Київський національний університет імені Тараса Шевченка
Факультет радіофізики, електроніки та комп'ютерних систем
Навчальна дисципліна «Комп'ютерні системи»

Звіт з лабораторної роботи №1
на тему «**Дослідження кількості інформації при різних варіантах кодування**»

Роботу виконав
Студент 3 курсу
КІ, група МА
Мормуль Олексій
Володимирович

Київ 2019

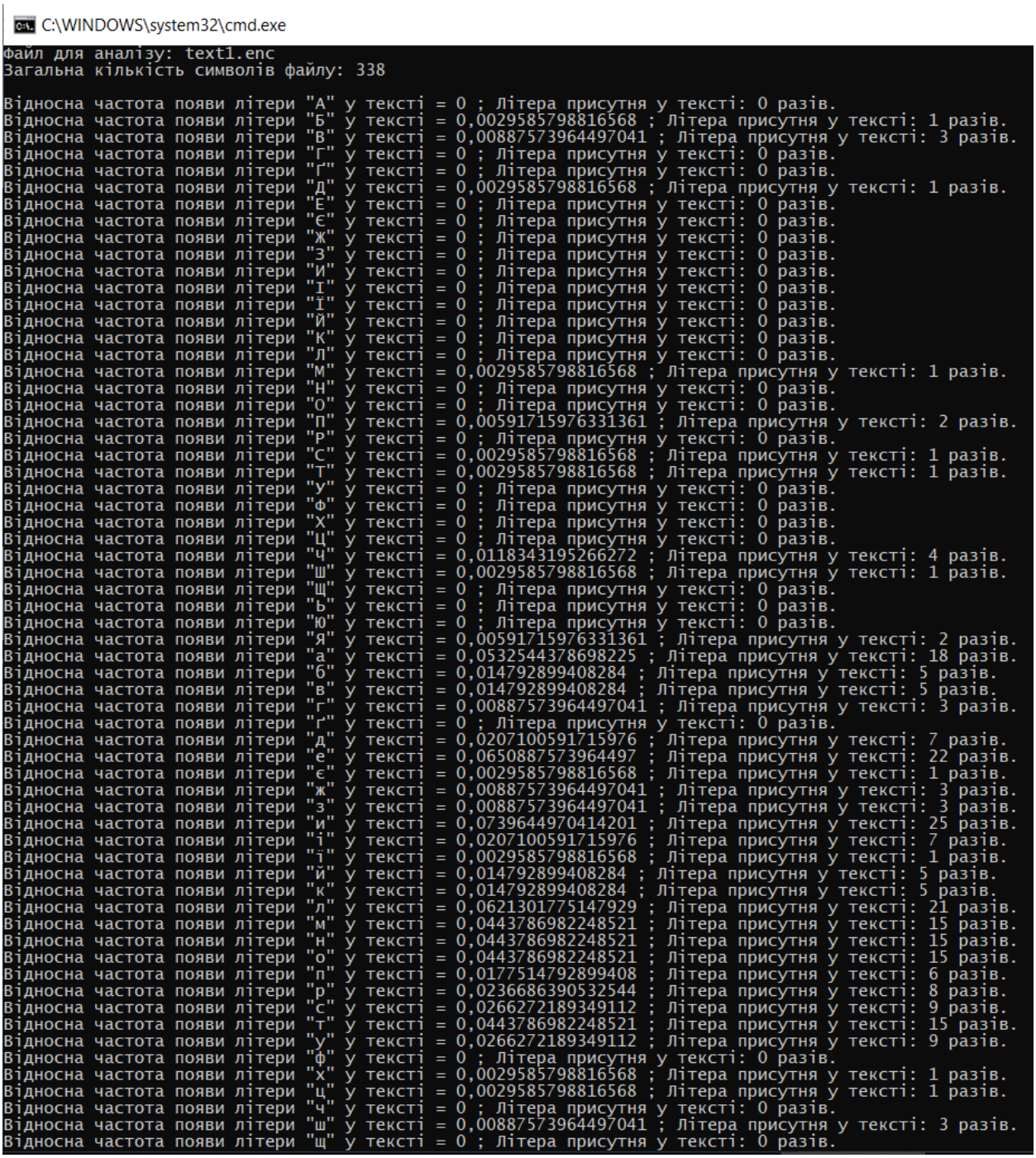
Мета: Дослідити імовірнісні параметри української мови для оцінки кількості інформації текстів. Дослідити вплив різних методів кодування інформації на її кількість.

Хід роботи

Дослідження кількості інформації в тексті

1. Оберіть 3 текстових файла різного тематичного та лінгвістичного спрямування (файли також є у репозиторії): • Об’єкт SCP-173
• Гальмування двигуном
• Перекладений текст пісні Rick Astley - Never Gonna Give You Up (натисніть, щоб піймaтися на рікролл українською)

2. Переконайтесь, що тексти, які ви використовуєте є унікальними і не повторюються у ваших колег! Використовуйте наявні електронні засоби зв’язку та документообігу, щоб уникнути дублювання! Вдруге аналіз того самого тексту не зараховується!
• Заради цього створено мною цей документ
3. Код створеної програми міститься у репозиторії, посилання буде вкінці. Результат роботи програми:


C:\WINDOWS\system32\cmd.exe
Файл для аналізу: text1.enc
Загальна кількість символів файлу: 338
Відносна частота появи літери "А" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Б" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "В" у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "Г" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ґ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Д" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Е" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Є" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ж" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "З" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "И" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "І" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ї" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Й" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "К" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Л" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "М" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Н" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "О" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "П" у тексті = 0,00591715976331361 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Р" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "С" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Т" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "У" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ф" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Х" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ц" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ч" у тексті = 0,0118343195266272 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "Ш" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Щ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ь" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ю" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Я" у тексті = 0,00591715976331361 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "а" у тексті = 0,0532544378698225 ; Літера присутня у тексті: 18 разів.
Відносна частота появи літери "б" у тексті = 0,014792899408284 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "в" у тексті = 0,014792899408284 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "г" у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "ґ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "д" у тексті = 0,02071005917159776 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "е" у тексті = 0,0650887573964497 ; Літера присутня у тексті: 22 разів.
Відносна частота появи літери "є" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "ж" у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "з" у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "и" у тексті = 0,0739644970414201 ; Літера присутня у тексті: 25 разів.
Відносна частота появи літери "і" у тексті = 0,02071005917159776 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "ї" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "й" у тексті = 0,014792899408284 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "к" у тексті = 0,014792899408284 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "л" у тексті = 0,06213017751479289 ; Літера присутня у тексті: 21 разів.
Відносна частота появи літери "м" у тексті = 0,04437869822485207 ; Літера присутня у тексті: 15 разів.
Відносна частота появи літери "н" у тексті = 0,04437869822485207 ; Літера присутня у тексті: 15 разів.
Відносна частота появи літери "о" у тексті = 0,04437869822485207 ; Літера присутня у тексті: 15 разів.
Відносна частота появи літери "п" у тексті = 0,017751479289940828 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "р" у тексті = 0,023668639053254437 ; Літера присутня у тексті: 8 разів.
Відносна частота появи літери "с" у тексті = 0,026627218934911243 ; Літера присутня у тексті: 9 разів.
Відносна частота появи літери "т" у тексті = 0,04437869822485207 ; Літера присутня у тексті: 15 разів.
Відносна частота появи літери "у" у тексті = 0,026627218934911243 ; Літера присутня у тексті: 9 разів.
Відносна частота появи літери "ф" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "х" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "ц" у тексті = 0,0029585798816568 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "ч" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "ш" у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "щ" у тексті = 0 ; Літера присутня у тексті: 0 разів.

```
Відносна частота появи літери "щ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "ь" у тексті = 0,0177514792899408 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "ю" у тексті = 0,0236686390532544 ; Літера присутня у тексті: 8 разів.
Відносна частота появи літери "я" у тексті = 0,0177514792899408 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "," у тексті = 0,029585798816568 ; Літера присутня у тексті: 10 разів.
Відносна частота появи літери " " у тексті = 0,14792899408284 ; Літера присутня у тексті: 50 разів.
Відносна частота появи літери "(" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери ")" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "." у тексті = 0,00887573964497041 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери ":" у тексті = 0,00591715976331361 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "-" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери """ у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "0" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "1" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "2" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "3" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "4" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "5" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "6" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "7" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "8" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "9" у тексті = 0 ; Літера присутня у тексті: 0 разів.

Середня ентропія нерівноймовірного алфавіту у заданому тексті: 4,6186527830823
Кількість інформації у тексті: 195,138080085227
```

C:\WINDOWS\system32\cmd.exe

```
Файл для аналізу: text2.enc
Загальна кількість символів файлу: 4694

Відносна частота появи літери "А" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "Б" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "В" у тексті = 0,00213037920749893 ; Літера присутня у тексті: 10 разів.
Відносна частота появи літери "Г" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ґ" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Д" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Е" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Є" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ж" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "З" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "И" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "І" у тексті = 0,00106518960374947 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "Ї" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Й" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "К" у тексті = 0,00149126544524925 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "Л" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "М" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Н" у тексті = 0,00127822752449936 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "О" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "П" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "Р" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "С" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "Т" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "У" у тексті = 0,00149126544524925 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "Ф" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Х" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ц" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Ч" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ш" у тексті = 0,000213037920749893 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Щ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ь" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ю" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Я" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "а" у тексті = 0,0602897315722199 ; Літера присутня у тексті: 283 разів.
Відносна частота появи літери "б" у тексті = 0,0097997443544951 ; Літера присутня у тексті: 46 разів.
Відносна частота появи літери "в" у тексті = 0,0481465700894759 ; Літера присутня у тексті: 226 разів.
Відносна частота появи літери "г" у тексті = 0,00937366851299531 ; Літера присутня у тексті: 44 разів.
Відносна частота появи літери "ґ" у тексті = 0,000426075841499787 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "д" у тексті = 0,0217298679164891 ; Літера присутня у тексті: 102 разів.
Відносна частота появи літери "е" у тексті = 0,0308904985087346 ; Літера присутня у тексті: 145 разів.
Відносна частота появи літери "є" у тексті = 0,0066041755432467 ; Літера присутня у тексті: 31 разів.
Відносна частота появи літери "ж" у тексті = 0,00489987217724755 ; Літера присутня у тексті: 23 разів.
Відносна частота появи літери "з" у тексті = 0,0117170856412441 ; Літера присутня у тексті: 55 разів.
Відносна частота появи літери "и" у тексті = 0,0449510012782275 ; Літера присутня у тексті: 211 разів.
Відносна частота появи літери "і" у тексті = 0,0521942905837239 ; Літера присутня у тексті: 245 разів.
Відносна частота появи літери "ї" у тексті = 0,010012782275245 ; Літера присутня у тексті: 47 разів.
Відносна частота появи літери "й" у тексті = 0,0151256923732424 ; Літера присутня у тексті: 71 разів.
Відносна частота появи літери "к" у тексті = 0,0387729015764806 ; Літера присутня у тексті: 182 разів.
Відносна частота появи літери "л" у тексті = 0,0285470813804857 ; Літера присутня у тексті: 134 разів.
Відносна частота появи літери "м" у тексті = 0,0232211333617384 ; Літера присутня у тексті: 109 разів.
Відносна частота появи літери "н" у тексті = 0,0566680869194717 ; Літера присутня у тексті: 266 разів.
Відносна частота появи літери "о" у тексті = 0,0692373242437154 ; Літера присутня у тексті: 325 разів.
Відносна частота появи літери "п" у тексті = 0,0138474648487431 ; Літера присутня у тексті: 65 разів.
Відносна частота появи літери "р" у тексті = 0,0364294844482318 ; Літера присутня у тексті: 171 разів.
Відносна частота появи літери "с" у тексті = 0,0443118875159778 ; Літера присутня у тексті: 208 разів.
Відносна частота появи літери "т" у тексті = 0,0328078397954836 ; Літера присутня у тексті: 154 разів.
Відносна частота появи літери "у" у тексті = 0,0281210055389859 ; Літера присутня у тексті: 132 разів.
Відносна частота появи літери "ф" у тексті = 0,0031955688112484 ; Літера присутня у тексті: 15 разів.
Відносна частота появи літери "х" у тексті = 0,00745632722624627 ; Літера присутня у тексті: 35 разів.
Відносна частота появи літери "ц" у тексті = 0,00703025138474649 ; Літера присутня у тексті: 33 разів.
Відносна частота появи літери "ч" у тексті = 0,0106518960374947 ; Літера присутня у тексті: 50 разів.
Відносна частота появи літери "ш" у тексті = 0,0066041755432467 ; Літера присутня у тексті: 31 разів.
```

```
Відносна частота появи літери "щ" у тексті = 0,00191734128674904 ; Літера присутня у тексті: 9 разів.
Відносна частота появи літери "ь" у тексті = 0,0249254367277375 ; Літера присутня у тексті: 117 разів.
Відносна частота появи літери "ю" у тексті = 0,00511291009799744 ; Літера присутня у тексті: 24 разів.
Відносна частота появи літери "я" у тексті = 0,0157648061354921 ; Літера присутня у тексті: 74 разів.
Відносна частота появи літери "," у тексті = 0,0136344269279932 ; Літера присутня у тексті: 64 разів.
Відносна частота появи літери " " у тексті = 0,137835534725181 ; Літера присутня у тексті: 647 разів.
Відносна частота появи літери "(" у тексті = 0,00489987217724755 ; Літера присутня у тексті: 23 разів.
Відносна частота появи літери ")" у тексті = 0,00489987217724755 ; Літера присутня у тексті: 23 разів.
Відносна частота появи літери "." у тексті = 0,0129953131657435 ; Літера присутня у тексті: 61 разів.
Відносна частота появи літери ":" у тексті = 0,00063911376224968 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "-" у тексті = 0,00106518960374947 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери """ у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "0" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "1" у тексті = 0,0034086067319983 ; Літера присутня у тексті: 16 разів.
Відносна частота появи літери "2" у тексті = 0,00127822752449936 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "3" у тексті = 0,00063911376224968 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "4" у тексті = 0,00127822752449936 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "5" у тексті = 0,00149126544524925 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "6" у тексті = 0,00063911376224968 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "7" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "8" у тексті = 0,000852151682999574 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "9" у тексті = 0,00106518960374947 ; Літера присутня у тексті: 5 разів.

Середня ентропія нерівноймовірного алфавіту у заданому тексті: 4,78265636880096
Кількість інформації у тексті: 2806,22362439397
```

```
Файл для аналізу: text3.enc
Загальна кількість символів файлу: 3584

Відносна частота появи літери "А" у тексті = 0,000558035714285714 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Б" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "В" у тексті = 0,000558035714285714 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Г" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ґ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Д" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Е" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Є" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ж" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "З" у тексті = 0,00167410714285714 ; Літера присутня у тексті: 6 разів.
Відносна частота появи літери "И" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "І" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ї" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Й" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "К" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Л" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "М" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Н" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "О" у тексті = 0,00334821428571429 ; Літера присутня у тексті: 12 разів.
Відносна частота появи літери "П" у тексті = 0,00139508928571429 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери "Р" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "С" у тексті = 0,00111607142857143 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "Т" у тексті = 0,000837053571428571 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "У" у тексті = 0,00223214285714286 ; Літера присутня у тексті: 8 разів.
Відносна частота появи літери "Ф" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Х" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ц" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ч" у тексті = 0,000558035714285714 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "Ш" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Щ" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "Ь" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Ю" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "Я" у тексті = 0,000837053571428571 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "а" у тексті = 0,0778459821428571 ; Літера присутня у тексті: 279 разів.
Відносна частота появи літери "б" у тексті = 0,0145089285714286 ; Літера присутня у тексті: 52 разів.
Відносна частота появи літери "в" у тексті = 0,0398995535714286 ; Літера присутня у тексті: 143 разів.
Відносна частота появи літери "г" у тексті = 0,00697544642857143 ; Літера присутня у тексті: 25 разів.
Відносна частота появи літери "ґ" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "д" у тексті = 0,0231584821428571 ; Літера присутня у тексті: 83 разів.
Відносна частота появи літери "е" у тексті = 0,0412946428571429 ; Літера присутня у тексті: 148 разів.
Відносна частота появи літери "є" у тексті = 0,00641741071428571 ; Літера присутня у тексті: 23 разів.
Відносна частота появи літери "ж" у тексті = 0,00530133928571429 ; Літера присутня у тексті: 19 разів.
Відносна частота появи літери "з" у тексті = 0,0172991071428571 ; Літера присутня у тексті: 62 разів.
Відносна частота появи літери "и" у тексті = 0,0513392857142857 ; Літера присутня у тексті: 184 разів.
Відносна частота появи літери "і" у тексті = 0,0418526785714286 ; Літера присутня у тексті: 150 разів.
Відносна частота появи літери "ї" у тексті = 0,00725446428571429 ; Літера присутня у тексті: 26 разів.
Відносна частота появи літери "й" у тексті = 0,009765625 ; Літера присутня у тексті: 35 разів.
Відносна частота появи літери "к" у тексті = 0,0256696428571429 ; Літера присутня у тексті: 92 разів.
Відносна частота появи літери "л" у тексті = 0,0184151785714286 ; Літера присутня у тексті: 66 разів.
Відносна частота появи літери "м" у тексті = 0,0320870535714286 ; Літера присутня у тексті: 115 разів.
Відносна частота появи літери "н" у тексті = 0,0502232142857143 ; Літера присутня у тексті: 180 разів.
Відносна частота появи літери "о" у тексті = 0,0717075892857143 ; Літера присутня у тексті: 257 разів.
Відносна частота появи літери "п" у тексті = 0,0337611607142857 ; Літера присутня у тексті: 121 разів.
Відносна частота появи літери "р" у тексті = 0,0415736607142857 ; Літера присутня у тексті: 149 разів.
Відносна частота появи літери "с" у тексті = 0,0318080357142857 ; Літера присутня у тексті: 114 разів.
Відносна частота появи літери "т" у тексті = 0,0580357142857143 ; Літера присутня у тексті: 208 разів.
Відносна частота появи літери "у" у тексті = 0,0184151785714286 ; Літера присутня у тексті: 66 разів.
Відносна частота появи літери "ф" у тексті = 0,00334821428571429 ; Літера присутня у тексті: 12 разів.
Відносна частота появи літери "х" у тексті = 0,00334821428571429 ; Літера присутня у тексті: 12 разів.
Відносна частота появи літери "ц" у тексті = 0,00753348214285714 ; Літера присутня у тексті: 27 разів.
Відносна частота появи літери "ч" у тексті = 0,00892857142857143 ; Літера присутня у тексті: 32 разів.
Відносна частота появи літери "ш" у тексті = 0,00948660714285714 ; Літера присутня у тексті: 34 разів.
Відносна частота появи літери "щ" у тексті = 0,00251116071428571 ; Літера присутня у тексті: 9 разів.
```

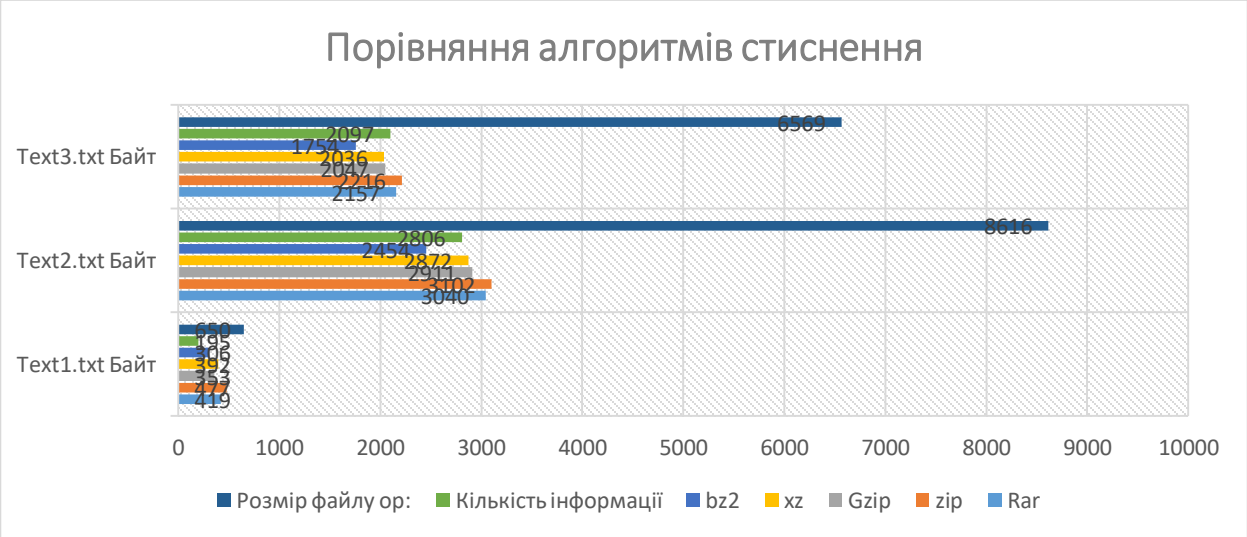
```
Відносна частота появи літери "ь" у тексті = 0,0139508928571429 ; Літера присутня у тексті: 50 разів.
Відносна частота появи літери "ю" у тексті = 0,00837053571428571 ; Літера присутня у тексті: 30 разів.
Відносна частота появи літери "я" у тексті = 0,0189732142857143 ; Літера присутня у тексті: 68 разів.
Відносна частота появи літери "," у тексті = 0,00864955357142857 ; Літера присутня у тексті: 31 разів.
Відносна частота появи літери " " у тексті = 0,132254464285714 ; Літера присутня у тексті: 474 разів.
Відносна частота появи літери "(" у тексті = 0,001953125 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери ")" у тексті = 0,001953125 ; Літера присутня у тексті: 7 разів.
Відносна частота появи літери "." у тексті = 0,0108816964285714 ; Літера присутня у тексті: 39 разів.
Відносна частота появи літери ":" у тексті = 0,000837053571428571 ; Літера присутня у тексті: 3 разів.
Відносна частота появи літери "-" у тексті = 0,00139508928571429 ; Літера присутня у тексті: 5 разів.
Відносна частота появи літери """ у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "0" у тексті = 0,00279017857142857 ; Літера присутня у тексті: 10 разів.
Відносна частота появи літери "1" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "2" у тексті = 0,00111607142857143 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "3" у тексті = 0,00111607142857143 ; Літера присутня у тексті: 4 разів.
Відносна частота появи літери "4" у тексті = 0,000558035714285714 ; Літера присутня у тексті: 2 разів.
Відносна частота появи літери "5" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "6" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "7" у тексті = 0 ; Літера присутня у тексті: 0 разів.
Відносна частота появи літери "8" у тексті = 0,000279017857142857 ; Літера присутня у тексті: 1 разів.
Відносна частота появи літери "9" у тексті = 0 ; Літера присутня у тексті: 0 разів.

Середня ентропія нерівноймовірного алфавіту у заданому тексті: 4,68133589749239
Кількість інформації у тексті: 2097,23848207659
```

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

Для цього завдання я використовував WinRAR та 7-Zip. Для кожного алгоритму я використовував рівень стиснення Normal.

Файл	Text1.txt Байт	Text2.txt Байт	Text3.txt Байт
Rar	419	3040	2157
Zip	477	3102	2216
Gzip	353	2911	2047
Xz	392	2872	2036
Bz2	306	2454	1754
Кількість інформації	195	2806	2097
Розмір файлу ор:	650	8616	6569



5. У результаті ідеального стиснення розмір файлу повинен бути рівним кількості інформації. Але у реальності розміри архівованих файлів у більшості випадків дещо більші за кількість інформації, окрім випадку з великим текстом.

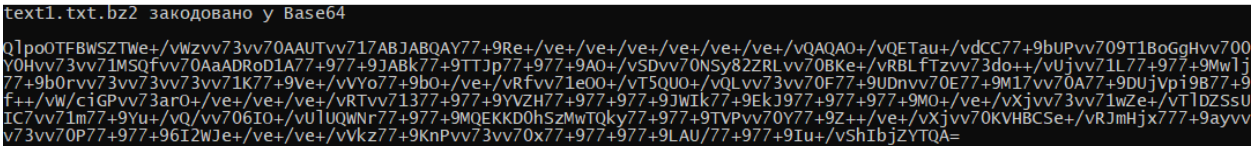
Це відбувається тому, що алгоритми архіваторів побудовані таким чином аби використати повторювані частини тексту. Виходячи з цього, формула розрахунку кількості інформації, використана для програми, не є досконалою, бо вона не враховує передбачення наступного шматочку текста.

• Хочу звернути уваги на алгоритм bzip2, який виявився найефективнішим у всіх випадках. А також навіть упорався «ідеальним стисненням» у випадку великого файлу, тобто стиснений архів має розмір навіть менший, ніж кі-сть інформації. У випадку з піснею, хоч файл і меншого розміру, ніж текст про SCP-173, але є багато частин повторюваного тексту (як-от приспів)

Дослідження способів кодування інформації на прикладі Base64

1. Ознайомтесь зі стандартом RFC4648

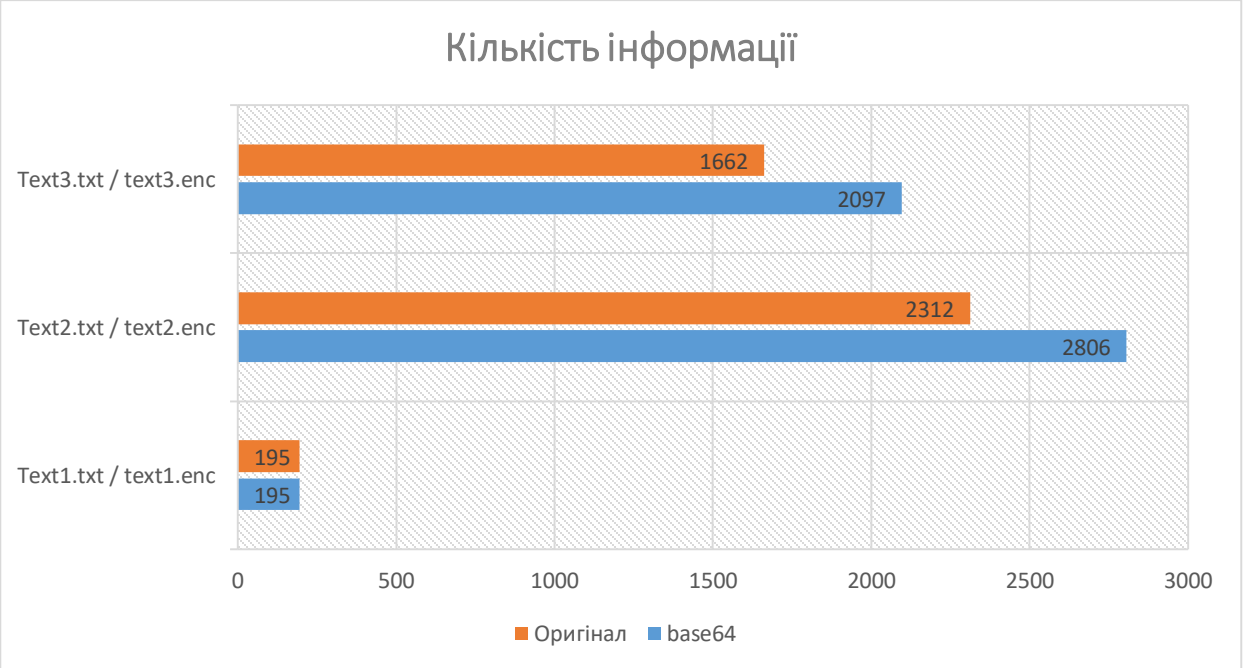
2. Для практичного засвоєння методу кодування, створіть програму, що кодує довільний файл в Base64 (шляхом реалізації алгоритму вручну, а не виклику бібліотечної функції)



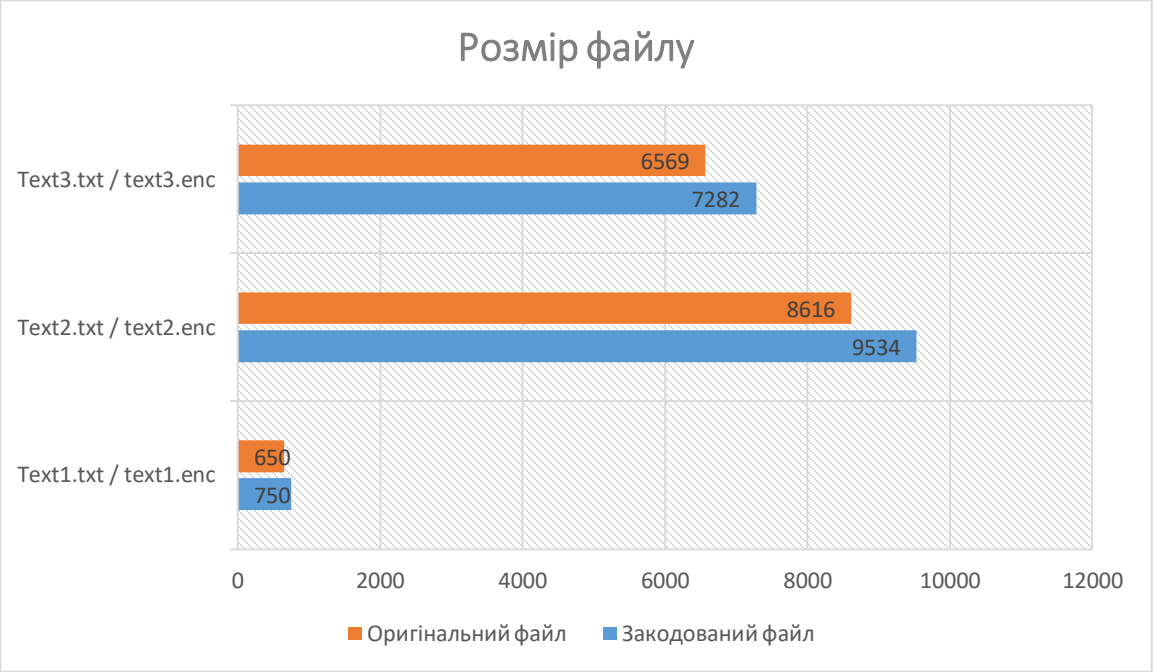
```
text3.txt.bz2 закодовано у Base64

QlpoOTFBWSZTWTXvv73vv73vv70ADO+/vV7vv70AEkDvv71+UO+/vUIQAAAQf++/vdWT77+977+977+977+977+9AAAA77+9B39NQO+/ve+/ve+/vV3vv71s
Vu+/ve+/ve+/ve+/vROO77+9A1NBDEZTaO+/vQh677+9AGnvv71NAA9TQGHvv73vv73vv73EjSbvv70j77+9eu+/ve+/vWhoDQADQAxBKe+/ve+/vUgAaAAA
AAAAABnvv71I77+9ae+/ve+/ve+/vTVPyKPvv73vv71PUO+/vT1AAAAN77+977+9AAAwAAAAAAAAAEiTvv73RoTTIkn7vv73vv73vv708Ue+/vTQNAe+/vQOA
Bu+/ve+/vS1LCu+/vT7vv73vv73vv73vv70yPe+/vXbvv73vv73dtV4XcV/vv73vv73vv70f77+977+9MwAsA1fvv73Iqkjvv705ETLvv73vv73vv73eme+/
vS4FMu+/vVdKZe+/ve+/vWN677+977+9fBTvv73vv7OT77+977+9Jlrvv70o77+9SzQzZjnvv73vv73vv70dUhHvv73vv73vv70U77+9XS3vv73vv70EWu+/
vQ/vv73Tqu+/vVd9IdmNT++/ve+/ve+/ve+/ve+/vd2V77+9PmtJCx7vv71F77+9UhDvv71j77+9ch5ESkLvv71kJ++/vULvv71OVDlH77+977+9ZHU377+9
Eyrvv71fW++/vTnvv71Fde+/vSM877+977+9dW19Exbvv71P77+977+9UQjvv71gfu+/vQbvv7Oc77+9BO+/ve+/vRrvv73vv73vv73vv70VPTlZ77+9bu+/
vXlw77+977+9fO+/ve+/ve+/vWOWEe+/vSrvv70mb3ElYRVO77+9eO+/vUO677+9OWHvv73ZuHBl77+9Wyjvv73vv70bF1Pvv73vv73vv70WWwlMSu+/vTBr
ce+/ve+/ve+/vUgbeO+/vVEoce+/vTXvv70dOO+/ve+/vQ5XBy15LXdV77+977+9agZ6Ve+/ve+/vVIk77+977+977+9FBIgC++/ve+/vUVMNu+/ve+/vUt9
bnHvv73vv70s77+977+9NWXvv713G++/ve+/vS4A77+9Qu+/vWfvv71bBu+/ve+/ve+/ve+/vTkpCnrvv73vv714Envvv73vv70SeFp6Gu+/vToOJO+/ve+/
ve+/ve+/vQXvv73vv73doOIQMu+/vUTvv70T77+9Ee+/vWczD1o+Je+/vSjvv73vv71477+9JAJAeTEJJO+/ve+/vQ3vv70kDe+/vRoaaEgEHmwQ77+9Dzbv
v71C77+9Vu+/vURCWe+/vQATEhDvv73vv70IUO+/vXlD77+9fn9EeBHvv73vv73vv70CdTYqGglIJC7vv73vv70DMO+/vSlqO++/vRJmJO+/vQXvv71D77+9
77+977+9C++/vT94X++/ve+/ve+/ve+/vSjvv73vv71iJe+/ve+/vS5677+9Au+/vRggOu+/vQ1IXiXvv73vv71977+977+977+9fe+/vVVXfu+/ve+/vWdj
JO+/vQYsP++/vRDvv71I77+9wo/vv73IgQNcSO+/vR3vv70L77+977+9au+/vUnvv71jF++/vQI6UHQM77+9I++/ve+/ve+/ve+/vWvvv71TVz3vv73v
v73vv73vv70X77+9EADvv71i77+977+92oJgPsOYaO+/vQzvv71m77+9d++/vR7WoSlg77+9cTLvv73vv71S77+9O++/ve+/ve+/vW3vv70+Ee+/vWhD
77+9CQhiZCvvv71pRO+/ve+/ve+/ve+/vS7vv71ZVe+/vRPvv714I2/vv7O1Xe+/vWjvv71YQnMm77+9Ut6TBu+/ve+/ve+/vQlo77+977+977+9wFnvv715
77+977+977+977+977+9ZTwv77+9P++/ve+/vSrvv70V77+977+977+977+977+9eO+/vTLvv73vv71Re++/ve+/ve+/ve+/ve+/vXkEIu+/vV7vv70W77+9
LO+/ve+/vRjvv71a77+9De+/vcyF77+977+9Au+/vXAsKO+/vfOOs7jvv70z77+9M++/vS7vv70A77+9YVzvv70IGODvv73vv73vv73vv73vv707Re+/
vSVK77+9Z2F2Zu+/ve+/vQZBHUcOVVTvv70C77+9Xe+/ve+/ve+/vSbvv71c77+9UkIOce+/vU3vv70377+9JO+/vdyBOO+/vUDvv717Me+/ve+/vR1z656B
aVpmYe+/vVHvv71C77+977+9Yu+/ve+/vWDvv73vv70577+9OV/vv71e77+9TV8Ezr4g77+9OO+/ve+/vTJM77+9ZO+/vXMD77+977+9cu+/ve+/ve+/ve+/
ve+/ve+/ve+/vSZWWu+/vSJ777+977+977+977+9Oe+/vTbvv717Du+/vRgiTO+/vRQpMu+/ve+/ve+/vTPvv73Xne+/vWoFSwl177+977+9I3bvv70S77+9
EAtw77+93qfvv73vv71o77+9LjYxJmnvv70DUO+/vSfvv71iLjga1owh77+9fu+/ve+/vVjEtO+/vdqqYe+/ve+/ve+/vSnvv73vv73vv71mSTvvv71OBWDv
V73vv73vv70kUEDvv71q77+977+9FHNo77+9L2xr77+977+9aEvvv73xl7KWVmnfnyBZfBO4Ve+/vXHGqxzvv703bd2FLu+/ve+/ve+/vSzvv73vv71s77+9
MO+/vU3vv73vv70NKO+/vTvvv73vv70wcw4M77+9U++/ve+/ve+/ve+/ve+/ve+/ve+/vQbvv73vv73vv70V77+9EO+/vT3vv73vv71k77+9QBtnCe+/
vUIQ77+9L++/vWtFyJrvv73vv70V77+9G++/ve+/vQOX77+977+977+977+9KRzvv73NtSJN77+977+977+9Ee+/ve+/ve+/ve+/ve+/vW5Za3Lvv71s77+9
77+9Yu+/ve+/vXbvv73vv70577+9Be+/ve+/ve+/ve+/veW8jmrvv70477+9Ae+/ve+/ve+/vRvvv707amfvv709GEcC77+977+9ee+/vWjvv70L77+977+9
TO+/ve+/vVMiNGsb77+9dmIhc37vv73vv73vv73vv71uEDrvv70N77+977+9Ql3vv73SlNOEWu+/vQLvv73vv71P77+9Vu+/ve+/ve+/vW8MFRbvv70hGDkm
CALvv70Y77+977+9Tu+/vUEIYi4hKyjvv73vv73vv70MMFTvv70BeEBJKnbvv73vv71rJu+/ve+/ve+/vSgrLEguMO+/vSTvv73Equ+/vQww77+9MO+/vVYY
XBgK77+9UidaQO+/ve+/vVzvv71jGxbZo1bvv719aAk377+9SdycUm8NOe+/vT1I77+90b4KY++/vSYcDO+/vVLvv73vv70477+977+9DDPv
v71dYDRY77+9GO+/vQLvv71q77+977+977+977+9OBw4FO+/ve+/ve+/vTxw77+9Qu+/vURG77+977+9LUzvv73vv73vv73vv73vv70f77+977+9c++/
ve+/vV9/77+9clfvv70fQQQu77+9Tt6z77+9VO+/ve+/ve+/vQ7vv71w77+977+9GydWcg5DD++/vTPvv73vv73vv70Da2VE77+9FDMC77+977+9Yu+/vXU4
Ie+/vQTvv71177+9cO+/vUXvv73vv70QRnHvv73vv70r77+9Hu+/vXVhVe+/ve+/vXjvv71777+977+977+9HAgQNRAgThjvv73vv70T77+9Xu+/ve+/vTfv
v70X77+9Ze+/ve+/vREJ77+9Me+/vWZZwqYlEe+/vWTvv71xe++/ve+/vWVO2Zmnvv709Ku+/ve+/vRQX77+977+9Yu+/vRhhJO+/ve+/vXLvv73vv73vv70T
Nu+/vSPvv70a77+9VSlZGzBnDinvv70cKO+/ve+/ve+/ve+/ve+/vX9J77+93zYC77+9GDLvv71he++/vXMVTl3Guu+/vQbvv70U77+9YO+/ve+/ve+/vVIV
dDIt77+977+977+9dhQRA++/vTXvv73vv73vv73vv73ckU4UJA1m77+977+977+9
```


3. Порівняння текстових файлів та їх версій, закодованих у base64:



	Text1.txt / text1.enc	Text2.txt / text2.enc	Text3.txt / text3.enc
Base64	290	2806	2097
Оригінал	195	2312	1662



Можна помітити, що розмір закодованих файлах зріс. Це пов'язано з алгоритмом кодування base64 – перетворення, наприклад 3 октетів (по 8 біт) у 4 секстети (по 6 біт), що збільшує розмір на третину.

Найефективнішим алгоритмом стиснення виявився bzip2.

Висновок

У ході виконання лабораторної роботи ознайомився з поняттям ентропії інформації та пов’язаних з ентропією понять. Порівняв алгоритми стиснення – обрав кращий з них для випадків, коли треба буде зекономити місце на носії. Закріпив теоретичні знання практикою. Теоретично та практично ознайомився з алгоритмом кодування Base64, його перевагами та недоліками.

Код програм, звіт та текстові файли, використані у роботі містяться у репозиторії за посиланням: (тут)