**ИНОБРНАУКИ 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 |  | Герасимов А. М. |
| Преподаватель |  | Родионова Е. А. |

Санкт-Петербург

2023

Оглавление

[Формулировка задания. 3](#_Toc32031)

[Описание методов и оценка временной сложности. 3](#_Toc8031)

[Пример работы программы. 3](#_Toc21061)

[Текст программы. 3](#_Toc16533)

# 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

1. HA time: 22,106,564 ; 510,896,475 bytes
2. AC time: 4,535,171 ; 8 bytes
3. LZ78 time: 83,632,069 ; 186,289,003 bytes
4. MTF time: 85,531,969

+ HA time: 69,250,879

Total: 154,782,848

Size: 736,934,975 bytes

1. MFT time: 85,810,845

+ AC time: 11299642

Total: 97,110,487

Size: 8 bytes

1. RLE time: 26,610,154

+ MFT time: 146,965,303

+ RLE time: 122,760,424

+ HA time: 184,775,473

Total: 481,111,354

Size: 1,581,475,260 bytes

1. RLE time: 27,294,209

+ MFT time: 153,475,788

+RLE time: 117,817,025

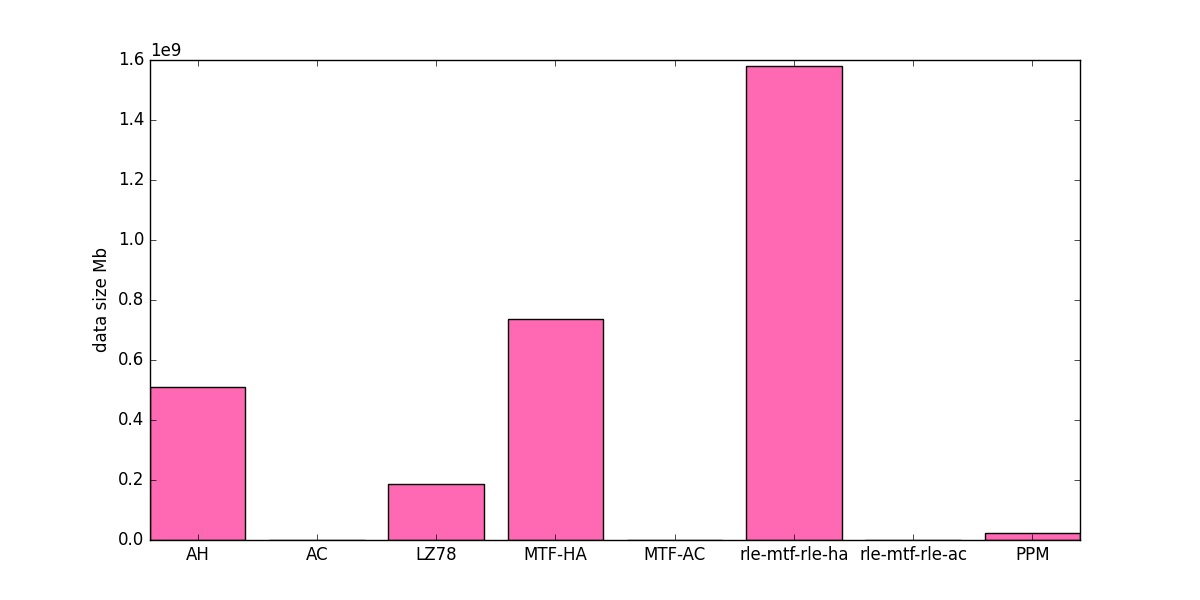
+AC time: 36,476,154

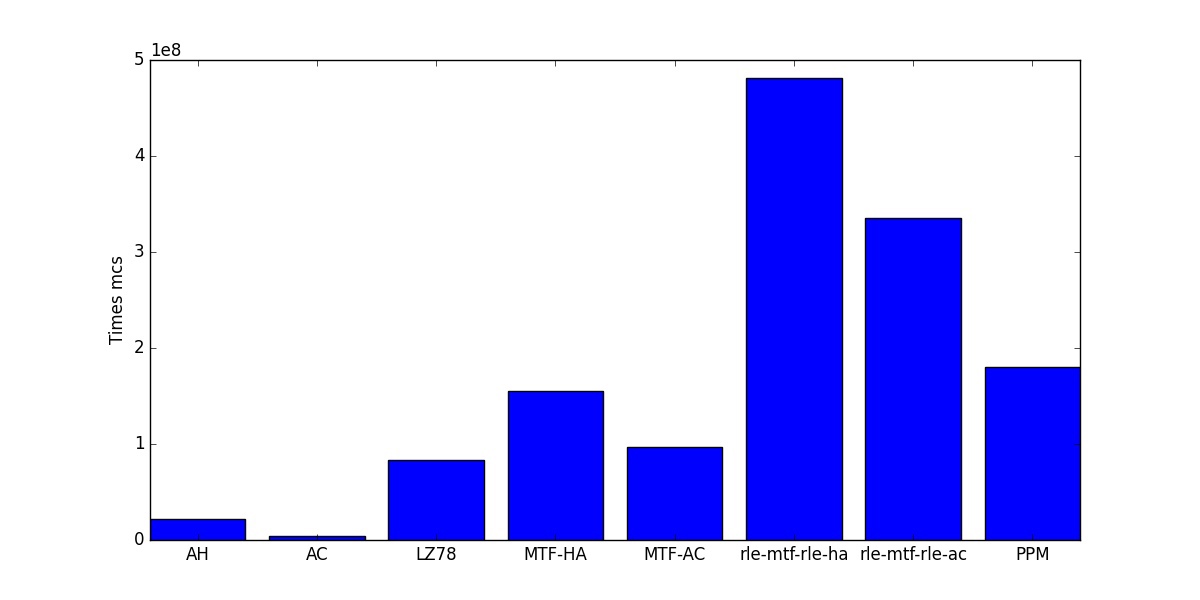
Total: 335,063,176

Size: 8 bytes

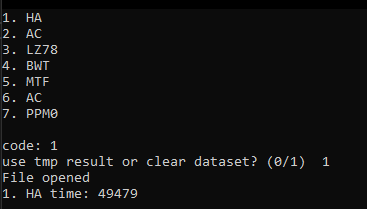
1. ppm time: 179,718,984

Ppm size: 24372304 bytes





# Пример работы программы.



# Текст программы.

[Ссылка на github](https://github.com/ArtGer2/Frogram/tree/main/Sem2_lab3)