Київський національний університет імені Тараса Шевченка
Факультет комп’ютерних наук та кібернетики
Кафедра інтелектуальних програмних систем
Математичні основи захисту інформації

Лабораторна робота №3
“Застосування методів частотного аналізу мови для дешифрування”

Виконали студенти 3-го курсу
Групи ІПС-32

Роботу виконали:
Шифрувальники:
Ольховатий Ігор
Ковальов Володимир
Тряско Софія
Дешифрувальники:
Цілинко Олександр
Бондар Юлія
Волик Артем

Київ 2023

Завдання:

Ми маємо вхідний текст, який мусимо зашифрувати за допомогою власного шифрувального алгоритму. А зашифрований текст ми маємо дешифрувати за допомогою методів частотного аналізу

Розв'язання

Шифрування:

Маємо вхідний текст:

The material given in this work is considered there due to the fact that last time the idea that science has exhausted itself and its role is decreasing every day, appears in the media and among specialists. This situation is especially surprising when a person who graduated from some University has access to the Internet, believes that everything he needs is in the Internet and there is no need for scientific research. Various pandemics demonstrates us the absurdity of such a thought. It is obvious that the progress in virology is absolutely important and necessary. But achievement in other areas of science, in particular in computer science, is very important also (may be less obvious). The relationship between algebra, computer science and programming is considers in this work; moreover we made an attempt to describe a certain influence of algebra on computer science and on the contrary computer science on algebra and other fields.

We do not pretend on completeness of the material, however it is almost impossible to give a full description of algebra applications in computer science and programming.

Для шифрування використаємо модифікований шифр Цезаря зі зміщенням у 3 (внесено достатньо стохастичних змін, щоб шифрувальники і не здогадалися, що за основу був взятий шифр Цезаря).

Original	Encoded
A	X
B	E
C	I
D	K
E	H
F	F
G	J
H	G
I	O
J	M
K	V
L	L
M	R
N	Q
O	P
P	S
Q	T
R	U
S	N
T	W
U	D
V	Y
W	C
X	Z
Y	B
Z	A

Варто зауважити, що кожен символ було закодовано у верхньому регістрі,

а спеціальні символи та пробіли було прибрано. Для автоматизації шифрування було написано тривіальну програму мовою Python:

```
encoding = {
    "A": "X",
    "B": "E",
```

```
    "C": "I",
    "D": "K",
    "E": "H",
    "F": "F",
    "G": "J",
    "H": "G",
    "I": "O",
    "J": "M",
    "K": "V",
    "L": "L",
    "M": "R",
    "N": "Q",
    "O": "P",
    "P": "S",
    "Q": "T",
    "R": "U",
    "S": "N",
    "T": "W",
    "U": "D",
    "V": "Y",
    "W": "C",
    "X": "Z",
    "Y": "B",
    "Z": "A",
    "        "
    : "", " ":
    "", ",":
    "", ":":
    "", ";":
    "", "!":
    "", "?":
    "", "(":
    "", ")":
    "", "\n":
    ""
}


def encode(message):
    encoded_message = ""
    for letter in message:
        encoded_message += encoding[letter.upper()]
```

```
    return encoded_message



if __name__ == '__main__':
    message = """The material given in this work is considered there due to the
fact that last time the idea that science has exhausted itself and its role is
decreasing every day, appears in the media and among specialists. This
situation is especially surprising when a person who graduated from some
University has access to the Internet, believes that everything he needs is
in the Internet and there is no need for scientific research. Various
pandemics demonstrates us the absurdity of such a thought.
It is obvious that the progress in virology is absolutely important and
necessary. But achievement in other areas of science, in particular
in computer science, is very important also (may be less obvious).
The relationship between algebra, computer science and
programming is considers in this work; moreover we made an
attempt to describe a certain influence of algebra on computer
science and on the contrary computer science on algebra and other
fields.
We do not pretend on completeness of the material, however it is
almost impossible to give a full description of algebra applications
in computer science and programming."""
    print(encode(message))
```

Дешифрування:

Маємо наступний шифро-текст, який мусимо дешифрувати:

WGHRXWHUOXLJOYHQOQWGONCPUVONIPQNOKHUHKWGHUHKD
HWPWGHFXIWW
GXWLXNWWORHWGHOKHXWGXWNIOHQIHGXNHZGXDNWHKOW
NHLFXQKOWNUPL
HONKHIUHXNOQJHYHUBKXBXSSHXUNOQWGHRHKOXXQKXRPQJ
NSHIOXLONWNW
GONNOWDXWOPQONHNSHIOXLLBNDUSUONOQJCGHQXSHUNPQC
GPJUXKDXWHK
FUPRNPRHDQOYHUNOWBGXNXIIHNNWPWGHOQWHUQHWEHLOH
YHNWGXWHYHU
BWGOQJGHQHHKNONOQWGHOQWHUQHWXQKWGHUHONQPQHH
KFPUNIOHQWOF
OIUHNHXUIGYXUOPDNSXQKHROINKHRPQNWUXWHNDNWGHXEN
DUKOWBPFNDIG

XWGPDJGWOWONPEYOPDNWGXWWGHSUPJUHNNOQYOUPLPJBON
XENPLDWHLB
ORSPUWXQWXQKQHIHNNXUBEDWXIGOHYHRHQWOQPWGHUXUH
XNPFNIOHQIHO
QSXUWOIDLXUOQIPRSDWHUNIOHQIHONYHUBORSPUWXQWXLNP
RXBEHLHNNPEY
OPDNWGHUHLXWOPQNGOSEHWCHHQXLJHEUXIPRSDWHUNIOHQI
HXQKSUPJUXR
ROQJONIPQNOKHUNOQWGONCPUVRPUHPYHUCHRXKHXQXWWHR
SWWPKHNIUOE
HXIHUWXOQOQFLDHQIHPFXLJHEUXPQIPRSDWHUNIOHQIHXQKPQ
WGHIPQWUXUBI
PRSDWHUNIOHQIHPQXLJHEUXXQKPWGHUFOHLKNCHKPQPWSUH
WHQKPQIPRSLH
WHQHNNPFWGHRXWHUOXLGPCHYHUOWONXLRPNWORSPNNOEL
HWPJOYHXFDLL
KHNIUOSWOPQPFXLJHEUXXSSLOIXWOPQNOQIPRSDWHUNIOHQIH
XQKSUPJUXRR OQJ

Напишемо програму на C++, яка рахує кількість числову характеристику кожного символу(як часто він зустрічається в рядку).

Отримаємо наступні данні:

B 1.19177

C 0.758397

D 2.38353

E 1.40845

F 1.30011

G 3.79198

H 13.4345

I 4.44204

J 1.84182

K 3.14193

L 2.92524

N 8.01733

O 8.88407

P 6.50054

Q 7.1506

R 2.92524
S 2.70856
U 6.60888
V 0.108342
W 8.77573
X 7.80065
Y 1.40845
Z: 0

Правдоподібним видається, що H співвідноситься з E.
Далі букви W,O з буквами - {T,I}
Букви N,X з буквами -{A,O,S}
Подивимось комбінації з двох букв
HU: 0.0249187
WG: 0.0249187
OQ: 0.0184182
GH: 0.0173348
WH: 0.0173348
PQ: 0.0162514
HQ: 0.0162514
ON: 0.0151679
XW: 0.0140845
NO: 0.0130011
Помітимо, що очевидно якась з двох комбінацій HU або WG співвідноситься з TH, бо TH найчастіша пара англійських літер які йдуть поруч. Але HU не може бути TH, тому що H-E. Отже WG - TH. Отже W=T, G=H.Плюс до цього враховуючи, що H-E. Виходить слово THE. Далі, враховуючи, що ER досить популярна, то можемо співвіднести її з HU. Отже U-R, що схоже на правду, бо R має частоту 6.7, як і U. З цього ще випливає, що буква O відповідає букві I, що теж схоже бо вони мають схожі частотні характеристики. Ще можемо помітити таку пару букв як XW, якщо X-O, то отримаємо буквосполучення OT,або ST, яке не є популярним, але якщо X-A, то отримаємо буквосполучення AT, яке є популярним. Отже, N або S або O. Також можемо припустити після цього що OQ відповідає популярній діграфу IN, бо воно не може відповідати IT, бо W-T. Отже Q-N, що схоже на правду відповідно до частоти Q-7.15, а у

N-7.0.

Подивимось на буквосполучення ON. Якщо N-O, то виходить діграф IO, який не є популярним, але якщо N-S, то виходить IS. До того ж, розглянемо слово в першому рядку WGON, якщо це одне слово то це дуже схоже на слово THIS. Отже можна припустити, що N-S. Далі дивимось на пару PQ, Q-N. Отже серед популярних диграфів є ON. Можемо зробити припущення,що P-O.

Отже зараз можемо точно сказати, що

W-T

G-H

U-R

H-E

O - I

X- A

N- S

Q-N

P- O

На даний момент текст має вигляд

THErATERIAljIyENINTHIScORvISiONSIkEREkTHEREkdETOTHEfAiTTH
ATlASTTIrETHEIk
EATHATSiIENiEHASEzHAdSTEkITSElfANkITSROlEISkEiREASINjEyERb
kAbAssEARSINT
HErEkIAANkArONjSsEiIAlISTSTHISSITdATIONISESsEiIAllbSdRsRISINjc
HENAsERSONcH
OjRAkdATEkfROrSOrEdNIyERSITbHASAiiESSTOTHEINTERNETeElIEyE
STHATEyERbTHI
NjHENEEkSISINTHEINTERNETANkTHEREISNONEEkfORSiIENTIfIiRES
EARiHyARIOdSs
ANkErIiSkErONSTRATESdSTHEAeSdRkITbOfSdiHATHOdjHTITISOeyIOd
STHATTHEsROj
RESSINyIROlOjbISAeSOldTElbIrsORTANTANkNEiESSARbedTAiHIEyErE
NTINOTHERARE
ASOfSiIENiEINsARTIidlARINiOrsdTERSiIENiEISyERbIrsORTANTAlSOrA
beElESSOeyIOdS
THERElATIONSHIseETcEENAljEeRAiOrsdTERSiIENiEANksROjRArrINjIS
iONSIkERSINTHI

ScORvrOREOyERcErAkEANATTErsTTOkESiRIeEAiERTAININfldENiEOf
AljEeRAONiOrsdT
ERSiIENiEANkONTHEiONTRARbiOrsdTERSiIENiEONAljEeRAANkOTHE
RfIElkScEkONOT
sRETENkONiOrslETENESSOfTHErATERIAlHOcEyERITISAlrOSTIrsOSSIe
lETOjIyEAfdllkE
SiRIsTIONOfAljEeRAAsslIiATIONSINiOrsdTERSiIENiEANksROjRArrINj
Де великі букви це ті що розшифрованні, а малі оригінальні.

Розглянемо частину третього рядка HASAiiESSTOTHEINTERNET. Тут однозначно можна виділити наступні слова HAS AiiESS TO THE INTERNE. З контексту можна припустити що слово AiiESS це слово ACCESS(це відповідає і теоретичній частоті букви C і випливає з контексту).

Тобто виходить, що I-C.

THErATERIAljIyENINTHIScORvISCONSIkEREkTHEREkdETOTHEfACTT
HATlASTTIrETHE
IkEATHATSCIENCEHASEzHAdSTEkITSElfANkITSROlEISkECREASINjE
yERbkAbAssEAR
SINTHErEkIAANkArONjSsECIAlISTSTHISSITdATIONISESsECIAllbSdRs
RISINjcHENAsER
SONcHOjRAkdATEkfROrSOrEdNIyERSITbHASACCESSTOTHEINTERNE
TeElIEyESTHAT
EyERbTHINjHENEEkSISINTHEINTERNETANkTHEREISNONEEkfORSCI
ENTIfICRESEAR
CHyARIOdSsANkErICSkErONSTRATESdSTHEAeSdRkITbOfSdCHATHOd
jHTITISOeyIOdS
THATTHEsROjRESSINyIROlOjbISAeSOldTElbIrsORTANTANkNECESSA
RbedTACHIEyErE
NTINOTHERAREASOfSCIENCEINsARTICdlARINCOrsdTERSCIENCEISy
ERbIrsORTANTAl
SOrAbeElESSOeyIOdSTHERElATIONSHIseETcEENAljEeRACOrsdTERSCI
ENCEANksROj
RArrINjISCONSIkERSINTHIScORvrOREOyERcErAkEANATTErsTTOkES
CRIeEACERTAINI
NfldENCEOfAljEeRAONCOrsdTERSCIENCEANkONTHECONTRARbCOrs
dTERSCIENCE
ONAljEeRAANkOTHERfIElkScEkONOTsRETENkONCOrslETENESSOfTH

ErATERIAlHOcE
yERITISAlrOSTIrsOSSIelETOjIyEAfdllkESCRIsTIONOfAljEeRAAsslICATI
ONSINCOrsdTER
SCIENCEANksROjRArrINj

Розглянемо перший рядок і розіб’ємо його на явні слова

THE rATERIAljIyEN IN THIS cORvISCONSIkEREk THERE kdE TO THE fACT THAT lASTTIrE THE Ik. Можна помітити слово fACT. Ймовірно тут закодовано слово

FACT.(зрозуміло з контексту). F-F

THE rATERIAljIyEN IN THIS cORv IS CONSIkEREk THERE kdE TO THE FACT THAT Розглянемо букву K,D. Подивимось якій букві вона може потенційно відповідати: M,C,D,F,H,L,U. Зразу можна зробити висновок, що букви C,H не підходять, бо вони вже відповідають іншим. Залишились M,D,F,L,U. Розглянемо усі можливі пари. Перебравши усі можливі варіанти отримали що лише пара K-D, D-U. Утворюють осмисленне слово DUE. Отже на даний момент маємо, що:

W-T
G-H
U-R
H-E
O - I
X- A
N- S
Q-N
P- O
K-D
D-U
I-C
F-F

Отже перший рядок має вигляд THE rATERIAljIyEN IN THIS cORv IS CONSIDERED

THERE DUE TO THE FACT THAT IDEA THAT SCIENCE HAS EzHAUSTED ITSElF AND

ITS ROlE IS DECREASINjEyERbDAbAssEARSIN

Подивившись на слово EzHAUSTED можна зрозуміти з контексту, що це слово

EXHAUSTED. Отже
Z-X.
THE rATERIAljIyEN IN THIS cORv IS CONSIDERED THERE DUE TO THE FACT THAT lASTTIrE THE IDEA THAT SCIENCE HAS EXHAUSTED ITSElF AND ITS ROlE IS DECREASINjEyERbDAbAssEARSINTHErEDIAANDArONjSsECIAlISTST HISSITUATIONIS ESsECIAllbSURsRISINjcHENAsERSONcHOjRADUATEDFROrSOrEUNIyERSITbHASACC ESSTOTHEINTERNETeElIEyESTHATEyERbTHINjHENEEDSISINTHEINTERNETANDTHE REISNONEEDFORSCIENTIFICRESEARCHyARIOUSsANDErICSDErONSTRATESUSTHE AeSURDITbOFSUCHATHOUjHTITISOeyIOUSTHATTHEsROjRESSINyIROlOjbISAeSOlUT ElbIrsORTANTANDNECESSARbeUTACHIEyErENTINOTHERAREASOFSCIENCEINsARTI CUlARINCOrsUTERSCIENCEISyERbIrsORTANTAlSOrAbeElESSOeyIOUSTHERElATIONS HIseETcEENAljEeRACOrsUTERSCIENCEANDsROjRArrINjISCONSIDERSINTHIScORvrO REOyERcErADEANATTErsTTODESCRIeEACERTAININFlUENCEOFAljEeRAONCOrsUTE RSCIENCEANDONTHECONTRARbCOrsUTERSCIENCEONAljEeRAANDOTHERFIElDScE DONOTsRETENDONCOrslETENESSOFTHErATERIAlHOcEyERITISAlrOSTIrsOSSIelETOjI yEAFUllDESCRIsTIONOFAljEeRAAsslICATIONSINCOrsUTERSCIENCEANDsROjRArrINj Подивимось на слово ROlE. Тут можна побачити, що скоріш за все тут було слово ROLE(зрозуміло з контексту).
Отже L-L
Тепер можна побачити, що в першому рядку слово rATERIAlL - це слово MATERIAL.
Отже R-M.
THE MATERIAL jIyEN IN THIS cORv IS CONSIDERED THERE DUE TO THE FACT THAT

LAST TIME THE IDEA THAT SCIENCE HAS EXHAUSTED ITSELF AND ITS ROLE IS

DECREASINj

З граматичної форми і контексту зрозуміло, що слово DECREASINj - DECREASING.

Отже J-G.

Якщо розглянути слово GIyEN і порівняти можливі варіанти відповідностей для y. То отримаємо, що єдиний варіант який підходить: Y-V.

THE MATERIAL GIVEN IN THIS cORv IS CONSIDERED THERE DUE TO THE FACT THAT

LAST TIME THE IDEA THAT SCIENCE HAS EXHAUSTED ITSELF AND ITS ROLE IS

DECREASING EVERb DAb AssEARS IN THE MEDIA AND AMONG SsECIALISTS THIS

SITUATION IS

EVERb DAb в цій парі слів очевидно, що B-Y.

SsECIALISTS - це слово SPECIALISTS. Отже S-P.

Отже маємо наступну таблицю відповідностей

W-T
G-H
U-R
H-E
O - I
X- A
N- S
Q-N
P- O
K-D
D-U
I-C
F-F
Z-X.
L-L
R-M
J-G

Y-V
B-Y
S-P
THE MATERIAL GIVEN IN THIS cORv IS CONSIDERED THERE DUE TO THE FACT THAT
LAST TIME THE IDEA THAT SCIENCE HAS EXHAUSTED ITSELF AND ITS ROLE IS
DECREASING EVERY DAY APPEARS IN THE MEDIA AND AMONG SPECIALISTS THIS
SITUATION IS ESPECIALLY SURPRISING cHENA PERSON cHO GRADUATED FROM
SOME UNIVERSITY HAS ACCESS TO THE INTERNET eELIEVES THAT EVERYTHING
HE NEEDS IS IN THE INTERNET AND THERE IS NO NEED FOR SCIENTIFIC
RESEARCH VARIOUS PANDEMICS DEMONSTRATES US THE AeSURDITY OF SUCH A
THOUGHT IT IS OeVIOUS THAT THE PROGRESS IN VIROLOGY IS AeSOLUTELY
IMPORTANT AND NECESSARY eUT ACHIEVEMENTINO THER ARE
ASOFSCIENCEINPARTICULARINCOMPUTERSCIENCEISVERYIMPORTANTALSOMAYeE
LESSOeVIOUSTHERELATIONSHIPeETcEENALGEeRACOMPUTERSCIENCEANDPROG
RAMMINGISCONSIDERSINTHIScORvMOREOVERcEMADEANATTEMPTTODESCRIeEA
CERTAININFLUENCEOFALGEeRAONCOMPUTERSCIENCEANDONTHECONTRARYCO
MPUTERSCIENCEONALGEeRAANDOTHERFIELDScEDONOTPRETENDONCOMPLETE
NESSOF THE MATERIAL HOcEVER IT IS ALMOST IMPOSSIeLE TO GIVE A FULL
DESCRIPTION OF ALGEeRA APPLICATIONS IN COMPUTER SCIENCE AND
PROGRAMMING
Подивимось на слова ALGEeRA OeVIOUS eELIEVES. З них стає очевидно, що на місці букви е має бути буква В

E-B

Далі, якщо подивитись на слова cORv HOcEVER можемо зробити висновок, що c-W, a v-K. Тоді маємо

W-T
G-H
U-R
H-E
O - I
X- A
N- S
Q-N
P- O
K-D
D-U
I-C
F-F
Z-X.
L-L
R-M
J-G
Y-V
B-Y
S-P
C-W
V-K

Дешифрованний текст:

THE MATERIAL GIVEN IN THIS WORK IS CONSIDERED THERE DUE TO THE FACT

THAT LAST TIME THE IDEA THAT SCIENCE HAS EXHAUSTED ITSELF AND ITS ROLE

IS DECREASING EVERY DAY APPEARS IN THE MEDIA AND AMONG SPECIALISTS

THIS SITUATION IS ESPECIALLY SURPRISING WHEN A PERSON WHO GRADUATED

FROM SOME UNIVERSITY HAS ACCESS TO THE INTERNET BELIEVES THAT

EVERYTHING HE NEEDS IS IN THE INTERNET AND THERE IS NO NEED FOR

SCIENTIFIC RESEARCH VARIOUS PANDEMICS DEMONSTRATES US THE ABSURDITY

OF SUCH A THOUGHT IT IS OBVIOUS THAT THE PROGRESS IN VIROLOGY IS

ABSOLUTELY IMPORTANT AND NECESSARY BUT ACHIEVEMENT IN OTHER AREAS

OF SCIENCE IN PARTICULAR IN COMPUTER SCIENCE IS VERY IMPORTANT ALSO

MAYBE LESS OBVIOUS THE RELATIONSHIP BETWEEN ALGEBRA COMPUTER

SCIENCE AND PROGRAMMING IS CONSIDERS IN THIS WORK MOREOVER WE MADE AN ATTEMPT TO DESCRIBE A CERTAIN INFLUENCE OF ALGEBRA ON COMPUTER

SCIENCE AND ON THE CONTRARY COMPUTER SCIENCE ON ALGEBRA AND OTHER

FIELDS WE DO NOT PRETEND ON COMPLETENESS OF THE MATERIAL HOWEVER IT

IS ALMOST IMPOSSIBLE TO GIVE A FULL DESCRIPTION OF ALGEBRA APPLICATIONS

IN COMPUTER SCIENCE AND PROGRAMMING

Література:

- Лекції з предмету “Математичні основи захисту інформації”
- https://www.101computing.net/frequency-analysis/
- https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BB%D1%84%D0%B0%D0%B2%D0%B8%D1%82