

Кузнецов Кирилл Олегович

Отчёт по курсовой работе

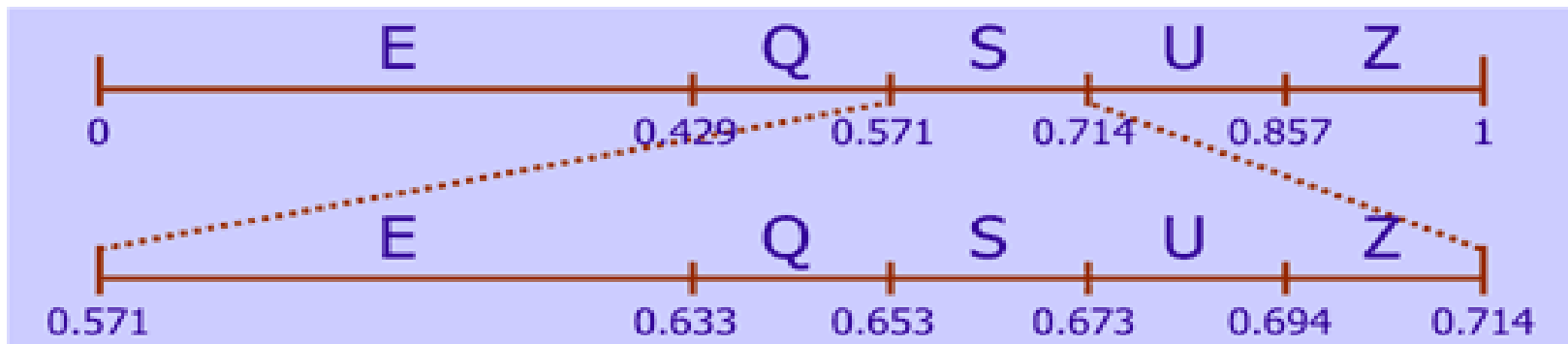
Реализация арифметического
кодирования

Задачи работы

- Реализация архиватора на основе алгоритма арифметического кодирования
- Оценка степени сжатия
- Сравнение с распространёнными архиваторами (RAR, ZIP, 7-zip - LZMA)

Описание алгоритма арифметического кодирования

- Кодируемому символу в зависимости от его вероятностной характеристикой ставиться в соответствие интервал



- На каждом шаге интервал сужается
- Код файла — число из итогового промежутка

Особенности реализации

- В качестве вероятностных характеристик символов используются их действительные частоты в кодируемом файле.
- Файлы кодируются и декодируются постепенно, то есть каждый раз записывается общая часть двоичного представления границ текущего интервала.

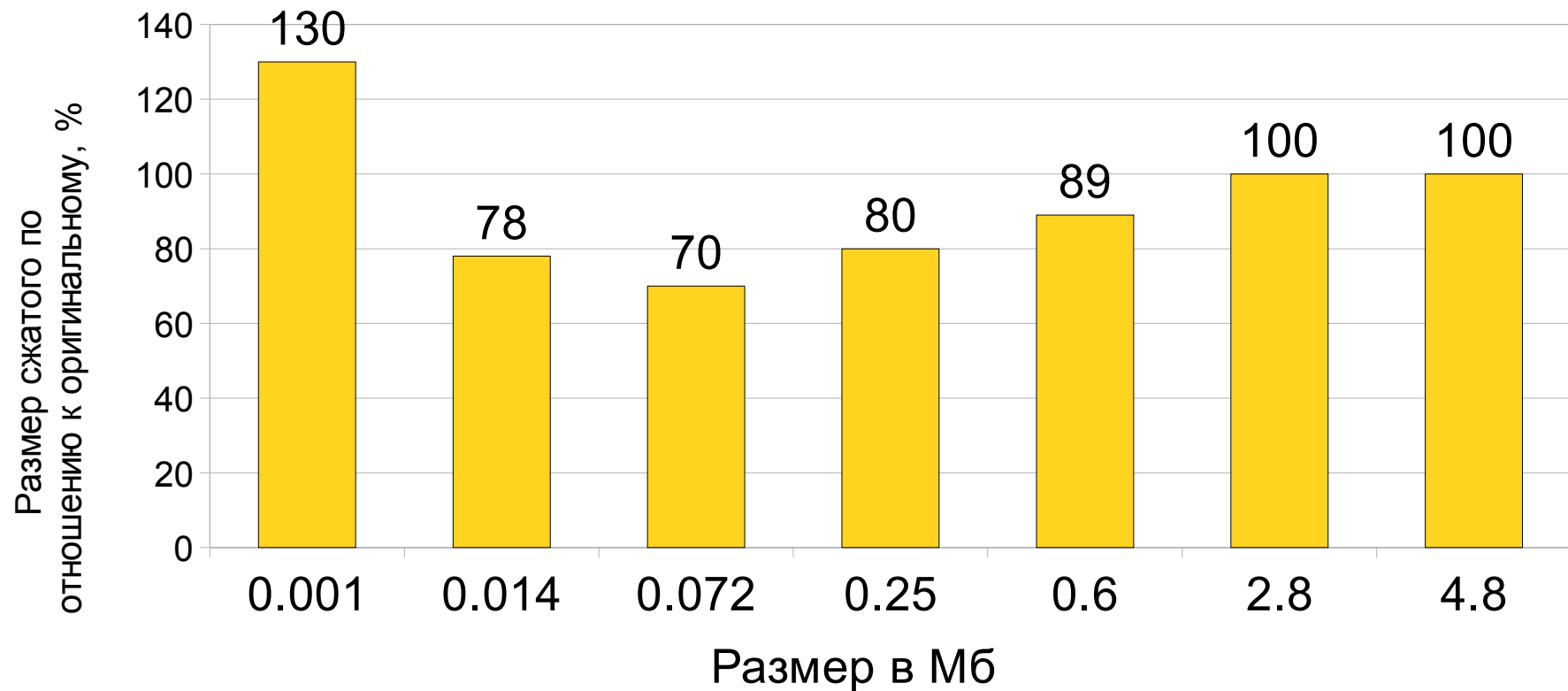
Пример: текущий интервал [13,14]

$$13 = {}_2 1101, 14 = {}_2 1110$$

можем записать 11

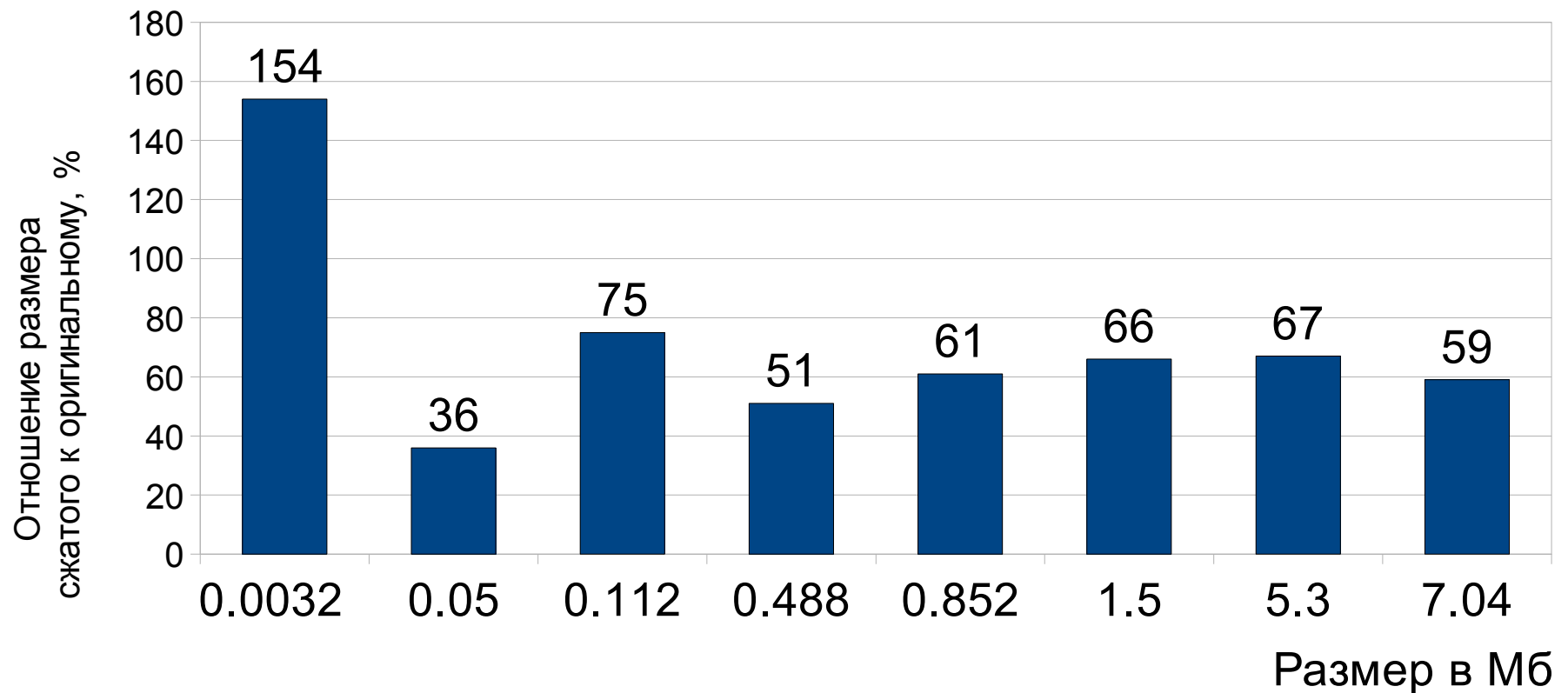
- В качестве границ интервалов используются целочисленные значения.

Оценка степени сжатия (исполняемые файлы)



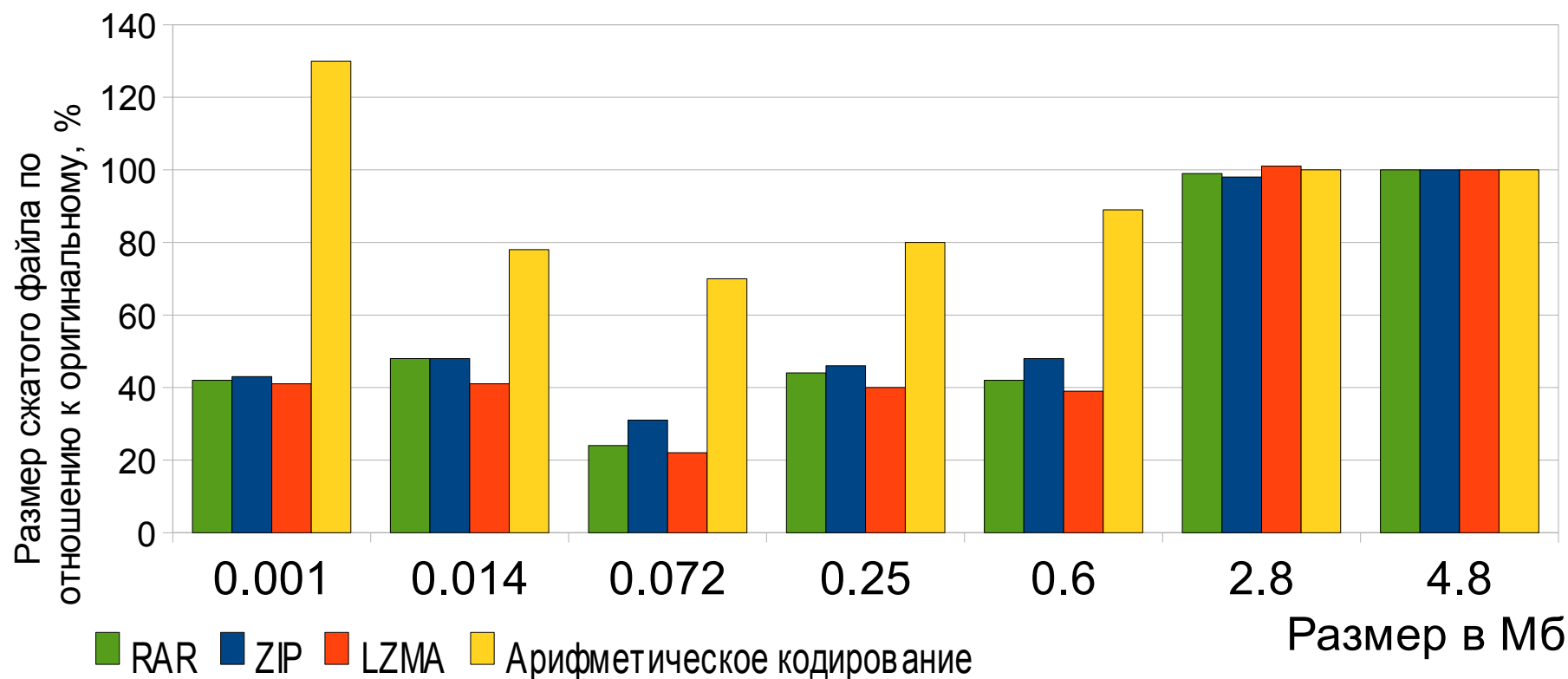
Размер сжатого файла 70%-100 % сжимаемого
(для файлов > 10 Кб)

Оценка сжатия (txt файлы)



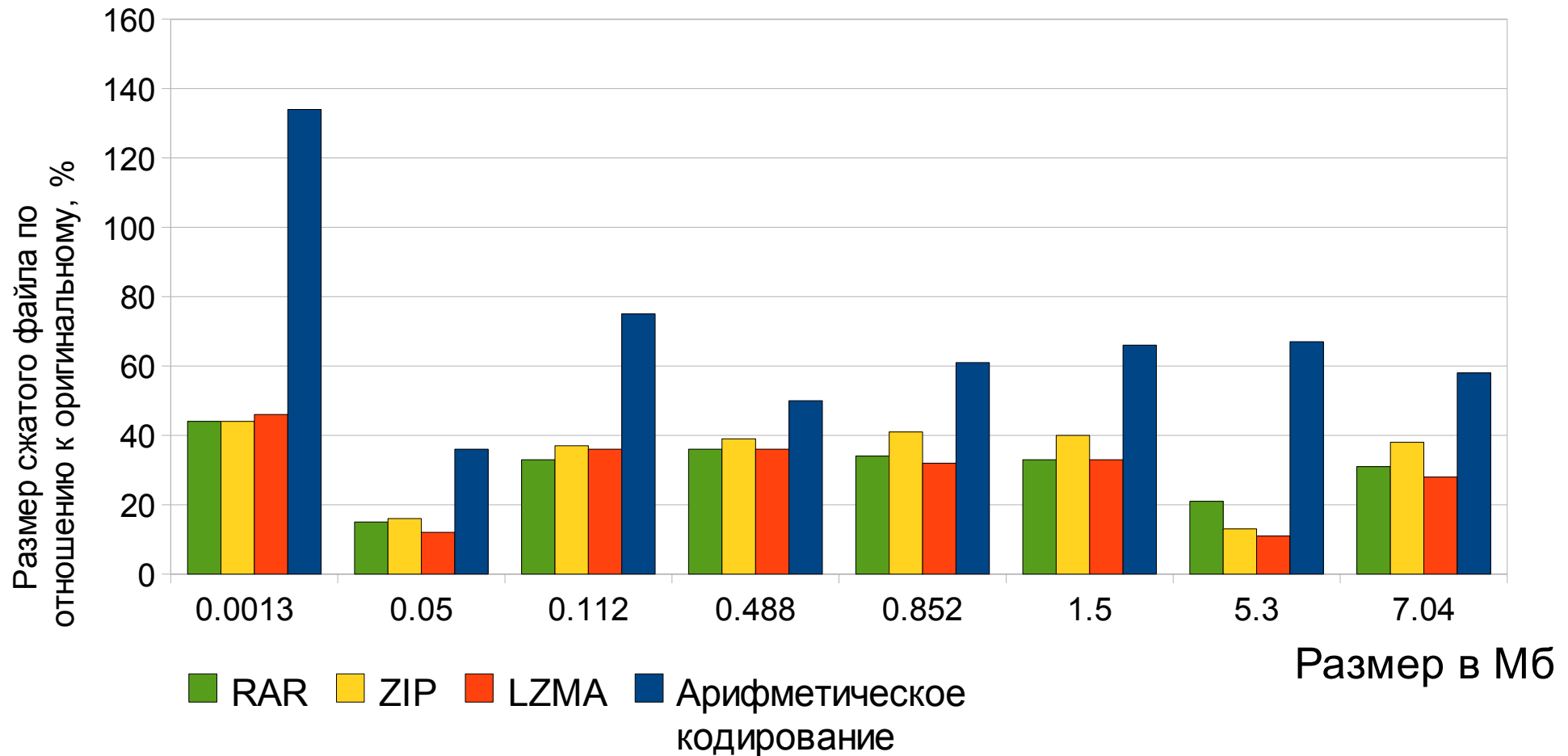
Размер сжатого 36% - 75% сжимаемого
(на файлах >10 Кб)

Сравнение с распространёнными алгоритмами (исполняемые файлы)



Файлы, сжатые реализованным архиватором,
в 1.5 - 3 раза больше сжатых конкурентами
(За исключением несжимаемых файлов)

Сравнение с распространёнными алгоритмами (txt файлы)



В этом случае ситуация практически не меняется.

Заключение

- Для маленьких(< 10 Кб) файлов размер закодированного файла превосходит размер итогового.
- На файлах определённого типа, например текстовых, сжатие выше чем в среднем, за счёт распределения символов в таких файлах, т.е. одни символы встречаются гораздо чаще, чем другие.
- Реализованный архиватор существенно проигрывает распространённым аналогам, объединяющими в себе различные методы сжатия.