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

Лабораторна робота №4
“Біграмне шифрування”
Виконали студенти 3-го курсу
Групи ІПС-32

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

Київ 2023

Завдання

Ми маємо вхідний текст, який мусимо зашифрувати за допомогою біграмного шифрування. А зашифрований текст ми маємо дешифрувати за допомогою методів частотного аналізу

Розв'язання

Процес 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 біграмне шифрування за допомогою матриці:

$$\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}$$

Розіб'ємо текст на блоки по 2 символи (оскільки розмірність матриці 2x2) та застосуємо шифрування, це буде виглядати наступним чином:

[A, B] * | 1 2 | = [A*1 + B*3, A*2 + B*4] = [A + 3B, 2A + 4B]
         | 3 4 |

Варто зазначити, що насправді операція множення (*) відбувається над елементами матриці та числовими представленнями символів тексту, де кожній літері латиського алфавіту за порядковим номером відповідає число від 0 до 26. Всі обчислення проводяться у полі за модулем 26.

Для автоматизації процесу шифрування використаємо наступний тривіальний програмний код мовою Python (також тут вказана функція для декодування за відомою матрицею):

```
import numpy as np


def number_to_letter(n):
    if 0 <= n <= 26:
        return chr(ord('A') + n)
    else:
        return None


def letter_to_number(letter):
    # Ensure the input is a single uppercase letter
    if len(letter) == 1 and 'A' <= letter <= 'Z':
        return ord(letter) - ord('A')
    else:
        return None


def bipartite_encrypt(text: str, matrix: np.ndarray, alphabet_size: int = 26):
    # Remove whitespaces, punctuation, \n and convert to uppercase
    text = "".join([letter.upper() for letter in text if letter.isalpha()])

    if len(text) % 2 != 0:
```

```
        text = text[:-1]

    encrypted_text = ""

    for i in range(0, len(text), 2):
        block = text[i: i + 2]
        vector = [letter_to_number(block[0]), letter_to_number(block[1])]

        vector_array, matrix_array = np.array(vector), np.array(matrix)
        encrypted_vector = np.dot(matrix_array, vector_array)
        encrypted_block = "".join([number_to_letter(int(val) % alphabet_size) for
val in encrypted_vector])
        encrypted_text += encrypted_block

    return encrypted_text


def bipartite_decrypt(text: str, matrix: np.ndarray, alphabet_size: int = 26):
    decrypted_text = ""
    inverse_matrix = np.linalg.inv(matrix)

    for i in range(0, len(text), 2):
        block = text[i:i+2]
        vector = [letter_to_number(block[0]), letter_to_number(block[1])]

        vector_array = np.array(vector)
        decrypted_vector = np.round(np.dot(inverse_matrix, vector_array))
        decrypted_block = ''.join([str(number_to_letter(int(val % alphabet_size)))
for val in decrypted_vector])

        decrypted_text += decrypted_block

    return decrypted_text


if __name__ == '__main__':
    matrix = np.array([[1, 2], [3, 4]])
    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."""
encrypted_message = bipartite_encrypt(message, matrix)
print("Encrypted text:", encrypted_message)

# write to file as UTF-8
with open("encrypted.txt", "w", encoding="utf-8") as file:
    file.write(encrypted_message)

decrypted_message = bipartite_decrypt(encrypted_message, matrix)
print("Decrypted text:", decrypted_message)

# write to file as UTF-8
with open("decrypted.txt", "w", encoding="utf-8") as file:
    file.write(decrypted_message)
```

Процес дешифрування:

Маємо наступний текст закодований за допомогою біграмного шифру

OUOVFLDPILDSTRRZVLOUKFMFVAKFSXPURXDPNLOUDPNLGLJW
OUTTGVYQHIAKC
HYQSHJXTYRXEZOUFLYXUBTIZBCHVNHIWPFIBCVMLRFCWMNJRF
VVCFBBKHDCCH

VLSVHOLEDWYHIOEZTGVLOUOVNLILNNDWCBFYLXKHILJAXNXN
FOUJNJUVRYBDKF
GTBWAZHFFMARKUPSQVFYRDRZTVDPITBKXMFVJZUVFISQHQORY
ZMBLGHOTGNJT
JCHGVOJUJJWOUCFSYDPZQWWLRUBHOXNHIFIHOLEOUVLBHRZQP
FQKFVLOUCFS
YDPZQTSWMOUDPCFFPBDQPSQNTYXUBSYXFOTDCEFZDXVVYPS
WFLXNNPMKDEN
PMCBPUSIFLGTWPOUEZDKTLBCNKDXARXVFLXMMTMGNJKFRHT
EWFXNHIYQTYOW
GBDCUJVLTEHQBYAXKFDRITTWFIFMSHFKWKNNTSWMZQOJUJZD
BLZTGVFOPFOVR
ZRYDEOUDPZDEZITHWAZRZOJVLPGWKOTBNZDVLSXFFZTDPYXUB
TICFDVDPWNFF
NTTSSYHFITMXBLLRGTITMWYPWPOUDPLRFLYPPUFOSKJXIRRZHF
SVAGGVYZXIFIT
GAZRZOJNNWEHQFVKJKDFYKFSXPURXDPQVSYFOGZNTUNNTUDH
OFWOVJZEZNA
YQOVUEJWPMYXPSNGGVDPTSVLVLMUGLTIUDFCDSHDRMBDSXFF
ZTDPYXUBTIEZ
WMBDOUKHBDSIZDEPYZXIFITGAZRZOJBDHFSVAGANWMTBTYGG
UBUGGZNLBDTB
OWJXRZTATIYZWYJXRZGTITKATYMXFIPSHFXMIRHOPSRYSPVQQX
RYFFQXQVIIJXG
BTREZNESMPMYXPSUEYPDEFCDSHDRMTVWYOTFLYPPUVLSXFFZ
TDPYXUBTIEZW
MOWGBRMWTVLLD

Спробуємо його розшифрувати

Розглянемо частоти з яким з’являються букви і біграми
Спочатку розглянемо біграми:

Bigram: OU Frequency: 8.23529 17

Bigram: DP Frequency: 8.82353 15

Bigram: VL Frequency: 7.05882 12

Bigram: RZ Frequency: 5.88235 10

Bigram: EZ Frequency: 4.70588 8

Bigram: KF Frequency: 4.11765 7

Bigram: FI Frequency: 4.11765 7

Природнім здається прийняти OU за TH. A VL або DP за ER або HE.

Є слово OUDP воно може бути THER або THIS. Припустимо,що це перший варіант

Є природнім припустити, що оскільки шифр є біграмним, то він використовує матрицю

переходу A= ((a,b),(c,d)) над якимось полем, $(x, y)^T$ а результат $(res1, res2)^T$ Оскільки тут

фігурують числа, а в тексті тільки букви, то природнім є припущення про те, що кожному символу відповідає його символ в юнікоді: А матриця переходу є матрицею переходу над якимось полем

Тоді отримуємо номери наших символів в юнікоді отримаємо

O: 79

U: 85

D: 68

P: 80

T- 84

H: 72 E:69
R: 82

Тепер розв’яжемо матричне рівняння отримаємо, що

84*a+72*b = 79

84*c+72*d = 85

69a+82*b=68
69*c+82*d=80

Розв’язавши цю систему отримаємо, що

A = (1,2,3,4). Отримали цілучисельну матрицю. Спробуємо декодувати увесь текст. Для цього напишемо програму на Python. Отримаємо, що декодований текст має вигляд

THEMATERIALGIVENINTHISWORKISCONSIDEREDTHEREDUETOTH
EFACTTHATLAST
TIMETHEIDEATHATSCIENCEHASEXHAUSTEDITSELFANDITSROLEIS
DECREASINGEV
ERYDAYAPPEARSINTHEMEDIAANDAMONGSPECIALISTSTHISSITUA
TIONISESPECIAL
LYSURPRISINGWHENAPERSONWHOGRADUATEDFROMSOMEUNIVE
RSITYHASACCE
SSTOTHEINTERNETBELIEVESTHATEVERYTHINGHENEEDSISINTHEI
NTERNETANDTH
EREISNONEEDFORSCIENTIFICRESEARCHVARIOUSPANDEMICSDEM
ONSTRATESUS
THEABSURDITYOFSUCHATHOUGHTITISOBVIOUSTHATTHEPROGRE
SSINVIROLOGYI
SABSOLUTELYIMPORTANTANDNECESSARYBUTACHIEVEMENTINOT
HERAREASOFSC
IENCEINPARTICULARINCOMPUTERSCIENCEISVERYIMPORTANTALS
OMAYBELESSOB
VIOUSTHERELATIONSHIPBETWEENALGEBRACOMPUTERSCIENCEA
NDPROGRAMMI
NGISCONSIDERSINTHISWORKMOREOVERWEMADEANATTEMPTTO
DESCRIBEACERT
AININFLUENCEOFALGEBRAONCOMPUTERSCIENCEANDONTHECON
TRARYCOMPUT
ERSCIENCEONALGEBRAANDOTHERFIELDSWEDONOTPRETENDONC
OMPLETENESS
OFTHEMATERIALHOWEVERITISALMOSTIMPOSSIBLETOGIVEAFULL
DESCRIPTIONOF
ALGEBRAAPPLICATIONSINCOMPUTERSCIENCEANDPROGRAMMING
Розділимо його на осмисленні слова і отримаємо:

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

Очевидно, такого не могло би статися якщо ми підібрали би іншу матрицю(ймовірність прямує до 0). Отже ми правильно знайшли матрицю переходу і правильно декодували текст

Література:

- Лекції з предмету “Математичні основи захисту інформації”
- https://en.wikipedia.org/wiki/Bigram
- https://towardsdatascience.com/data-representation-in-nlp-cc9460f855a7