**МИНОБРНАУКИ РОССИИ**

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**

**ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**

**«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**

**Кафедра САПР.**

**ОТЧЕТ**

**по лабораторной работе №3**

**по дисциплине «Алгоритмы и структуры данных»**

**Тема: «Алгоритмы сжатия без потерь».**

| Студентка гр. 1301 |  | Ищенко Д.О. |
| --- | --- | --- |
| Преподаватель |  | Родионова Е. А. |

Санкт-Петербург

2023

Оглавление

[Формулировка задания. 3](#_gjdgxs)

[Описание методов и оценка временной сложности. 3](#_30j0zll)

[Пример работы программы. 3](#_3znysh7)

[Текст программы. 3](#_2et92p0)

# Формулировка задания.

Реализовать следующие алгоритмы сжатия символьных данных:

1. Алгоритм Хаффмана (HA)
2. Кодирование длин серий (RKE)
3. Алгоритм Лемпеля-Зива (LZ78)
4. *Преобразование Барроуза-Уиллера (BWT)*
5. *MTF*
6. Арифметическое кодирование (AC)
7. PPM

Проанализировать эффективность и время сжатия на датасете enwik8 для процедур сжатия на основе:

1. Алгоритма Хаффмана
2. Арифметического кодирования
3. LZ78
4. BWT -> MTF -> HA
5. BWT -> MTF -> AC
6. RLE -> BWT -> MTF -> RLE -> HA
7. RLE -> BWT -> MTF -> RLE -> AC
8. PPM

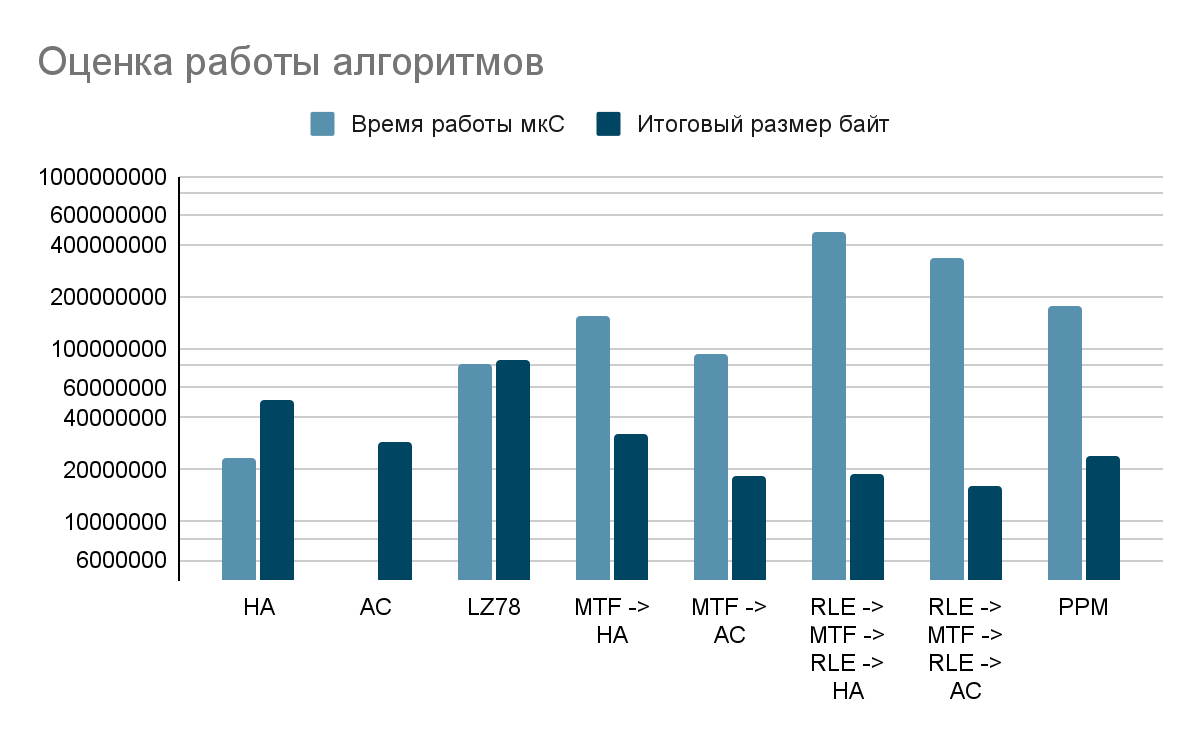
Результаты представить в виде или таблиц. При слишком больших временных затратах на сжатие взять первые 10% enwiki8.

# Описание методов и оценка временной сложности.

Поскольку BWT требует слишком много памяти (10Гб на 0.1% датасета) результаты представлены без него на 100% датасета. На 0.1% алгоритм отрабатывает за микросекунд.

Исходный размер файла: 100,000,000 bytes

| Название алгоритма | Время работы мкС | Итоговый размер байт |
| --- | --- | --- |
| HA | 23,157,264 | 51195637 |
| AC | 4,557,371 | 29152844 |
| LZ78 | 82,431,160 | 86289003 |
| MTF -> HA | 154,782,848 | 31997273 |
| MTF -> AC | 93,214,587 | 18220527 |
| RLE -> MTF -> RLE -> HA | 481,111,354 | 18961347 |
| RLE -> MTF -> RLE -> AC | 335,063,176 | 15998636 |
| PPM | 179,718,984 | 24372304 |



# Текст программы.

[Ссылка на github](https://github.com/DamirAhm/AADS/master/tree/2_lab3)