

РАЗРАБОТКА БИБЛИОТЕКИ АППРОКСИМАТИВНОГО ПОИСКА

Дмитрий Шеставин

Руководитель: Владислав Белогрудов

Цели работы

Изучение существующих алгоритмов
аппроксимативного поиска

Реализация базовых алгоритмов на
языке Python

Разработка библиотеки на языке C++ с
реализацией нескольких алгоритмов

Сравнение различных алгоритмов

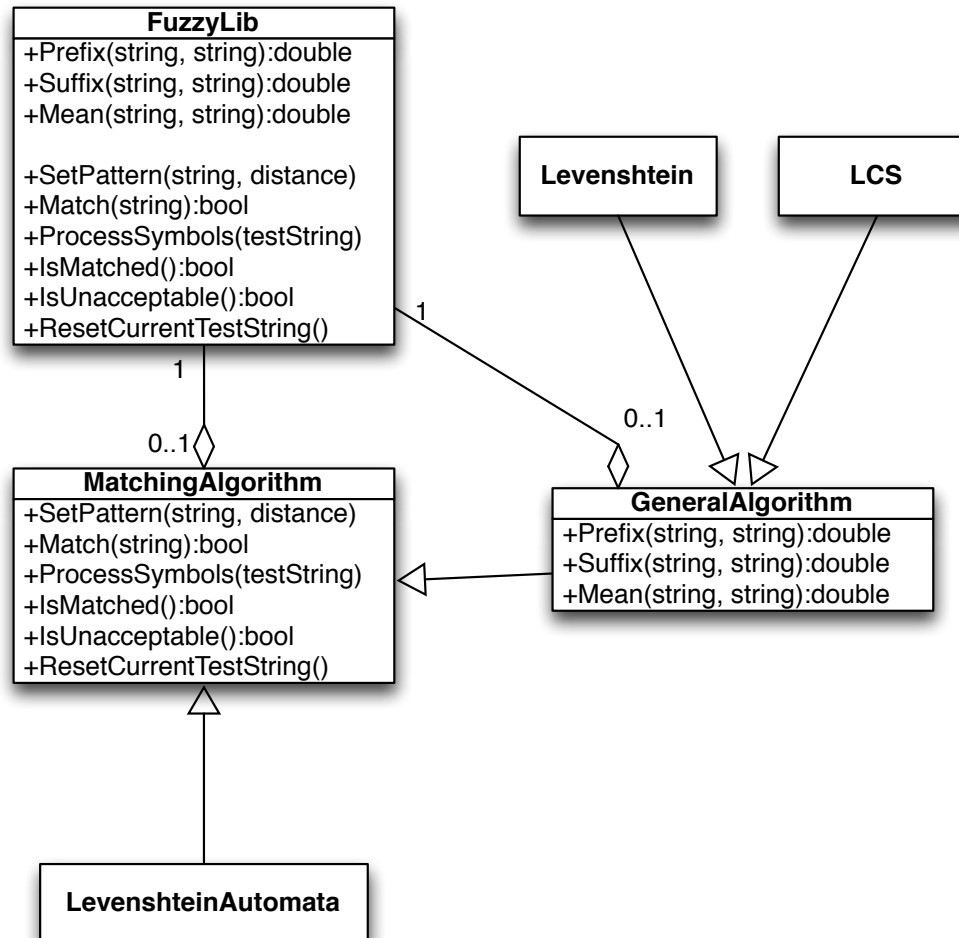
Решаемые задачи

Longest Common Subsequence

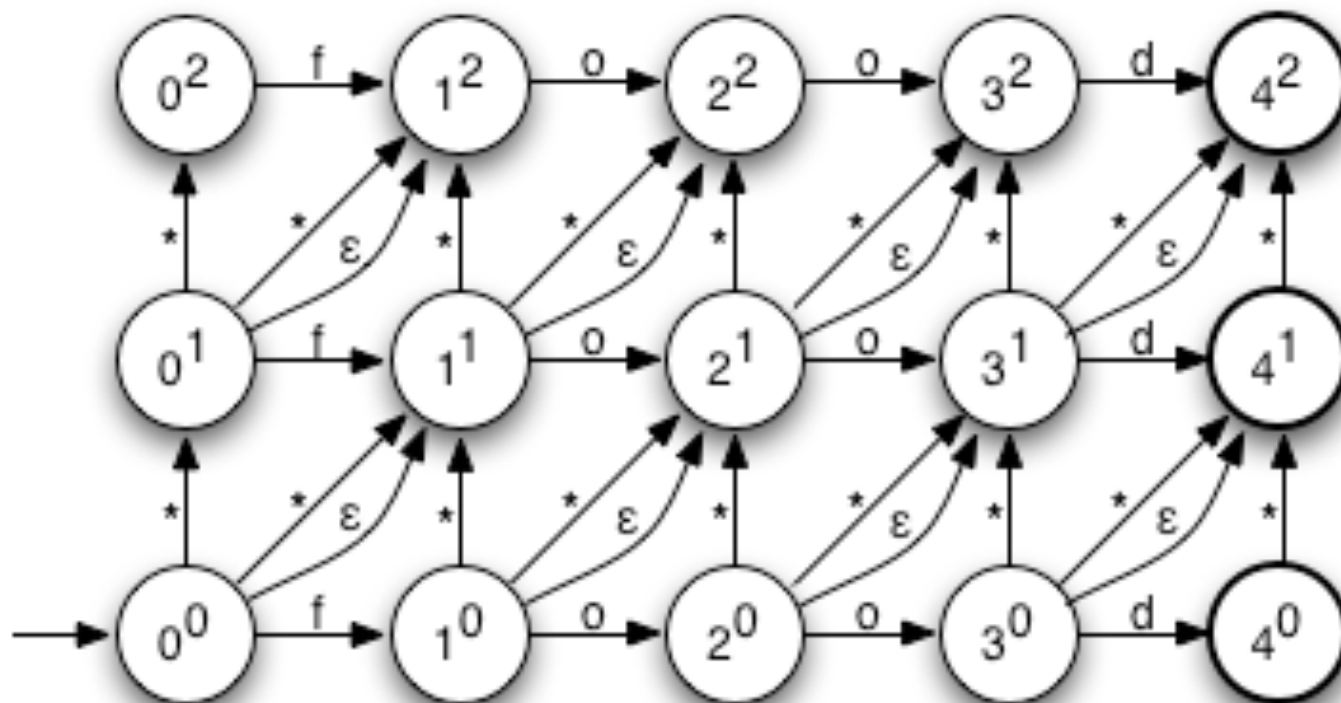
Levenshtein distance

- Dynamic programming
- Levenshtein automata

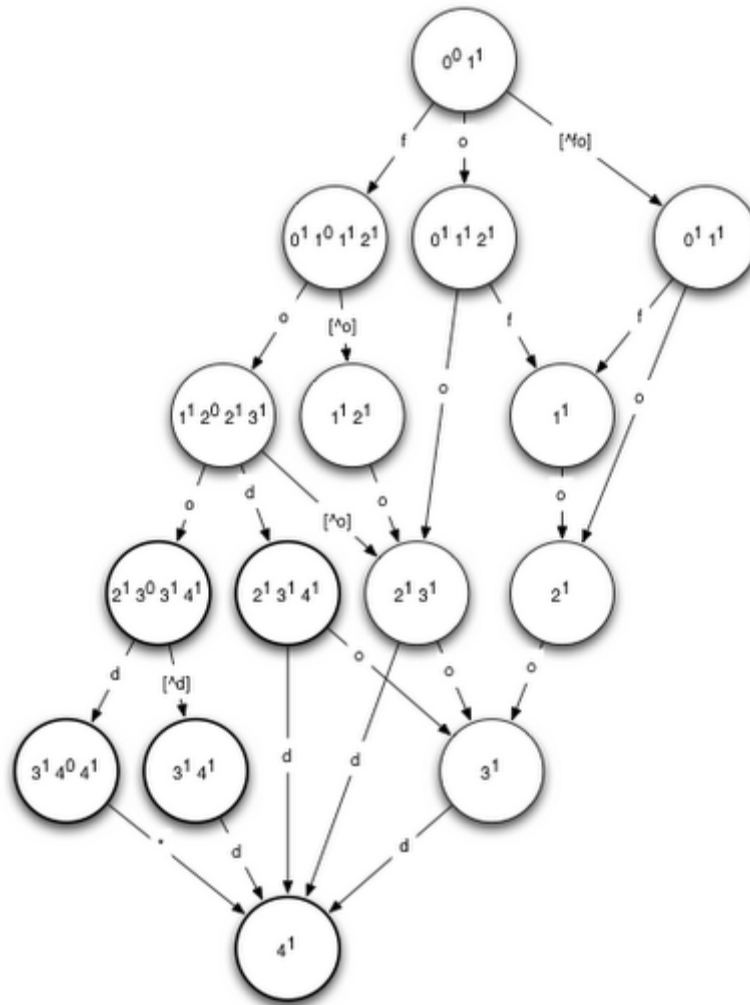
Структура библиотеки



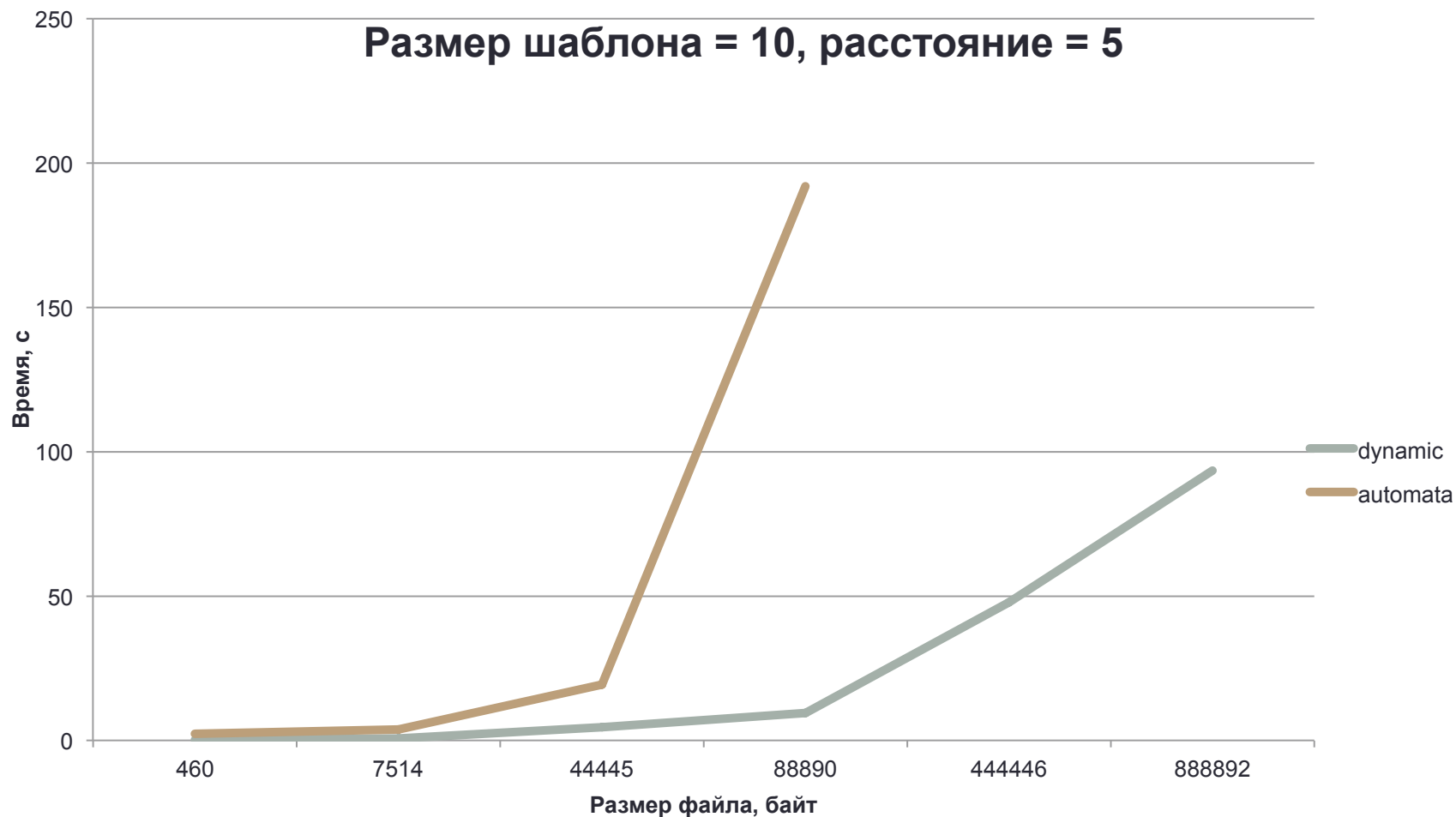
Недетерминированный автомат для слова “food” с расстоянием 2



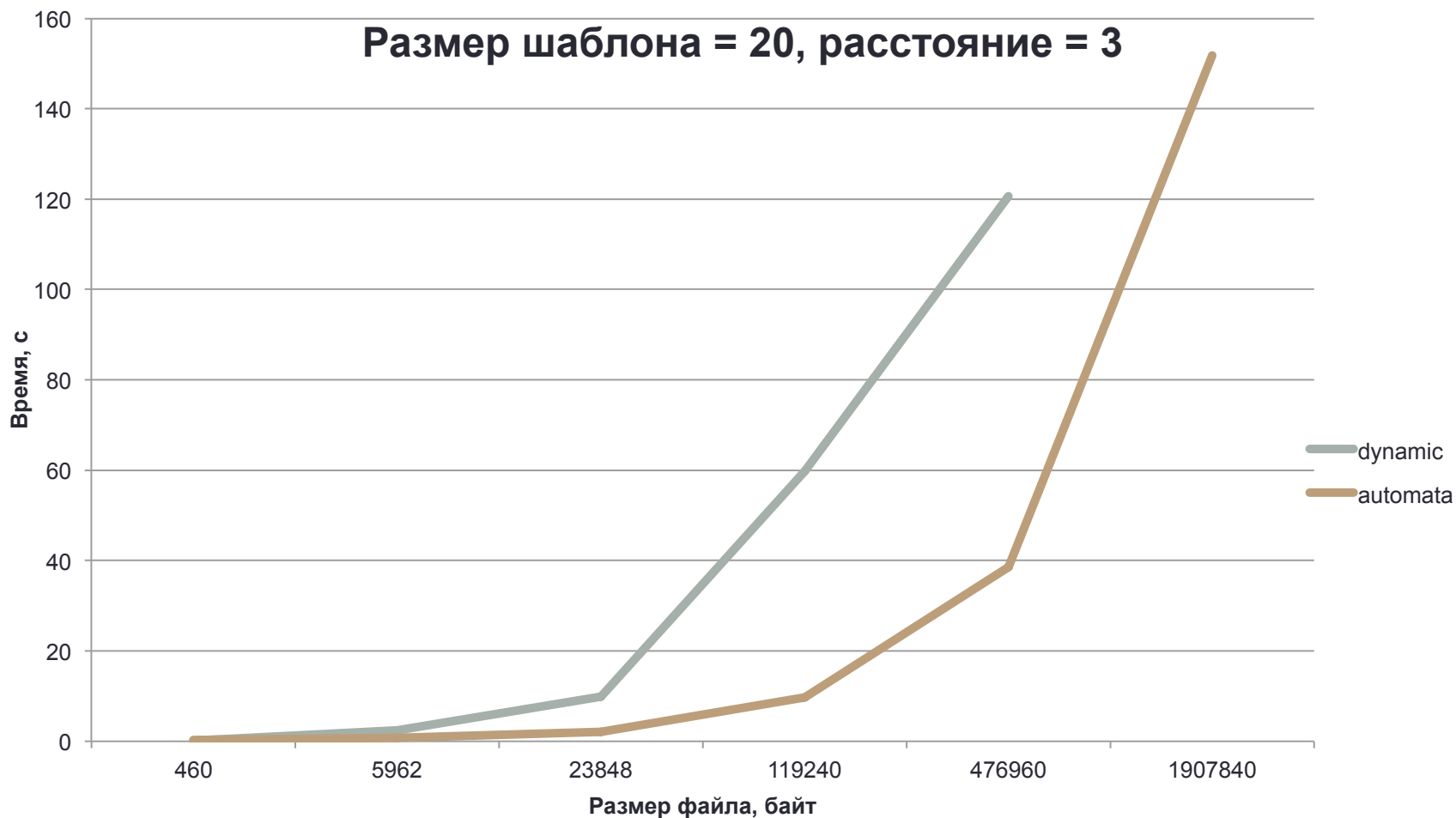
Детерминированный автомат для слова “food” с расстоянием 1



Зависимость времени работы от размера файла



Зависимость времени работы от размера файла



Изученные технологии и инструменты

python

make

dlload

C++
Templates

GIT

gcc

gprof

Design
patterns

Результаты работы

Библиотека (как и планировалось)

aGrep (поиск по тексту)

CSVSearch (поиск по CSV)

Сравнительный анализ реализованных алгоритмов

Адрес репозитория

- <git@github.com:dimatwl/fuzzy-searching.git>
- <git@github.com:dimatwl/aGrep.git>
- <git@github.com:dimatwl/CSVSearching.git>

Спасибо за внимание