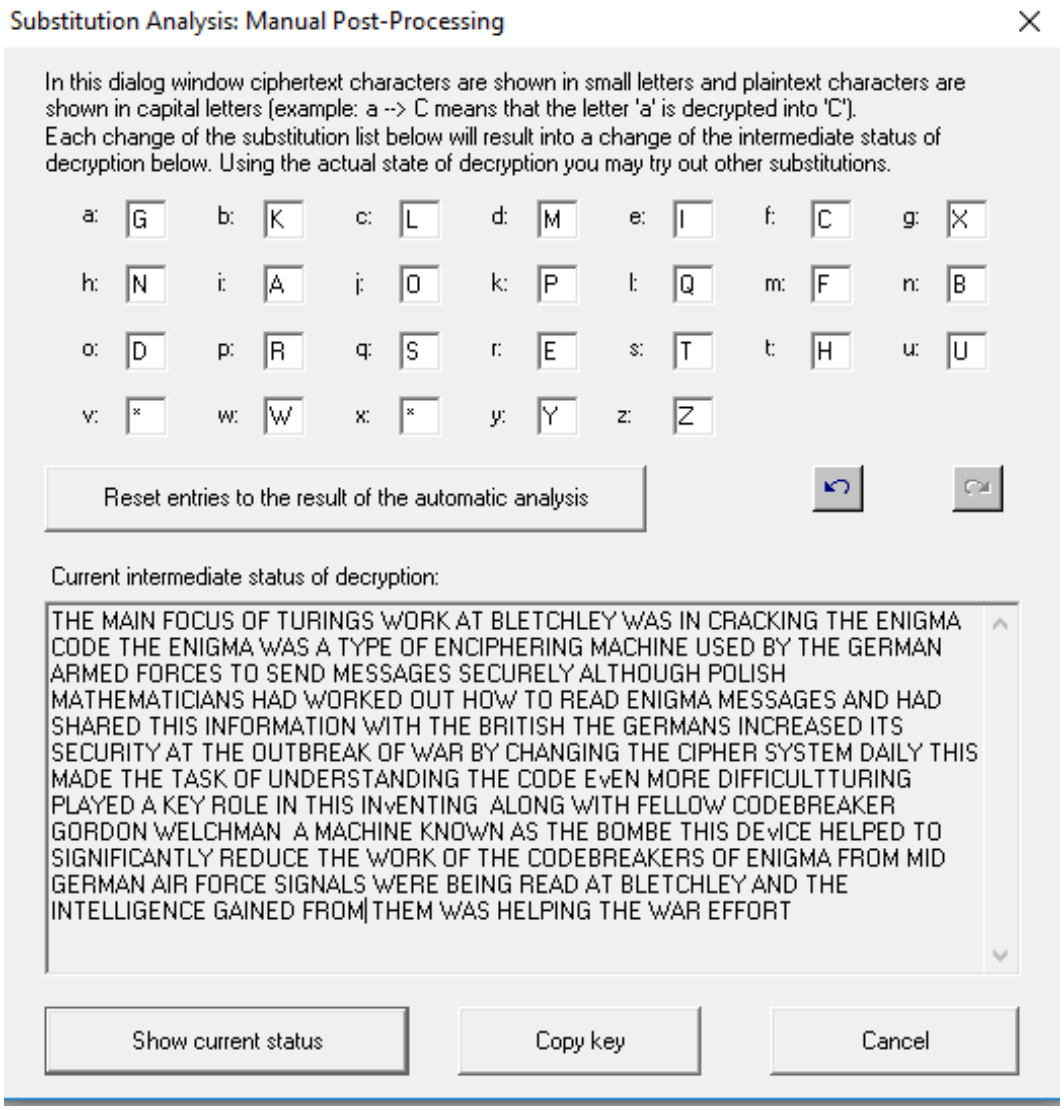
Δ) Για τον αλγόριθμο Affine παρατίθεται κώδικας σε γλώσσα C.

Ε) a) Στο κείμενο που μας δίνεται από το αρχείο : encr_substitution.txt Προσπαθούμε μέσω της εφαρμογής CrypTool να κάνουμε την κρυπτανάλυση βάσει των συχνοτήτων εμφανισης των γραμμάτων. Όπου με αυτό το τρόπο βρίσκουμε κάποιες από της λέξεις όπως: THE,SIDE,KEY,CODE αλλά δεν αποκρυπτογραφείται ολόκληρο το μήνυμα.



Εικόνα 11

b) Στην προσπάθεια εκ νέου να ανακτήσουμε το μήνυμα γνωρίζοντας ό,τι αφορά την κρύπτο μηχανή ENIGMA όπου βλέπουμε από το κείμενο ότι λέξη παραπλήσια ναι την λέξη ENIGMA – ENAGCO από την οποία συμπεραίνουμε πως το γράμμα I έχει αντικατασταθεί από το γράμμα A, το γράμμα M έχει αντικατασταθεί από το γράμμα C ,και το γράμμα A από το γράμμα O. Κάνουμε τις αντικαταστάσεις και προχωράμε σε επόμενα γράμματα.



Εικόνα 12

Το από-κρυπτογραφημένο μήνυμα παρουσιάζεται παρακάτω, κάνοντας αντικατάστασης σε λέξεις που μοιάζουν οικίες, καταλήγουμε σε αυτό το μήνυμα:

THE MAIN FOCUS OF TURINGS WORK AT BLETCHLEY WAS IN CRACKING THE ENIGMA CODE THE ENIGMA WAS A TYPE OF ENCIPHERING MACHINE USED BY THE GERMAN ARMED FORCES TO SEND MESSAGES SECURELY ALTHOUGH POLISH MATHEMATICIANS HAD WORKED OUT HOW TO READ ENIGMA MESSAGES AND HAD SHARED THIS INFORMATION WITH THE BRITISH THE GERMANS INCREASED ITS SECURITY AT THE OUTBREAK OF WAR BY CHANGING THE CIPHER SYSTEM DAILY THIS MADE THE TASK OF UNDERSTANDING THE CODE EvEN MORE DIFFICULT TURING

PLAYED A KEY ROLE IN THIS INvENTING ALONG WITH FELLOW CODEBREAKER GORDON WELCHMAN A MACHINE KNOWN AS THE BOMBE THIS DEvICE HELPED TO SIGNIFICANTLY REDUCE THE WORK OF THE CODEBREAKERS OF ENIGMA FROM MID GERMAN AIR FORCE SIGNALS WERE BEING READ AT BLETCHLEY AND THE INTELLIGENCE GAINED FROM THEM WAS HELPING THE WAR EFFORT

Z) Βάση του κρυΠτογραφημένο κειμένου που έχει κρυπτογραφηθεί με αλγόριθμο :Vigeneve κάνουμε κρυπτανάλυση εστιάζοντας στο μήκος του κλειδιού.

Το κρυπτογραφημένο κείμενο: hitciwtwvzzxciwdeopchfvlppopw
Παρατηρώντας τους χαρακτήρες και την επανεμφάνιση τους βλέπουμε ότι το τριγραμμα : *ciw* εμφανίζεται δύο φορές στο κείμενο. Η απόσταση ανάμεσα στην πρώτη εμφάνιση του τριγράμματος μέχρι εκεί που επαναλαμβάνεται για δεύτερη φορά είναι πιθανό να μας δώσει το κλειδί ή τους διαιρέτες του. Με αυτό τον τρόπο συμπεραίνουμε ό,τι πιθανό μέγεθος κλειδιού είναι το 9 και ο διαιρέτης του 3.
Επομένως «σπάμε» το μήνυμα σε στήλες παρατηρώντας τις πιθανές εκδοχές.
Ξεκινάμε τις δοκιμές με πιθανό κλειδί το 3.

H	I	T
C	I	W
T	W	V
Z	Z	X
C	I	W
D	E	O
P	C	H
F	V	L
P	P	O
P	W	

x	x	x

Εφαρμόζουμε Caesar σε κάθε στήλη και βασιζόμενη στις συχνότητες στην
Πρώτη στήλη έχουμε p à e (ολίσθηση 11)
Δεύτερη στήλη : i à e (ολίσθηση 4)

Τρίτη στήλη :

W	E	T
R	E	W
I	S	V
O	V	X
R	E	W
S	A	O
E	Y	H
U	R	L
E	L	O
E	S	

WE(T)**RE**(W)**IS**(V)**OV**(X)**RE**(W)**SA**(O)**EY**(H)**UR**(L)**EL**(O)**ES**

Το αποκρυπτογραφημένο μήνυμα:

WEAREDISCOVEREDSAVEYOURSELVES

Όπου κλειδί είναι το : LET

Και σημαίνει : **WE ARE DISCOVERED SAVE YOURSELVES**

(Εμεις αποκαλυφθηκαμε, σωστε τους εαυτους σας)

Πηγές που χρησιμοποιήθηκαν:
https://www.dcode.fr/caesar-cipher
http://www.multiwingspan.co.uk/cipher.php?page=caesar
Για την εικόνα:
http://giuliodagostino.com/2018/03/cryptography-cryptanalysis-tools-part-1/

Ambel Basha,
Αθήνα Μάιος 2018