

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

Хід роботи

I. Дослідження кількості інформації в тексті

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

Обрані тексти:

- Фрагмент листа Нестора Івановича Махно 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

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

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:

```

Введите название файла: mahno
-----
File size = 7365 bytes
The amount of information = 2405,1949 bytes
Common entropy = 4,76866
-----
Frequency of symbols:
Frequency of symbol "/n" in text = 0,003%
Frequency of symbol "/r" in text = 0,003%
Frequency of symbol "a" in text = 0,065%
Frequency of symbol "б" in text = 0,015%
Frequency of symbol "в" in text = 0,039%
Frequency of symbol "г" in text = 0,011%
Frequency of symbol "д" in text = 0,024%
Frequency of symbol "е" in text = 0,035%
Frequency of symbol "ж" in text = 0,008%
Frequency of symbol "з" in text = 0,020%
Frequency of symbol "и" in text = 0,049%
Frequency of symbol "й" in text = 0,012%
Frequency of symbol "к" in text = 0,025%
Frequency of symbol "л" in text = 0,031%
Frequency of symbol "м" in text = 0,020%
Frequency of symbol "н" in text = 0,061%
Frequency of symbol "о" in text = 0,081%
Frequency of symbol "п" in text = 0,019%
Frequency of symbol "р" in text = 0,040%
Frequency of symbol "с" in text = 0,039%
Frequency of symbol "т" in text = 0,039%
Frequency of symbol "у" in text = 0,029%
Frequency of symbol "х" in text = 0,010%
Frequency of symbol "ц" in text = 0,009%
Frequency of symbol "ч" in text = 0,010%
Frequency of symbol "ш" in text = 0,005%
Frequency of symbol "щ" in text = 0,006%
Frequency of symbol "ь" in text = 0,022%
Frequency of symbol "ю" in text = 0,011%
Frequency of symbol "я" in text = 0,016%
Frequency of symbol "е" in text = 0,001%
Frequency of symbol "i" in text = 0,050%
Frequency of symbol "I" in text = 0,012%

```

Алфавіт другого файлу:

```

Введите название файла: bodriyar
-----
File size = 17094 bytes
The amount of information = 5649,9574 bytes
Common entropy = 4,82541
-----
Frequency of symbols:
Frequency of symbol "/n" in text = 0,000%
Frequency of symbol "/r" in text = 0,000%
Frequency of symbol "a" in text = 0,064%
Frequency of symbol "б" in text = 0,012%
Frequency of symbol "в" in text = 0,037%
Frequency of symbol "г" in text = 0,012%
Frequency of symbol "д" in text = 0,027%
Frequency of symbol "е" in text = 0,044%
Frequency of symbol "ж" in text = 0,006%
Frequency of symbol "з" in text = 0,012%
Frequency of symbol "и" in text = 0,043%
Frequency of symbol "й" in text = 0,011%
Frequency of symbol "к" in text = 0,024%
Frequency of symbol "л" in text = 0,029%
Frequency of symbol "м" in text = 0,025%
Frequency of symbol "н" in text = 0,061%
Frequency of symbol "о" in text = 0,075%
Frequency of symbol "п" in text = 0,022%
Frequency of symbol "р" in text = 0,037%
Frequency of symbol "с" in text = 0,036%
Frequency of symbol "т" in text = 0,046%
Frequency of symbol "у" in text = 0,027%
Frequency of symbol "ф" in text = 0,003%
Frequency of symbol "х" in text = 0,006%
Frequency of symbol "ц" in text = 0,010%
Frequency of symbol "ч" in text = 0,008%
Frequency of symbol "ш" in text = 0,005%
Frequency of symbol "щ" in text = 0,005%
Frequency of symbol "ь" in text = 0,018%
Frequency of symbol "ю" in text = 0,008%
Frequency of symbol "я" in text = 0,020%
Frequency of symbol "е" in text = 0,007%
Frequency of symbol "i" in text = 0,051%
Frequency of symbol "I" in text = 0,007%
Frequency of symbol "?" in text = 0,000%

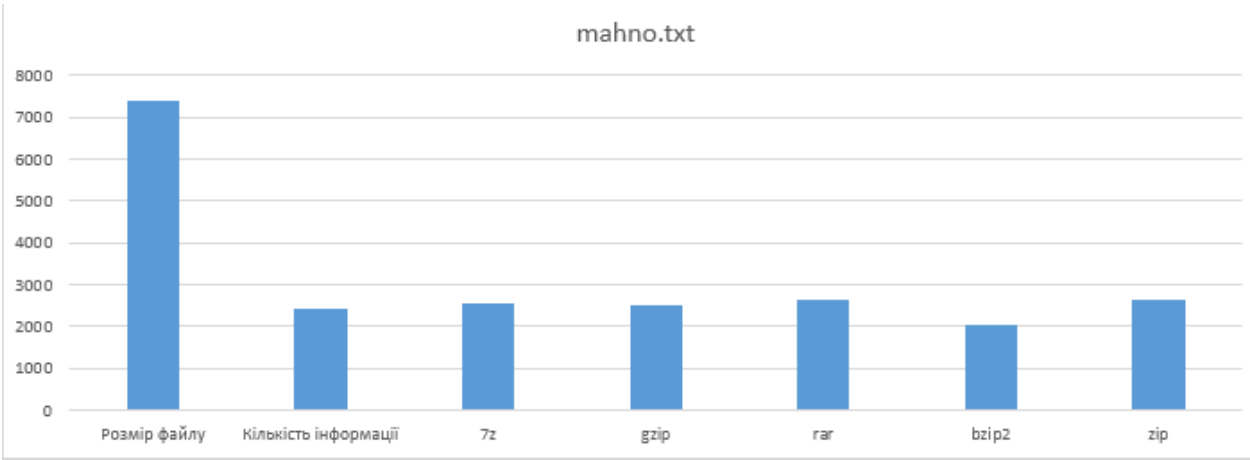
```

Алфавіт третього файлу:

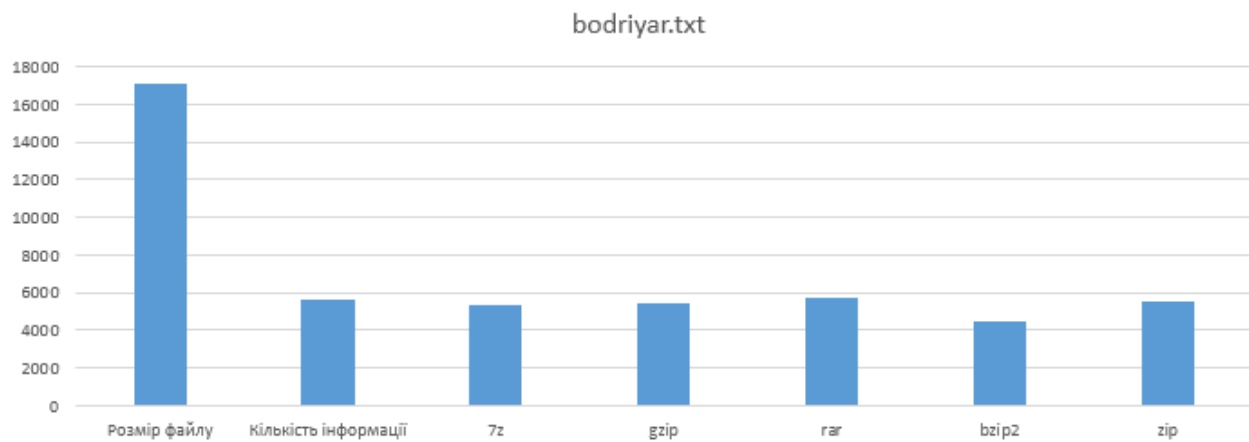
```
Введите название файла: kravchuk
-----
File size = 20634 bytes
The amount of information = 7046,0892 bytes
Common entropy = 4,88506
-----
Frequency of symbols:
Frequency of symbol "/n" in text = 0,007%
Frequency of symbol "/r" in text = 0,007%
Frequency of symbol "a" in text = 0,059%
Frequency of symbol "б" in text = 0,010%
Frequency of symbol "в" in text = 0,037%
Frequency of symbol "г" in text = 0,007%
Frequency of symbol "д" in text = 0,024%
Frequency of symbol "е" in text = 0,042%
Frequency of symbol "ж" in text = 0,006%
Frequency of symbol "з" in text = 0,017%
Frequency of symbol "и" in text = 0,052%
Frequency of symbol "й" in text = 0,007%
Frequency of symbol "к" in text = 0,029%
Frequency of symbol "л" in text = 0,022%
Frequency of symbol "м" in text = 0,029%
Frequency of symbol "н" in text = 0,053%
Frequency of symbol "о" in text = 0,074%
Frequency of symbol "п" in text = 0,023%
Frequency of symbol "р" in text = 0,035%
Frequency of symbol "с" in text = 0,033%
Frequency of symbol "т" in text = 0,038%
Frequency of symbol "у" in text = 0,023%
Frequency of symbol "ф" in text = 0,001%
Frequency of symbol "х" in text = 0,006%
Frequency of symbol "ц" in text = 0,007%
Frequency of symbol "ч" in text = 0,008%
Frequency of symbol "ш" in text = 0,006%
Frequency of symbol "щ" in text = 0,008%
Frequency of symbol "ь" in text = 0,014%
Frequency of symbol "ю" in text = 0,006%
Frequency of symbol "я" in text = 0,017%
Frequency of symbol "е" in text = 0,005%
Frequency of symbol "і" in text = 0,046%
Frequency of symbol "ї" in text = 0,008%
Frequency of symbol "?" in text = 0,000%
```

Графіки:

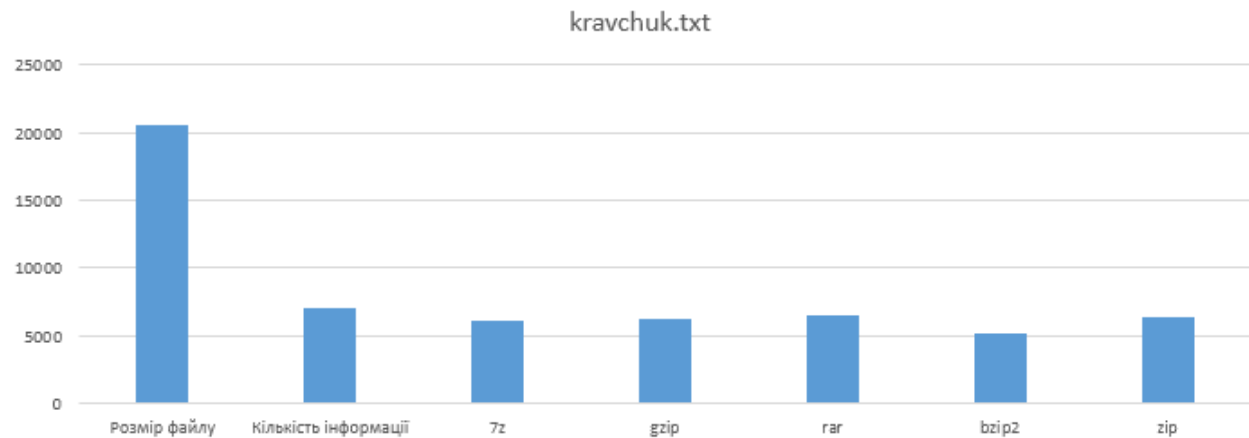
Перший файл:



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*.

II. 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](#)

Перший файл:

Третій файл:

[illegible]

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 кількість інформації в файлі.

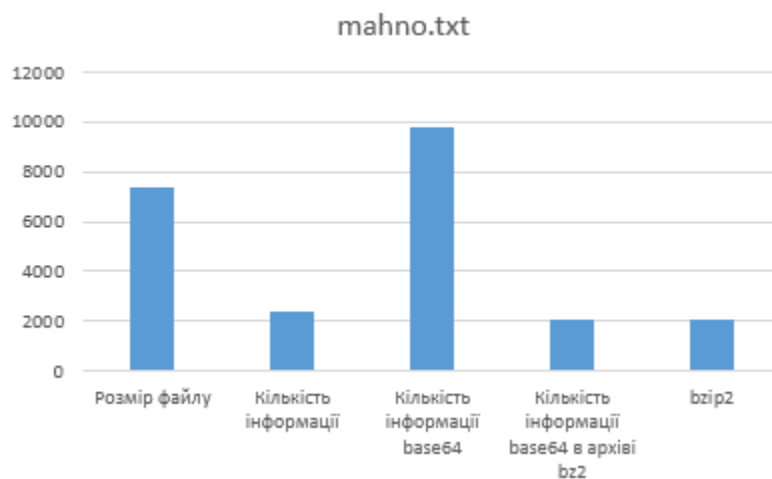
4. Закодуйте в Base64 стиснені кращим з алгоритмів текстові файли

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

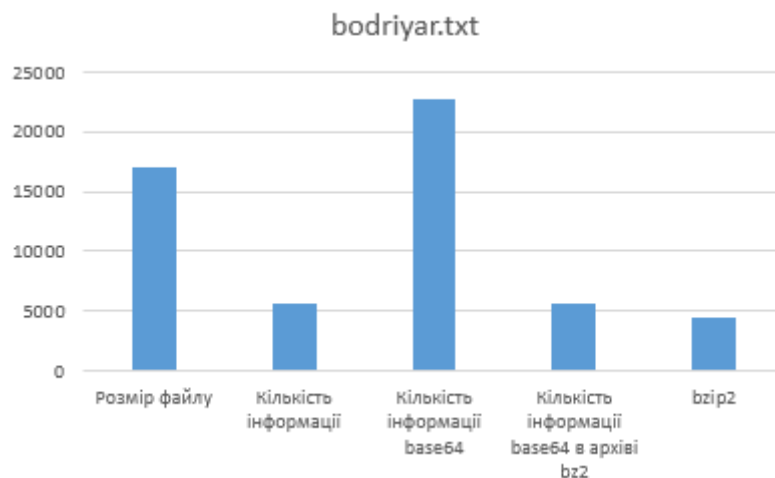
	Розмір файлу	Кількість інформації	Кількість інформації base64	Кількість інформації base64 в архіві bz2	bzip2
mahno.txt	5280	2405	9803	2041	2041
bodriyar.txt	2440	5650	22787	5681	4435
kravchuk.txt	3218	7046	27399	6611	5112

Графіки:

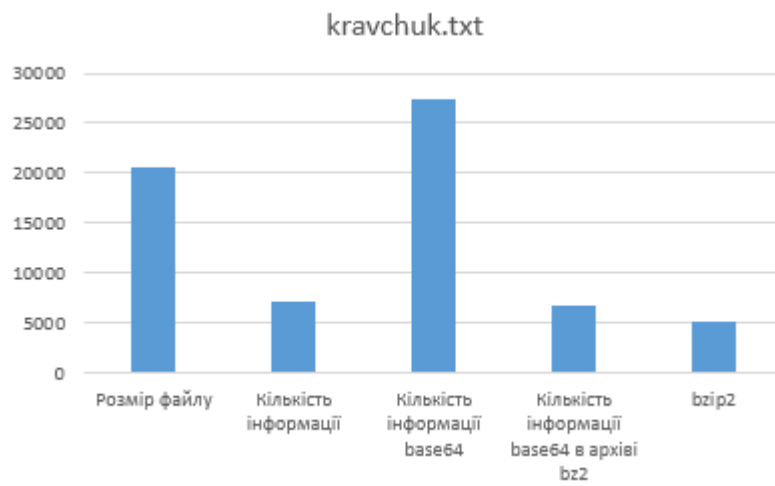
Перший файл:



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.