**Київський національний університет імені Тараса Шевченка**

**радіофізичний факультет**

Лабораторна робота № 1

**Тема:** “Дослідження кількості інформації при різних варіантах кодування”

Роботу виконав

студент 3 курсу

Комп’ютерної інженерії

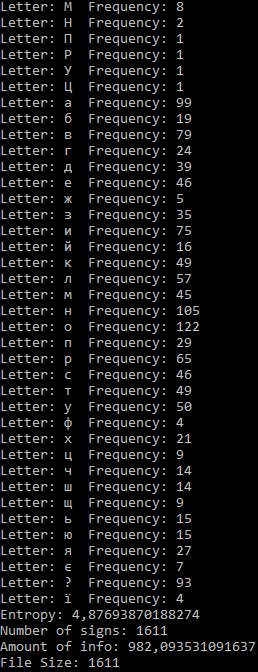
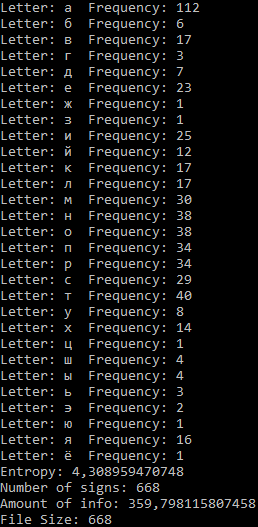
Мамотенко Сергій

Київ 2019

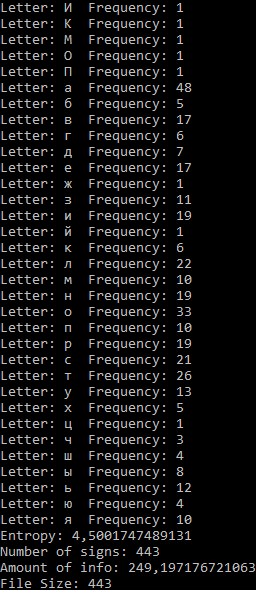
Хід роботи

1. Було обрано наступні 3 тексти:

* Технічні особливості двигуна УЗАМ-412;
* Кино – Мама Анархия
* Даниил Хармс – Вываливающиеся старухи



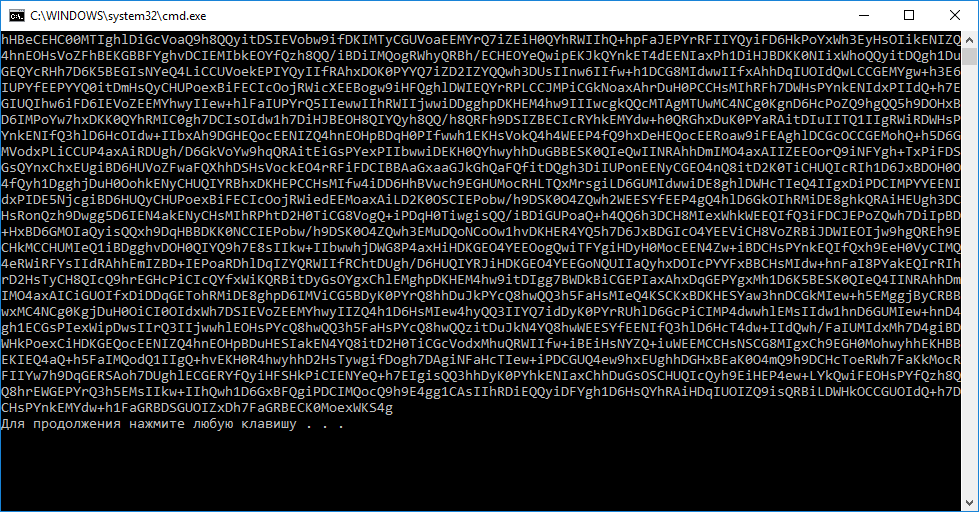
Перший текст Другий текст Третій текст



Можна бачити, що ентропія першого тексту(про технічні особливості двигуна) є найбільшою. Це пояснюється більшим об’ємом алфавіту.

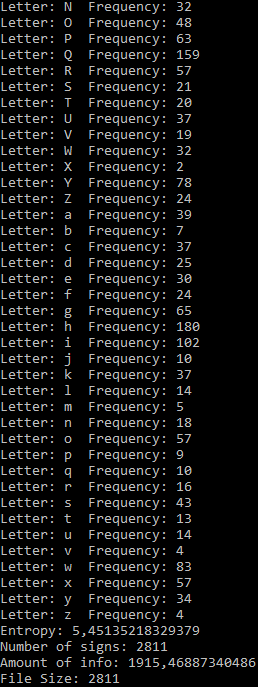
|  |  |  |  |
| --- | --- | --- | --- |
|  | Перший текст | Другий текст | Третій текст |
| Ентропія | 4.88 | 4.3 | 4.5 |
| Кількість знаків | 1611 | 668 | 443 |
| Кількість інформації | 982.1 | 359.8 | 249.2 |
| Розмір файлу | 1611 | 668 | 443 |
| RAR | 1036 | 326 | 331 |
| 7Z | 1146 | 384 | 376 |
| GZ | 956 | 259 | 261 |
| TAR | 3584 | 2560 | 2048 |

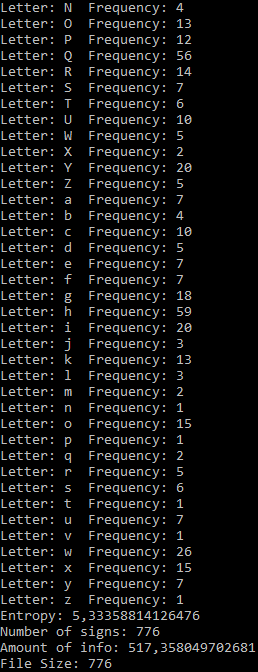
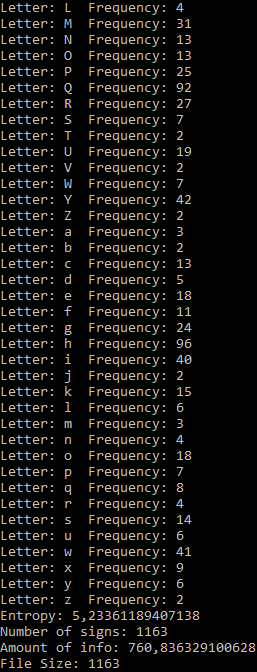
Найкраще компресує GZ.

1. 

Закодовані в BASE64 тексти:

Перший текст Другий текст Третій текст





|  |  |  |  |
| --- | --- | --- | --- |
|  | Перший текст | Другий текст | Третій текст |
| Ентропія | 5,45 | 5,34 | 5,23 |
| Кількість знаків | 2811 | 776 | 1663 |
| Кількість інформації | 1915 | 517 | 760 |
| Розмір файлу | 2811 | 776 | 1163 |
| GZ | 2811 | 1163 | 776 |

Було обрано архіватор GZ, як найбільш ефективний.

Внаслідок проведених операцій збільшилось число ентропії. Також збільшилась

кількість символів.

При цьому розмір файлу порівняно з минулими архівами збільшився приблизно в 1.5 рази.

**Посилання на Git репозиторій**

https://github.com/Serzhulka/lab1\_cs

**Висновки**

Було досліджено імовірнісні параметри української та російської мови для оцінки кількості інформації текстів. Досліджено вплив різних методів кодування інформації на її кількість.