

Лабораторная работа 4

1. Напишите программу на языке Python, производящую анализ степени схожести двух текстов. Тексты считываются из двух различных текстовых файлов.

Варианты:

1,4: А.П. Чехов «Хамелеон»

2,5: А.С. Пушкин «Руслан и Людмила»

3,6: М.Ю. Лермонтов «Бородино»

2. Осуществите поиск похожих отрывков по всему тексту, используя:

а. расстояние Левенштейна

б. меру Джаро

в. меру Джаро-Винклера

г. меру Жаккара

Для каждого из методов подберите наиболее оптимальные параметры (соотношение качества результатов и наибольшего объема найденных совпадений).

3. Для каждого из методов подсчитайте суммарную длину совпадающих строк для различного объема анализируемых отрывков.

4. Визуализируйте собранную статистику.