Київський національний університет імені Тараса Шевченка

Факультет радіофізики, електроніки і комп’ютерних систем

Звіт

З лабораторної роботи №1

Предмету «Комп’ютерні системи»

Студента 3-го курсу

Спеціальності КІ МА

Горелова Артема

Київ 2021

Хід роботи

І. Дослідження кількості інформації в тексті

1. Оберіть 3 текстових файла різного тематичного та лінгвістичного спрямування.

Обрані тексти:

* Фрагмент листа Нестора ІвановичаМахно з власних мемуарів (mahno.txt)
* Уривок з “Симулякри і симуляція” Жана Бодрійяра (bodriyar.txt)
* Фрагмент інтерв'ю Леоніда Макаровича Кравчука (kravchuk.txt)

[Посилання на тексти](https://github.com/prodampo4ky/CS_Labs/tree/main/lab1)

2. Переконайтесь, що тексти, які ви використовуєте є унікальними і не повторюються у ваших колег.

3. Створіть програму (будь-якою зручною для вас мовою), яка в якості вхідних даних приймає текстовий файл, та аналізує його вміст

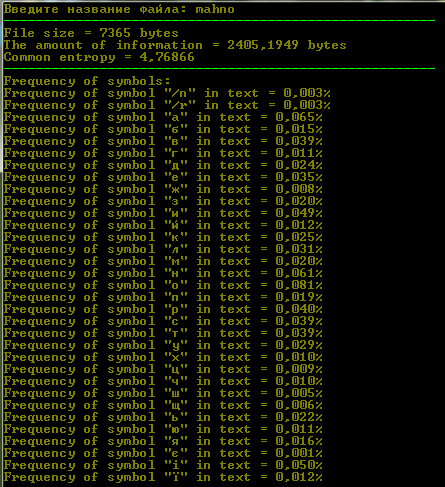
[Лістинг коду програми](https://github.com/prodampo4ky/CS_Labs/blob/main/lab1/lab1.1/Program.cs)

4. Проведіть стиснення кожного вхідного файлу за допомогою 5 різних алгоритмів стиснення

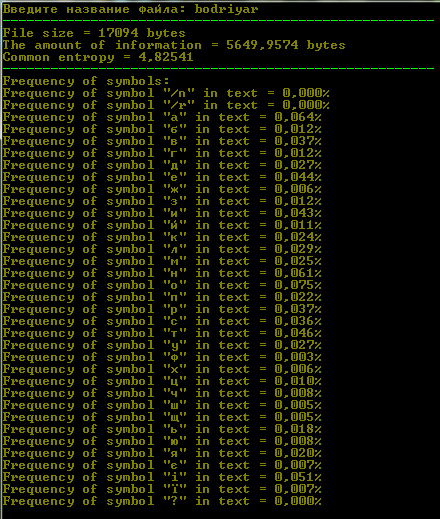
Дані у вигляді таблиці:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Назва файлу | Розмір файлу | Кількість інформації | 7z | gzip | rar | bzip2 | zip |
| mahno.txt | 7365 | 2405 | 2533 | 2492 | 2614 | 2041 | 2616 |
| bodriyar.txt | 17094 | 5650 | 5335 | 5413 | 5694 | 4435 | 5540 |
| kravchuk.txt | 20634 | 7046 | 6091 | 6204 | 6501 | 5112 | 6331 |

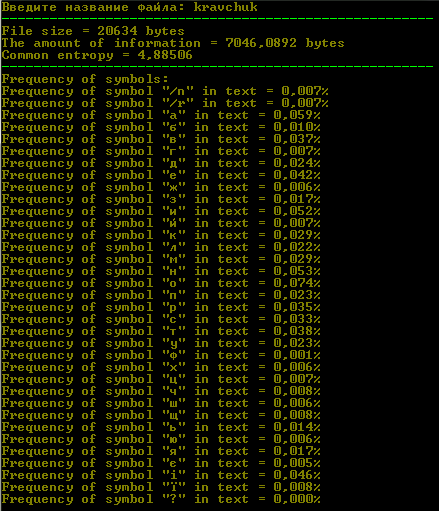
Алфавіт першого файлу:



Алфавіт другого файлу:

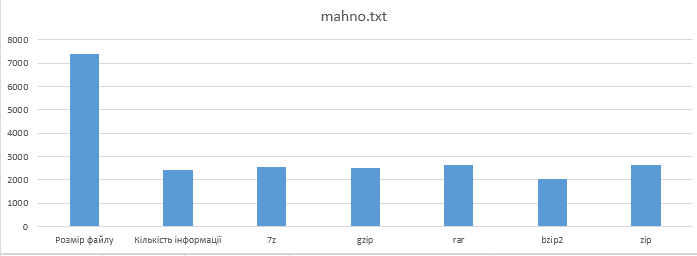


Алфавіт третього файлу:

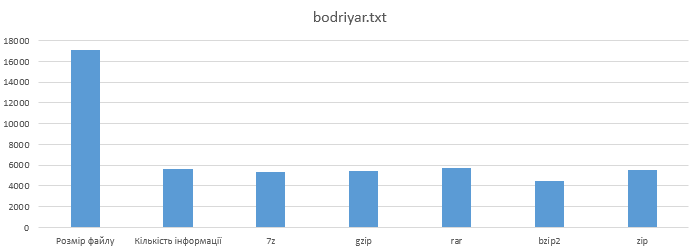


Графіки:

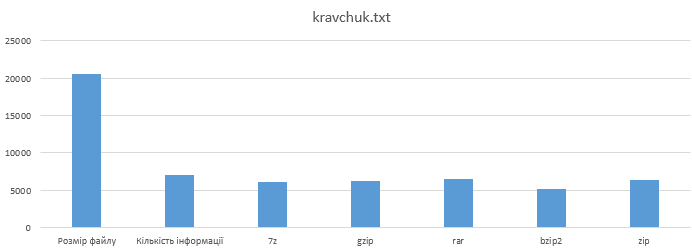
Перший файл:



Другий файл:



Третій файл:



Висновок

У всіх випадках найкращий результат показав алгоритм *bzip2*. На другому місці алгоритм *gzip*. Алгоритми *7z*, *rar* та *zip* продемонстрували найгірший результат; об’єм файлів, закодованих ними перевищував кількість інформації порівняно з *bz2* і *gz*.

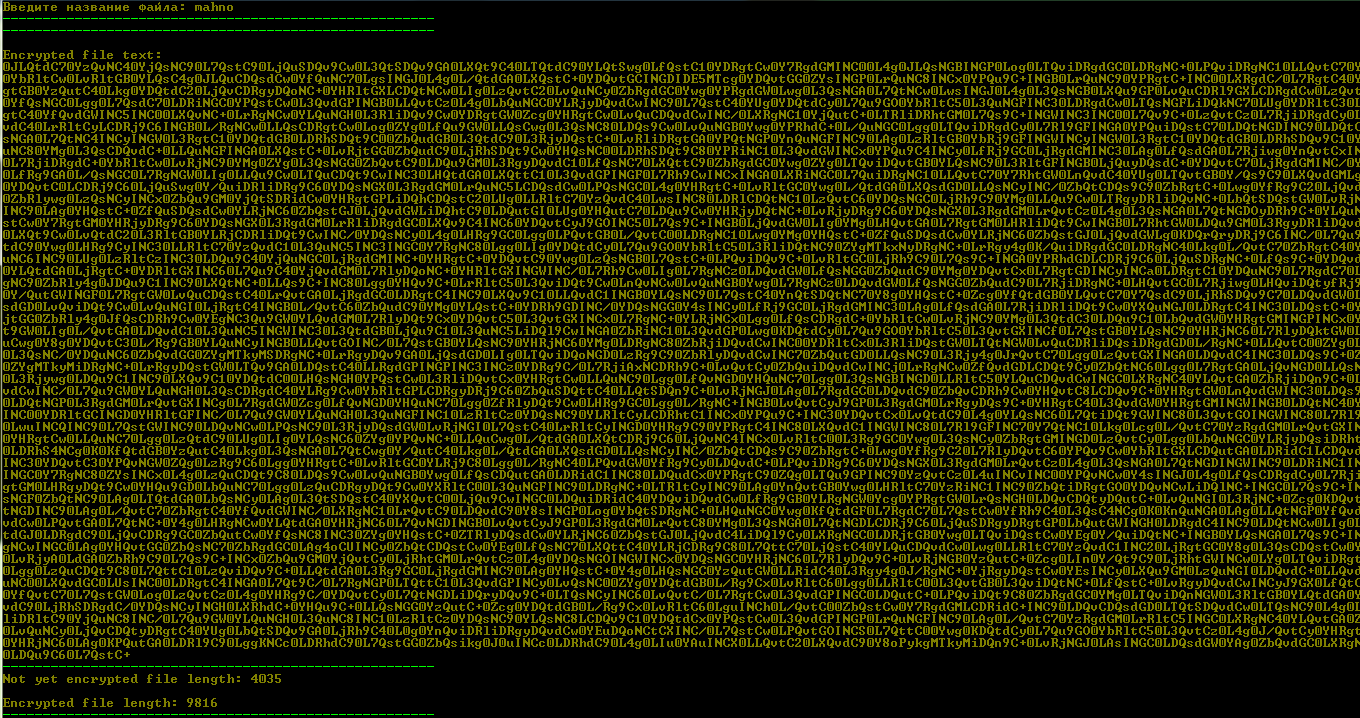
ІІ. Дослідження способів кодування інформації на прикладі Base64

1. Ознайомтесь зі стандартом RFC4648

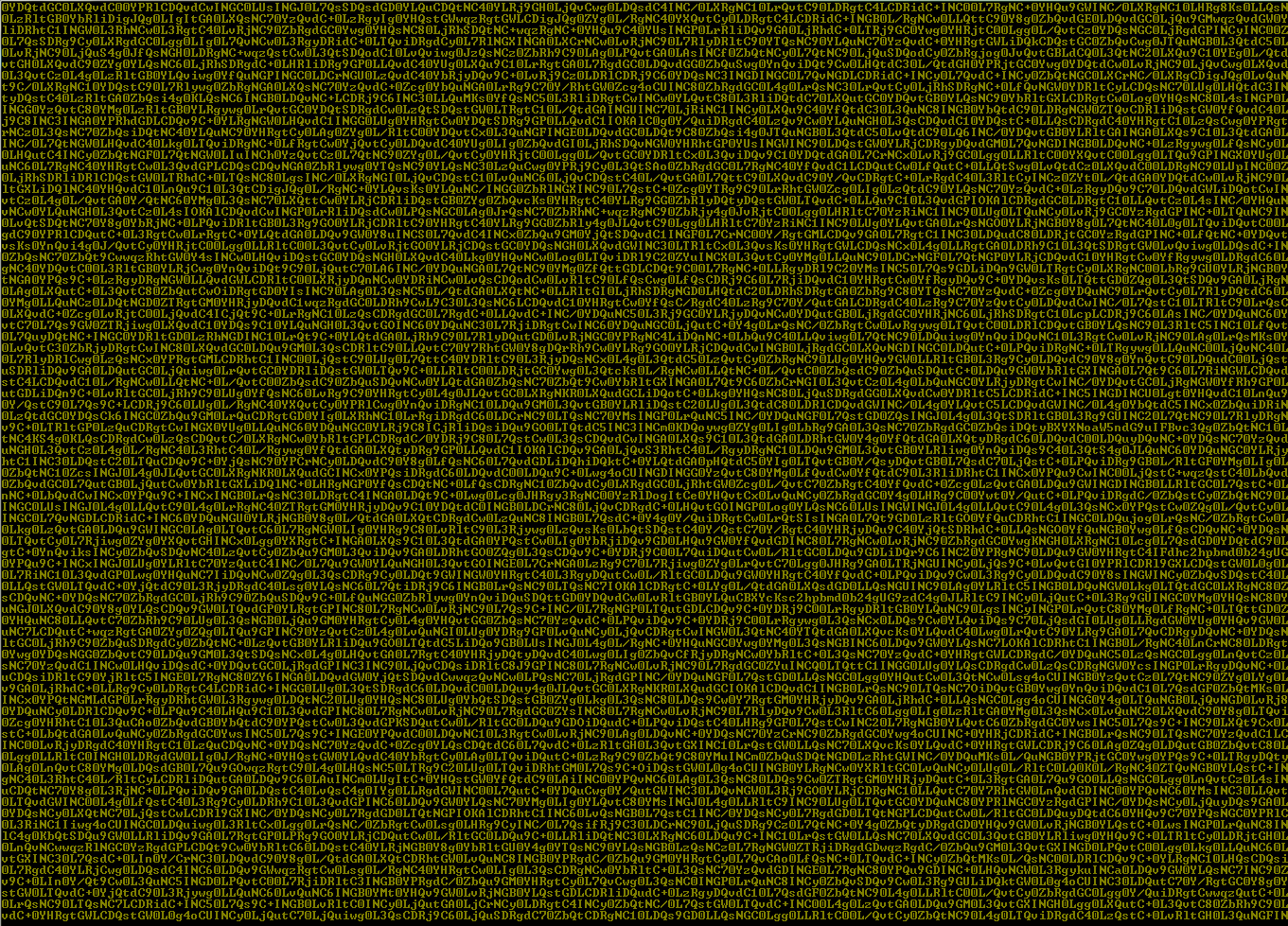
2. Для практичного засвоєння методу кодування, створіть програму, що кодує довільний файл в Base64

[Лістинг коду програми](https://github.com/prodampo4ky/CS_Labs/blob/main/lab1/lab1.2/Program.cs)

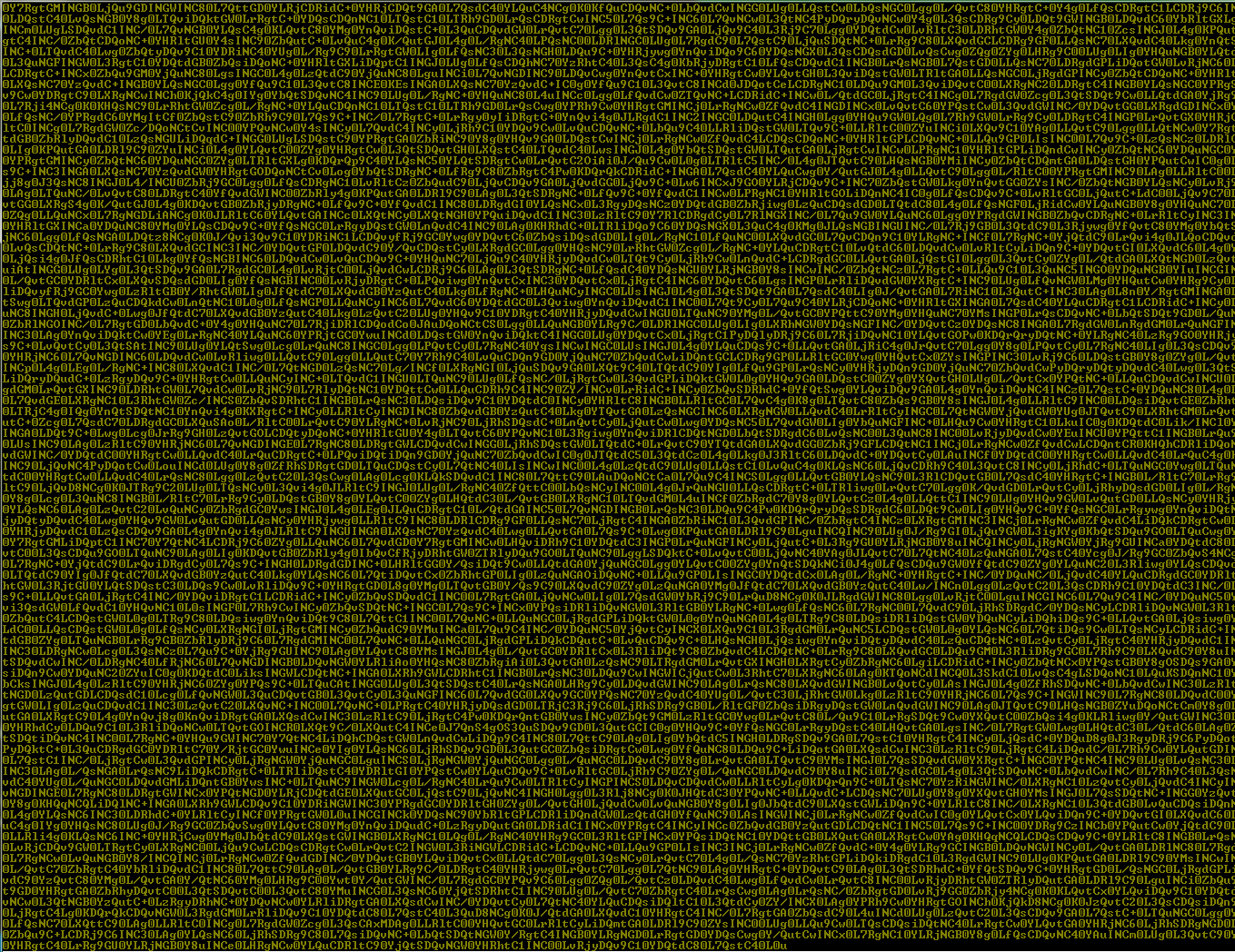
Перший файл:



Другий файл:



Третій файл:



3. Закодуйте в Base64 обрані вами текстові файли

Кодування виконано у минулому пункті.

|  |  |
| --- | --- |
|  | Кількість інформації(base64)/Кількість інформації(file) |
| mahno.txt | 1,333469 |
| bodriyar.txt | 1,333391 |
| kravchuk.txt | 1,(3) |

Згідно таблиці в середньому кількість інформації у файлі, що був перекодований у base64 більша у 1.33 рази ніж кількость інформації в файлі.

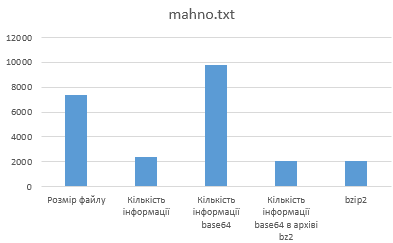
4. Закодуйте в Base64 стиснені кращим з алгоритмів текстові файли

Я обрав розширення .bz2, так як такі архіви мали найменший розмір серед усіх інших у моєму випадку.

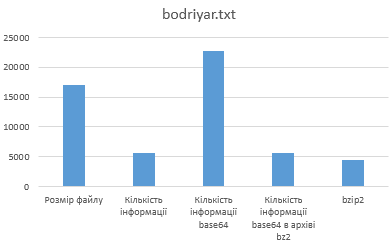
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Розмір файлу | Кількість інформації | Кількість інформації base64 | Кількість інформації base64 в архіві bz2 | bzip2 |
| mahno.txt | 5280 | 2405 | 9803 | 2041 | 2041 |
| bodriyar.txt | 2440 | 5650 | 22787 | 5681 | 4435 |
| kravchuk.txt | 3218 | 7046 | 27399 | 6611 | 5112 |

Графіки:

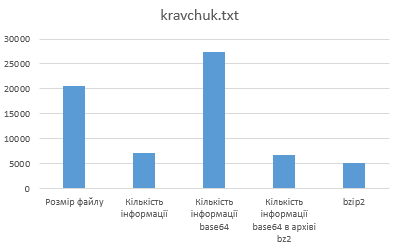
Перший файл:



Другий файл:



Третій файл:



**Висновок:** підчас виконання лабораторної роботи було досліджено частоти появи символу, ентропію нерівномірного алфавіту, як ці параметри впливають на кількість інформації в тексті. На практиці було виявлено, що кращим алгоритмом стискання в рамках лабораторної роботи був алгоритм *.bz2*. Також було реалізовано алгоритм кодування Base64, перевірена правильність його роботи.