

Лабораторна робота №1

Дослідження кількості інформації при різних варіантах кодування

Виконав студент

3 курсу СА-КІ

Герасименко Артем

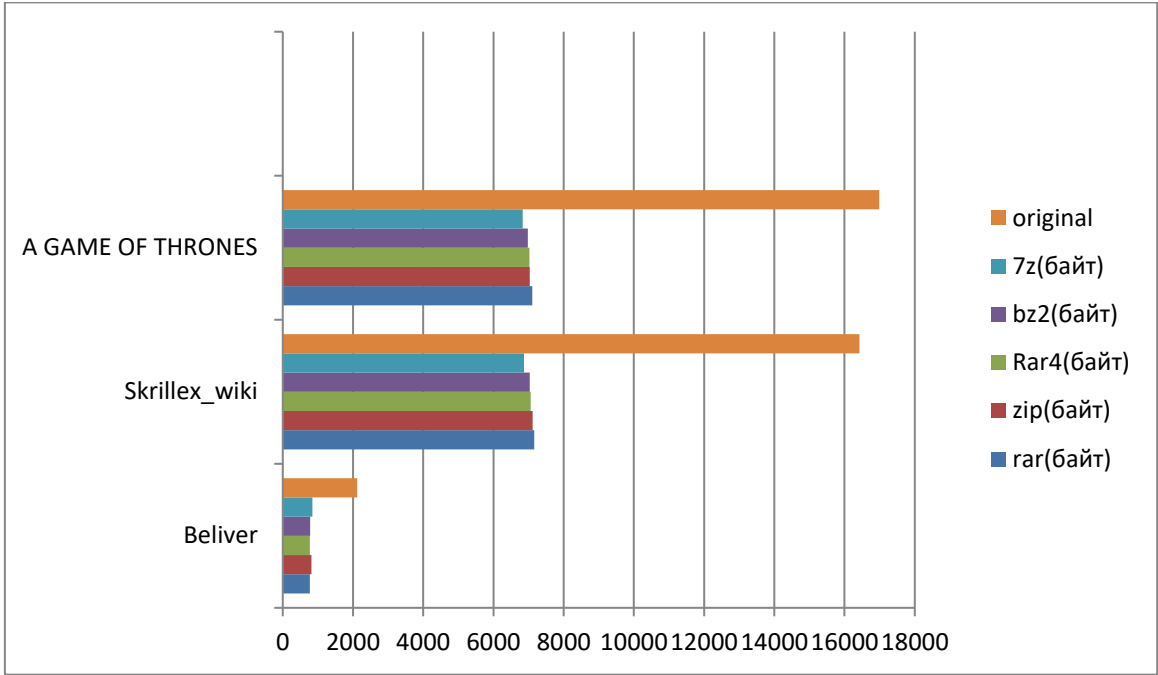
## 1. Дослідження кількості інформації в тексті

Оберіть 3 текстових файла різного тематичного та лінгвістичного спрямування (наприклад, вірш Тараса Шевченка “Мені тринадцятий минало”, “Казка про репку” Леся Подерв’янського та специфікацію інтерфейсу PCI)

Я обрав:

- 1)Текст пісні Beliver
- 2)Інформацію про Skrillex з сайту wikipedia
- 3)1 розділ з книги Гра престолів



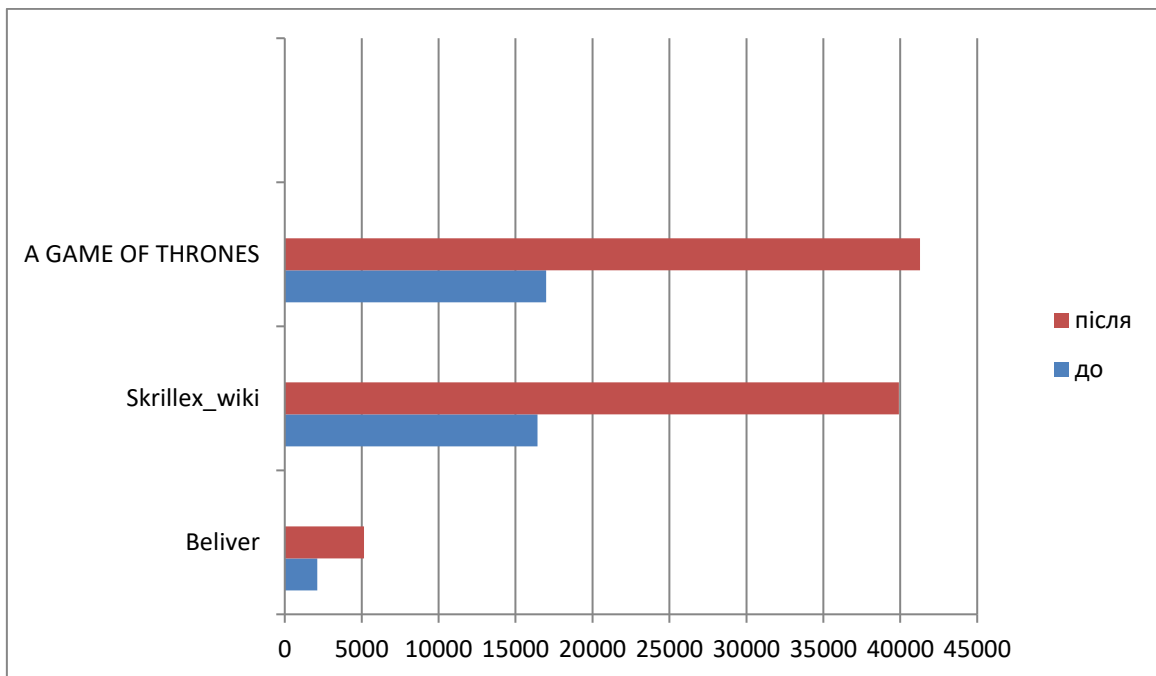


Дослідження способів кодування інформації на прикладі Base64

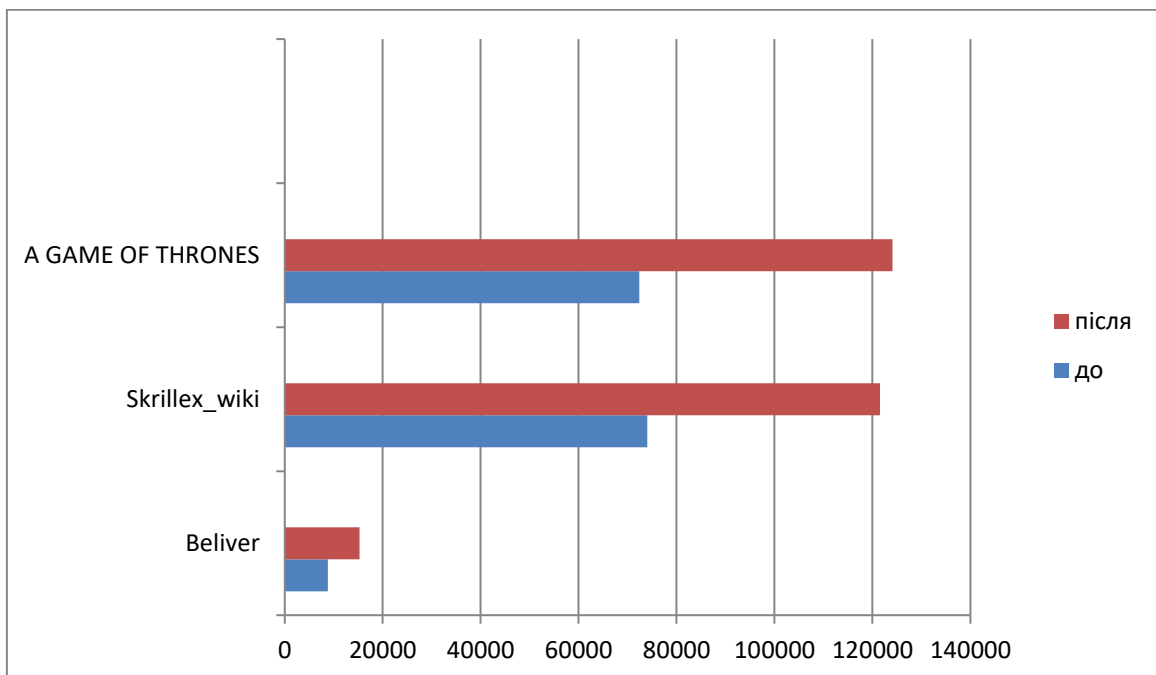
1) Аналіз закодованих файлів

| Назва             | Кількість символів | Ентропія | Кількість інформації | Розмір файлу(байт) |
|-------------------|--------------------|----------|----------------------|--------------------|
| Beliver           | 2836               | 5.378    | 15252                | 5,159              |
| Skrillex_wiki     | 21952              | 5.535    | 121526               | 39906              |
| A GAME OF THRONES | 22800              | 5.442    | 124082               | 41293              |

Розмір файлів до і після кодування:



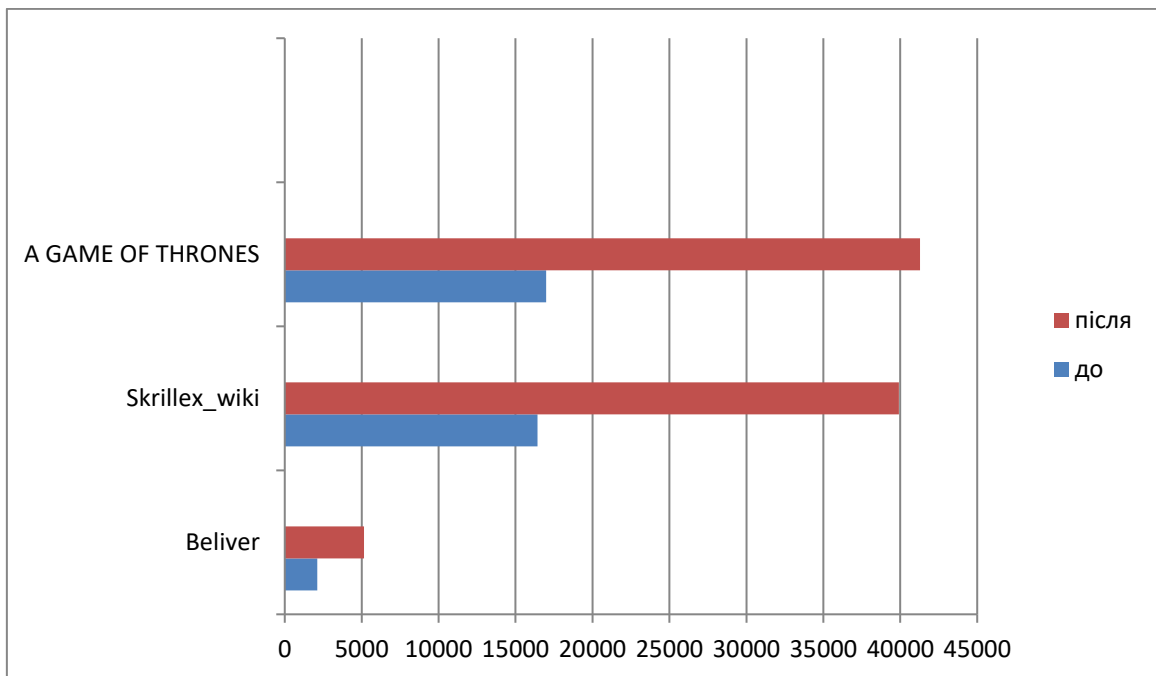
Кількість інформації у файлах до і після кодування:



## 2) Порівняння файлів після стиснення

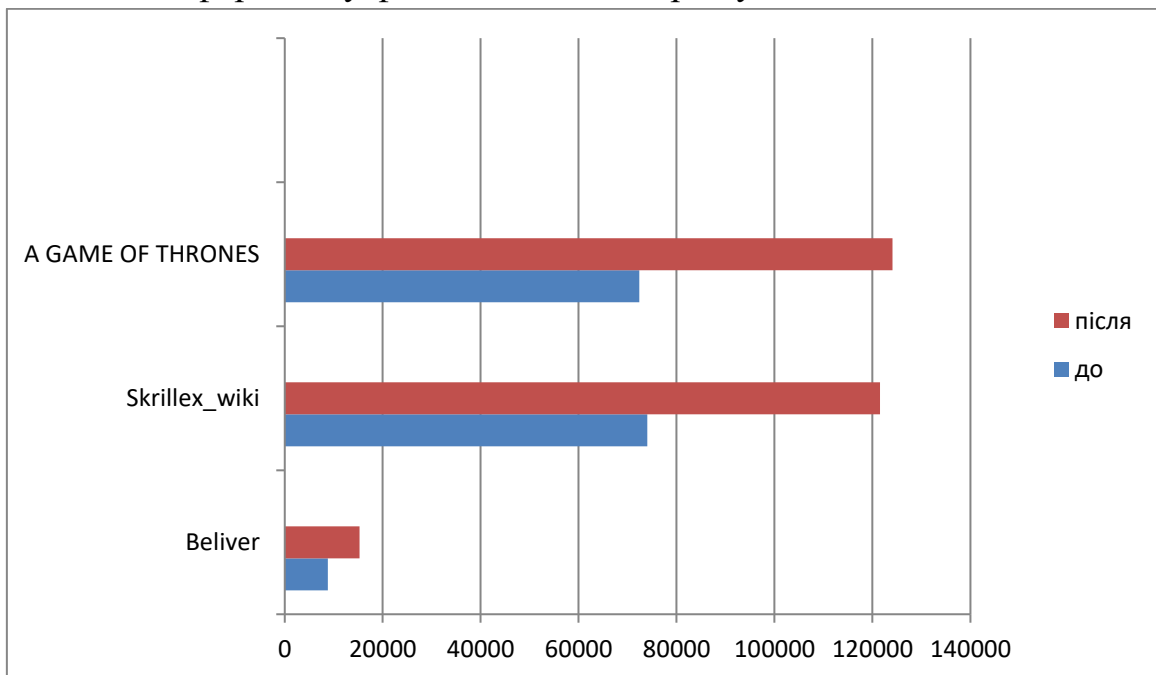
Найкращим інструментом для стиснення файлів, як можна побачити на першій діаграмі, виявився rar.

Розмір файлів до і після архівування:



Після кодування розмір файлів зростає приблизно у 1.5 рази

Кількість інформації у файлах до і після архівування:



Після архівування кількість інформації зростає приблизно у 1/3 раз.

Висновок: Під час лабораторної роботи я дослідив імовірнісні параметри української мови для оцінки кількості інформації текстів та вплив різних методів кодування інформації на її кількість.