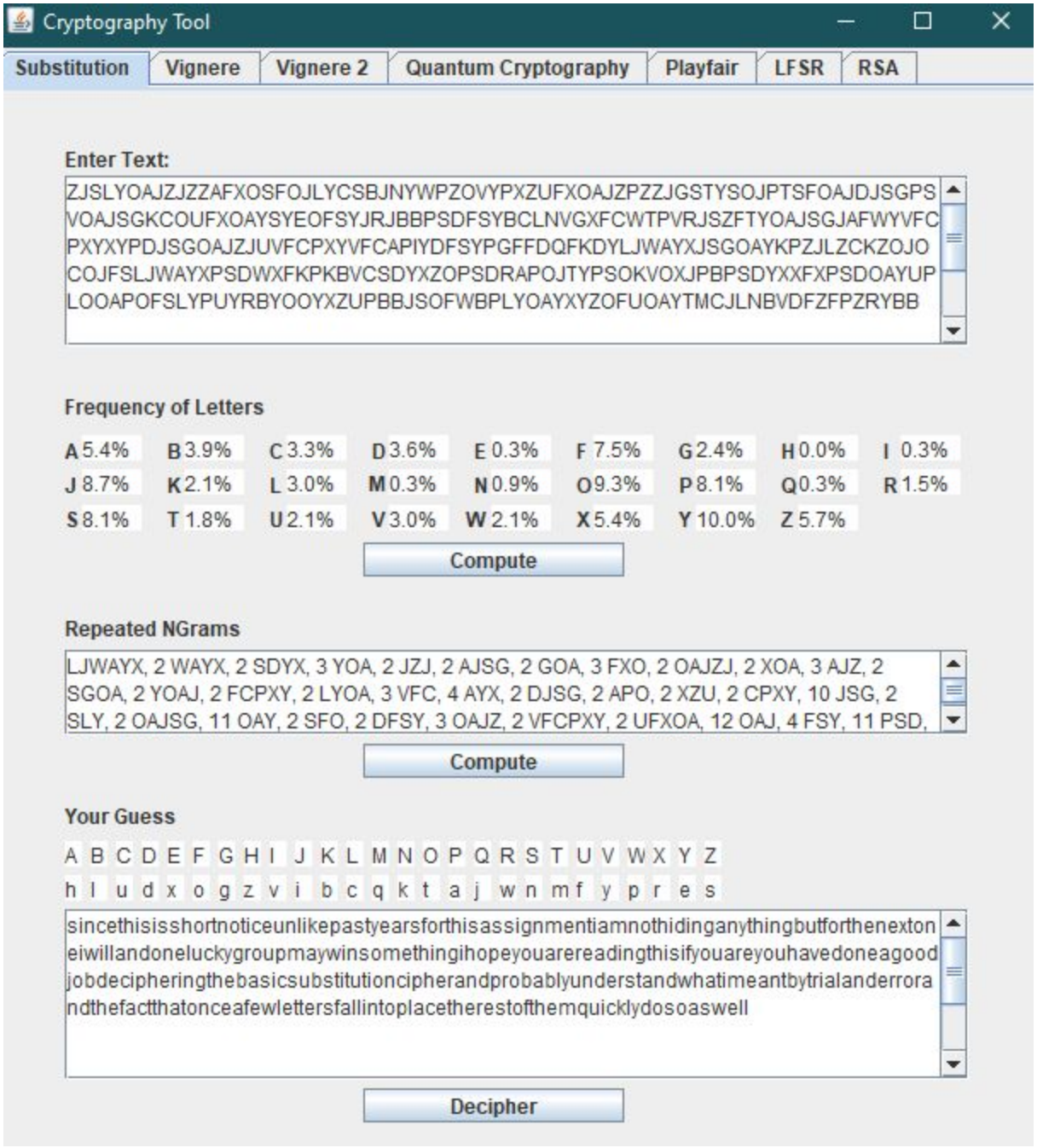
Maurizio Altamura
Cameron Lohr

CIS 3362 Homework #2:
Substitution Code Breaking, Vigenere
Due: Check WebCourses for the due date.
Directions: To be done in pairs. If you can't find someone to work with, you must submit individually. When you submit, please CLEARLY mark both group members on each document you submit.

1) Decode the following message, which was encrypted using the substitution cipher. Make sure to discuss all the steps you took, the key you arrived at, and the decoded message.

zjslyoajzjzzafxosfojlycsbjnywpzovypxzufxoajzpzzjgstysojptsfoajdjsgpsvoajsgkcoufxoaysyeofsyjrjbbpsdfsybclnvgxfcwtpvrjszftyoajsgjafwyvfcpxyxypdjsgoajzjuvfcpxyvfcapiydfsypgffdqfkdyljwayxjsgoaykpzjlzckzojocojfsljwayxpsdwxfkpkbvcsdyxzopsdrapojtypsokvoxjpbpsdyxxfxpsdoayuplooapofslypuyrbyooyxzupbbjsofwbplyoayxyzofuoaytmcjlnbvdfzfpzrybb

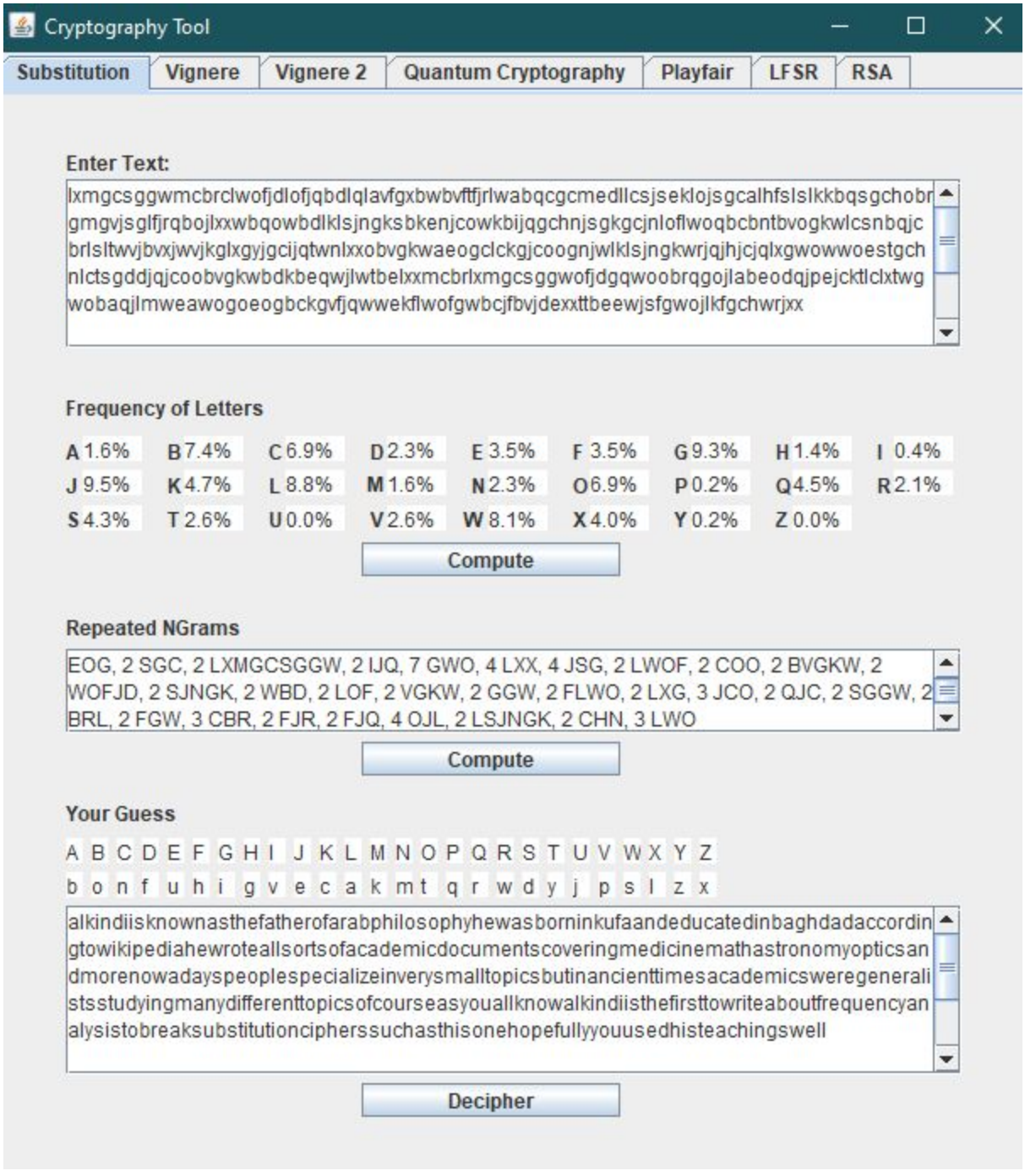
We started by plugging ‘e’ into the most common letter in the cipher, ‘y’. Then we looked at the most common 3 letter n-grams and which ones ended in ‘y’, which is ‘e’, because that three letter word would probably mean “the”. “Oay” appeared of frequently, and ended in “y”, and none of the other common n-grams ended in ‘y’. Then, looking at the 3 n-grams, we looked for “y-y”, because one of the most common 3 n-grams is “ere”, so we found “yxy”. By looking at the decryption so far, we could make out the word “letter”, so ‘b’ would be ‘l’. The letter after “letter” was ‘z’, which had a 5.7%, and “s” has a 6.3%, so it could probably be it. Then, by looking at the decryphered portion, we could see “--etherest--the--”, which we guessed to be “--e the rest of the --”. Then we saw “eth-s-sshort-ot”, and since both blanks before and after the s (5th character), is ‘j’ and it was probably a vowel, an ‘i’ would make sense. And we saw “-tri-l---error-”, which was probably, “trial and error”. And so on and so forth, till we solved the whole thing.


Cryptography Tool
Substitution
Vignere
Vignere 2
Quantum Cryptography
Playfair
LFSR
RSA
Enter Text:
ZJSLYOAJZJZZAFXOSFOJLYCSBJNYWPZOVYPXZUFXOAJZPZZJGSTYSOJPTSFOAJDJSGPS
VOAJSGKCOUFXOAYSYEOFSYJRJBBPSDFSYBCLNVGXFCWTPVRJSZFTYOAJSGJAFWYVFC
PXYXYPDJSGOAJZJUVFCPXYVFCAPIYDFSYPGFFDQFKDYLJWAYXJSGOAYKPZJLZCKZOJO
COJFSLJWAYXPSDWXFKPKBVCSDYXZOPSDRAPOJTYPSOKVOXJPBPSDYXXFXPSDOAYUP
LOOAPOFSLYPUYRBYOOYXZUPBBJSOFWBPLYOAYXYZOFUOAYTMCJLNBVDFZFPZRYBB
Frequency of Letters
A 5.4%
B 3.9%
C 3.3%
D 3.6%
E 0.3%
F 7.5%
G 2.4%
H 0.0%
I 0.3%
J 8.7%
K 2.1%
L 3.0%
M 0.3%
N 0.9%
O 9.3%
P 8.1%
Q 0.3%
R 1.5%
S 8.1%
T 1.8%
U 2.1%
V 3.0%
W 2.1%
X 5.4%
Y 10.0%
Z 5.7%
Compute
Repeated NGrams
LJWAYX, 2 WAYX, 2 SDYX, 3 YOA, 2 JZJ, 2 AJSG, 2 GOA, 3 FXO, 2 OAJZJ, 2 XOA, 3 AJZ, 2
SGOA, 2 YOAJ, 2 FCPXY, 2 LYOA, 3 VFC, 4 AYX, 2 DJSG, 2 APO, 2 XZU, 2 CPXY, 10 JSG, 2
SLY, 2 OAJSG, 11 OAY, 2 SFO, 2 DFSY, 3 OAJZ, 2 VFCPXY, 2 UFXOA, 12 OAJ, 4 FSY, 11 PSD,
Compute
Your Guess
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
h l u d x o g z v i b c q k t a j w n m f y p r e s
sincethisisshortnoticeunlikepastyearsforthisassignmentiamnothidinganythingbutforthenexton
eiwillandoneluckygroupmaywinsomethingihopeyouarereadingthisifyouareyouhavedoneagood
jobdecipheringthebasicsubstitutioncipherandprobablyunderstandwhatimeantbytrialanderrora
ndthefactthatonceafewlettersfallintoplacetherestofthemquicklydosoaswell
Decipher

2) Decode the following message, which was encrypted using the substitution cipher. Make sure to discuss all the steps you took, the key you arrived at, and the decoded message.

lxmgcsggwmcbrclwofjdlofjqbdlqlavfgxbwbvftfjrlwabqcgcmedllcsjseklojsgcalhfsls
lkkbqsgchobrgmgvjsglfjrqbojlxxwbqowbdlklsjngksbkenjcowkbijqgchnjsgkgcjnlofl
woqbcbntbvogkwlcsnbqjcbrlsltwvjbvxjwvjkglxgyjgcijqtwnlxxobvgkwaeogclckgjc
oognjwlklsjngkwrjqjhjcjqlxgwowwoestgchnlctsgddjqjcoobvgkwbdkbeqwjlwtbelxx
mcbrlxmgcsggwofjdgqwoobrqgojlabeodqjpejcktlclxtwgwobaqjlmweawogoeogbck
gvfjqwwekflwofgwbcjfbvjdexxttbeewjsfgwojlkfgchwrjxx

We started by assigning ‘j’ as ‘e’ because it had the highest percent. Then found a common tri-gram “jqj” which could be the common word “ere”. Then we started to look for ‘j’ and ‘q’ common tri-grams, and found “fjq”, “jqc”, and “ijq”.By comparing the percentages of ‘f’ and ‘i’, it was more likely that ‘f’ mapped to ‘h’ for the word “her”. Then, since we had ‘h’ and ‘e’, we looked for “the” and found “ofj”. Then we saw “ojl”, which was “th-” and ‘l’ has a high percent, so the thing that would have made sense was “tea” or “ten”. Because the percent was so high, we decided to put ‘l’ as ‘a’. Looking at what we decoded so far we can see a phrase “---a-the-ather--”, which probably has “the father” in it. And then the word after “father” seems to be “of”, and is confirmed by looking at the percent of ‘b’ and ‘o’. Then we saw “--ffere-tto--”, which looks like “different to--”. Then we saw “thefir-tto-ritea” which looks like “the first to write a”. Then, we kept seeing words fill in, and guessed the right keys.


Cryptography Tool
Substitution
Vignere
Vignere 2
Quantum Cryptography
Playfair
LFSR
RSA
Enter Text:
lxmgcsggwmcbrclwofjdlofjqbdlqlavfgxbwbvfftfjrlwabqcgcmedllcsjseklojsgcalhfslslkkbqsgchobr
gmgvjsglfjrqbojlxxwbqowbdlklsjngksbkenjcowkbijqgchnjsgkgcjnloflwoqbcbntbvogkwlcsnbqjc
brlsltwvjbvxjwvjkglxgyjgcijqtwnlxxobvgkwaeogclckgjcoognjwlklsjngkwrjqjhjcjqlxgwowwoestgch
nlctsgddjqjcoobvgkwbdkbeqwjlwtbelxxmcbrlxmgcsggwofjdgqwoobrqgojlabeodqjpejcktlclxtwg
wobaqjlmweawogoeogbckgvfjqwwekflwofgwbcjfbvjdexxttbeewjsfgwojlkfgchwrjxx
Frequency of Letters
A 1.6% B 7.4% C 6.9% D 2.3% E 3.5% F 3.5% G 9.3% H 1.4% I 0.4%
J 9.5% K 4.7% L 8.8% M 1.6% N 2.3% O 6.9% P 0.2% Q 4.5% R 2.1%
S 4.3% T 2.6% U 0.0% V 2.6% W 8.1% X 4.0% Y 0.2% Z 0.0%
Compute
Repeated NGrams
EOG, 2 SGC, 2 LXMGCSGGW, 2 IJQ, 7 GWO, 4 LXX, 4 JSG, 2 LWOF, 2 COO, 2 BVGKW, 2
WOFJD, 2 SJNGK, 2 WBD, 2 LOF, 2 VGKW, 2 GGW, 2 FLWO, 2 LXG, 3 JCO, 2 QJC, 2 SGGW, 2
BRL, 2 FGW, 3 CBR, 2 FJR, 2 FJQ, 4 OJL, 2 LSJNGK, 2 CHN, 3 LWO
Compute
Your Guess
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
b o n f u h i g v e c a k m t q r w d y j p s l z x
alkindiisknownasthefatherofarabphilosophyhewasborninkufaandeducatedinbaghdadaccordin
gtowikipediahewroteallsortsofacademicdocumentscoveringmedicinemathastronomyopticsan
dmorenowadayspeoplespecializeinverysmalltopicsbutinancienttimesacademicsweregenerali
stsstudyingmanydifferenttopicsofcourseasyouallknowalkindiisthefirsttowriteaboutfrequencyan
alysistobreaksubstitutionciphersuchasthisonehopefullyyouusedhisteachingswell
Decipher

3)Calculate the index of coincidence of both of the ciphertexts above. Please write a program to do so, include it, and show the letter frequencies the program calculated (as numbers, not percentages.) Express the index of coincidence as both a fraction and a decimal. Are you surprised by the results?

We wrote a program that finds the Index of Coincidence from two strings, Coincidence.c
Here is the full output we received for both strings:

Please enter the first string...
zjslyoajzjzzafxosfojlycsbjnywpzovypxzufxoajzpzzjgstysojptsfoajdjsgpsvoajsgkcou
fxoaysyeofsyjrjbbpsdfsybclnvgxfcwtpvrjszftyoajsgjafwyvfcpxyxypdjsgoajzjuvfcpx
yvfcapiydfsypgffdqfkdyljwayxjsgoaykpzjlzckzojocojfsljwayxpsdwxfkpkbvcsdyxz
opsdrapojtypsokvoxjpbpsdyxxfxpsdoayuplooapofslypuyrbyooyxzupbbjsofwbplyoa
yxyzofuoaytmcjlnbvdfzfpzrybb
Please enter the second string...
lxmgcsggwmcbrclwofjdlofjqbdlqlavfgxbwbvftfjrlwabqcgcmedllcsjseklojsgcalhfsls
lkkbqsgchobrgmgvjsglfjrqbojlxxwbqowbdlklsjngksbkenjcowkbijqgchnjsgkgcjnlofl
woqbcbntbvogkwlcsnbqjcbrlsltwvjbvxjwvjkglxgyjgcijqtwnlxxobvgkwaeogclckgjc
oognjwlklsjngkwrjqjhjcjqlxgwowwoestgchnlctsgddjqjcoobvgkwbdkbeqwjlwtbelxx
mcbrlxmgcsggwofjdgqwoobrqgojlabeodqjpejcktlclxtwgwobaqjlmweawogoeogbck
gvfjqwwekflwofgwbcjfbvjdexxttbeewjsfgwojlkfgchwrjxx

If a character doesn't appear, it is not present in the string!

String 0! The IoC of this string is 0.060532 decimal, 6572 / 108570 fraction.

The number of a's in String 0 is 18!
The number of b's in String 0 is 13!
The number of c's in String 0 is 11!
The number of d's in String 0 is 12!
The number of e's in String 0 is 1!
The number of f's in String 0 is 25!

The number of g's in String 0 is 8!
The number of i's in String 0 is 1!
The number of j's in String 0 is 29!
The number of k's in String 0 is 7!
The number of l's in String 0 is 10!
The number of m's in String 0 is 1!
The number of n's in String 0 is 3!
The number of o's in String 0 is 31!
The number of p's in String 0 is 27!
The number of q's in String 0 is 1!
The number of r's in String 0 is 5!
The number of s's in String 0 is 27!
The number of t's in String 0 is 6!
The number of u's in String 0 is 7!
The number of v's in String 0 is 10!
The number of w's in String 0 is 7!
The number of x's in String 0 is 18!
The number of y's in String 0 is 33!
The number of z's in String 0 is 19!

String 1! The IoC of this string is 0.059424 decimal, 10358 / 174306 fraction.

The number of a's in String 1 is 7!
The number of b's in String 1 is 31!
The number of c's in String 1 is 29!
The number of d's in String 1 is 10!
The number of e's in String 1 is 15!
The number of f's in String 1 is 15!
The number of g's in String 1 is 39!
The number of h's in String 1 is 6!
The number of i's in String 1 is 2!
The number of j's in String 1 is 40!
The number of k's in String 1 is 20!

The number of l's in String 1 is 37!
The number of m's in String 1 is 7!
The number of n's in String 1 is 10!
The number of o's in String 1 is 29!
The number of p's in String 1 is 1!
The number of q's in String 1 is 19!
The number of r's in String 1 is 9!
The number of s's in String 1 is 18!
The number of t's in String 1 is 11!
The number of v's in String 1 is 11!
The number of w's in String 1 is 34!
The number of x's in String 1 is 17!
The number of y's in String 1 is 1!

As shown by our output, the Index of Coincidence for the first string is 0.060532 decimal, 6572 / 108570 fraction. This isn't very surprising at all, considering the English language has an index of coincidence of .0667.
The same can be said about the second string. 0.059424 decimal, 10358 / 174306 fraction is not that far off of the English Index of Coincidence considering that every character is not present in the string as well.

4) Decode the following ciphertext that was encoded using the Vigenere cipher with a keyword from this wordlist:
https://www.ef.com/wwen/english-resources/english-vocabulary/top-1000-words/
To help you automate your task, I will guarantee that the substring "fast" will appear somewhere in the plaintext.

Here is the ciphertext:

Uaztkwfmazkohrispmazekvlrwtiexfofekmrcbnoaaaaxaqnmqrrajmzkvgmlvszsrtbm
qngigpwilparqsgtbplmteztuxtipbbhrrzijxngptuetegxbfmtvcgpxcfvmnghkgutuxdtrxz
javnimrgiiyxeeparqtnixsazqrghowomcaqrgedxuwgaqsyocidcbndpeomvmqjbxleutf
gxfhyeghorspvdqtuiyaupylfiylhihierrafthioehlqiuabismixzybuysrijparqszsfvl

We wrote a program to decode the Ciphertext, Vignere.c. The program reads in all the provided 1000 words, and brute forces each possible output for an entered string into a file called output.txt. Once you have the output.txt file, you are able to search the file for “fast” to find the final results:

Key: Management
Plaintext:
IAMTESTINGYOURCODINGSKILLSHERETOSEEIFYOUCANAUTOMATEREADINGINALISTO
FPOTENTIALKEYWORDSAPPLYTHEMTOTHECIPHERTEXTANDTHENAUTOMATICALLYSC
ANTHECIPHERTEXTFORAPARTICULARWORDTHELONGERTHISCIPHERTEXTISTHESLO
WERYOURPROGRAMWILLRUNBUTSOMETHINGLIKETHISWILLSTILLBEVERYFASTBECAU
SEIHAVEGIVENYOUSOFEWWORDSTOTRY

We have attached the output file as well as the program so you may verify the validity.