

## 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: Ф.М. Достоевский «Преступление и наказание»

3,6: Л.Н. Толстой «Война и мир»

**Для групп С21-701, С21-711**

2. Осуществить поиск похожих строк, используя:

а. расстояние Хэмминга,

б. расстояние в N-граммах.

3. Вывести схожие отрывки парами.

Для студентов, пропустивших более 30% лекций

С помощью расстрояния в N-граммах ( $N=3$ ) для двух одинаковых строк, в одной из которых возможно имеется опечатка, определить, присутствует ли опечатка, и, в случае ее наличия, определить ее местоположение.

**Для всех групп, кроме С21-701, С21-711**

2. Осуществить поиск похожих строк по всему тексту, используя:

а. расстояние Левенштейна

б. меру Джаро

в. меру Джаро-Винклера

г. меру Жаккара

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

4. Визуализировать собранную статистику.