



Algorithm LZW

tiny research



Idea

Основная идея

Заменять отдельные строки специальным кодом. Код занимает меньше места чем строка. -> Сжатие.

listen to your heart when he's calling for you
listen to your heart, there's nothing else you can do
i don't know where you're going and **i don't know** why
but **listen to your heart** before you tell him goodbye.

#33 when he's calling for you
#33, there's nothing else you can do
#34 where you're going and **#34** why
But **#33** before you tell him goodbye.

Словарь.

#33 - listen to your heart

#34 - i don't know

Problem

Основная проблема

- Нужно хранить отдельно словарь, и передавать его вместе со сжатым текстом.
- есть и другие проблемы

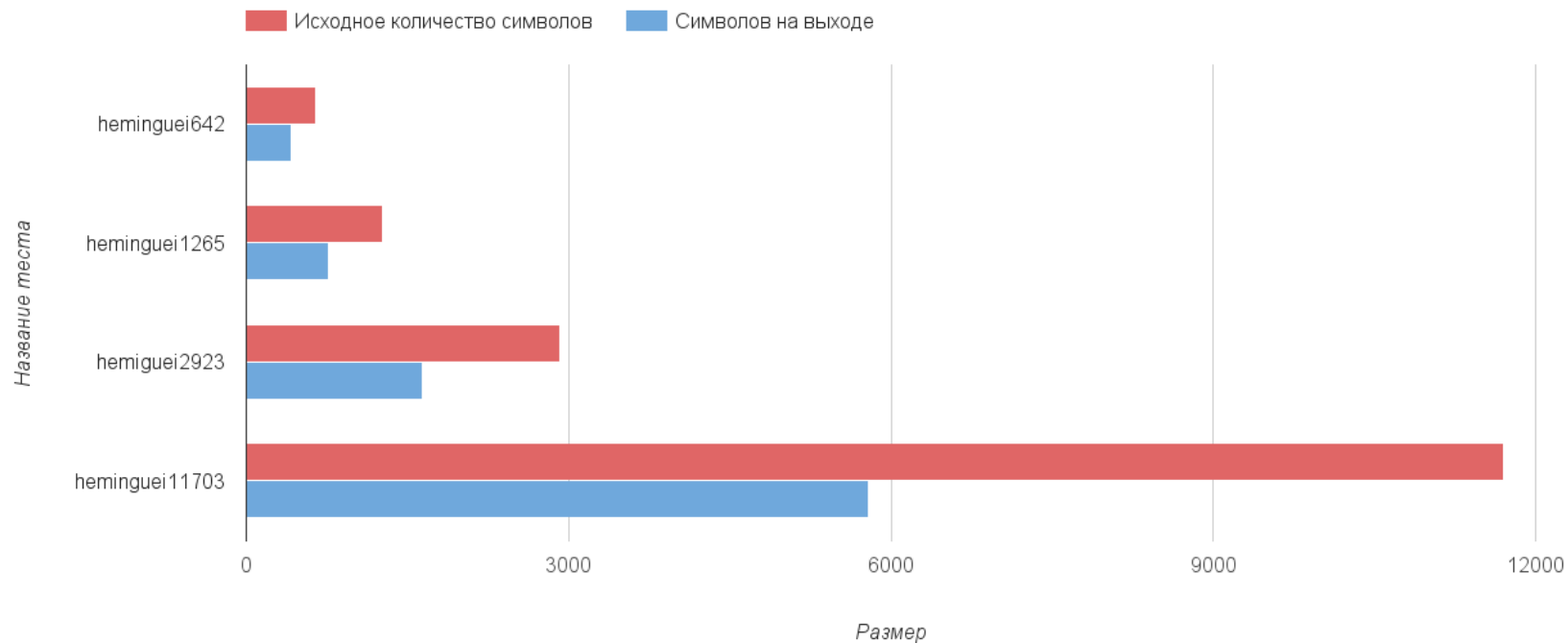
Решение

- сделать так, чтобы словарь был не нужен. Словарь однозначно восстанавливается по сжатому тексту.

How does it work?

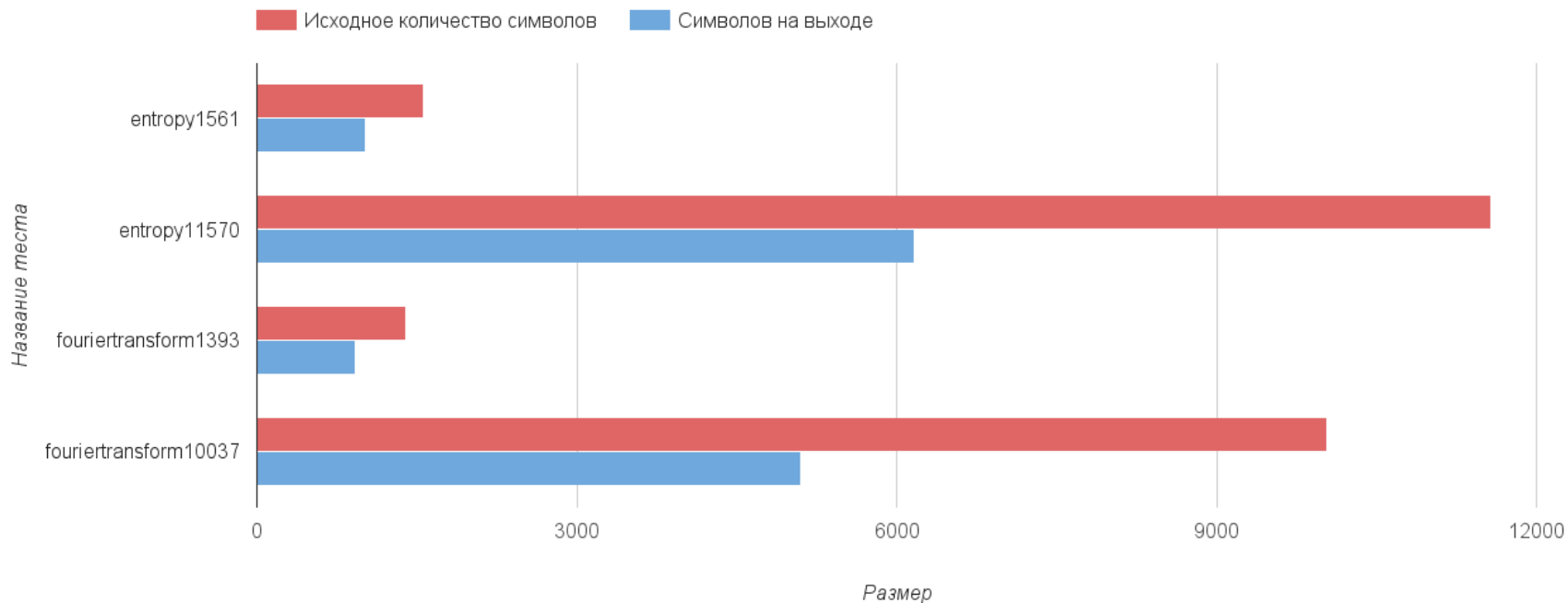
Research 1

Осмысленные тексты



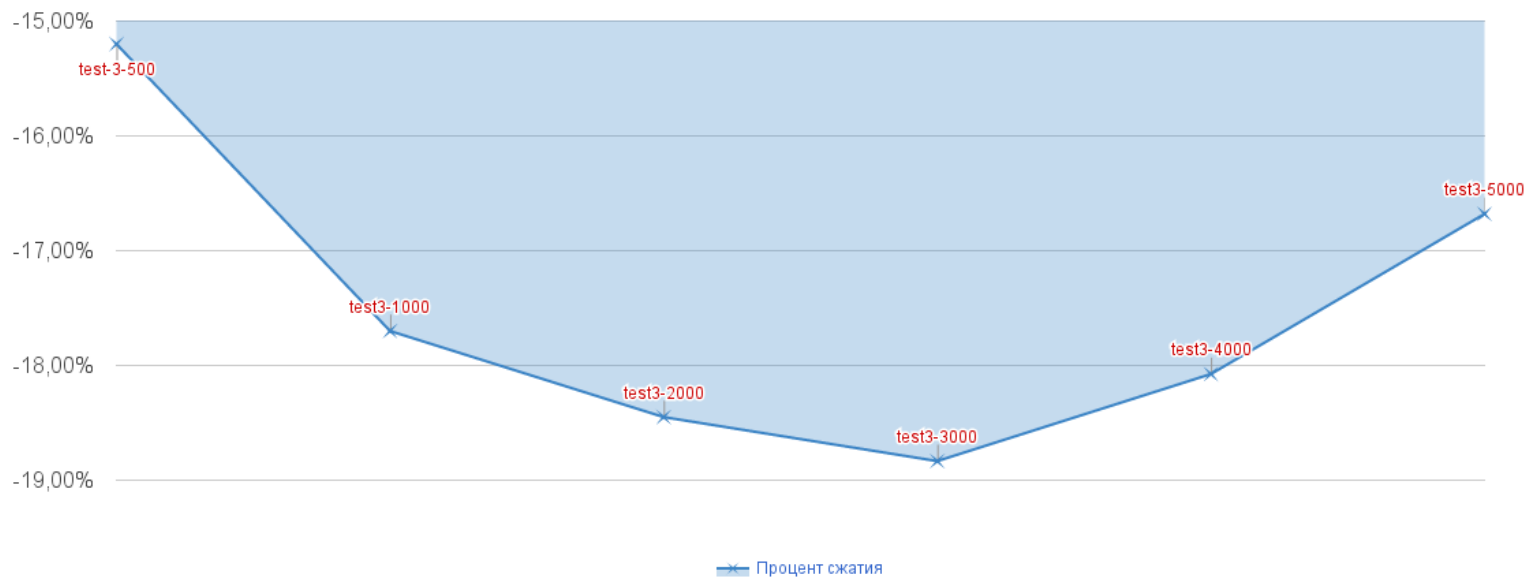
Research 2

Математические тексты



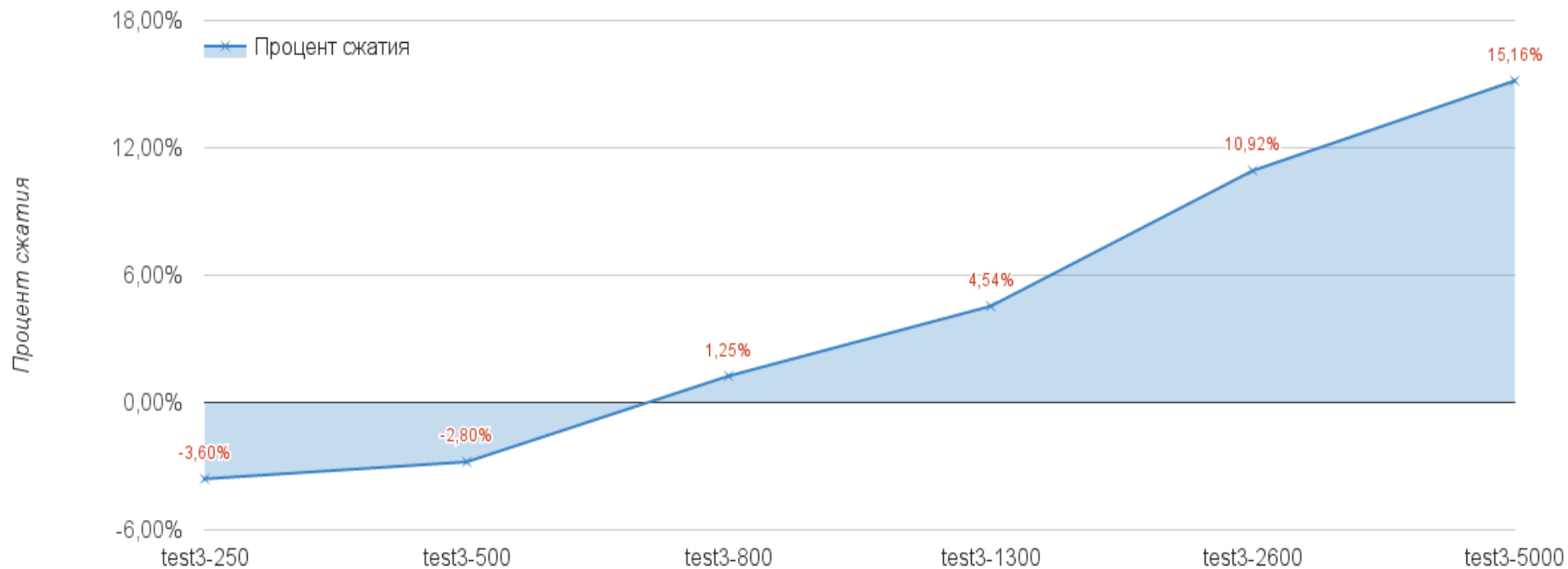
Research 3

Случайные строки (алфавит 89 символов)



Research 4

Случайные строки (алфавит 36 символов)



Summary

- Отлично подходит для сжатия осмысленных текстов любых данных, где есть часто встречаемые строки.
- Плохо работает на случайных данных.

Links

<https://goo.gl/TIk3Tq> - таблицы с данными для диаграмм

http://pm-pu.ru/stuff/codes/date_LZW.zip - исходные тексты + скрипт генерации случайных строк

<http://kobyshevsergey.pythonanywhere.com/lzw/> - реализация алгоритма, на которой проводилось тестирование