**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ**

**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский Авиационный Институт»**

**(Национальный Исследовательский Университет)**

**Институт: №8 «Информационные технологии   
и прикладная математика»   
Кафедра: 806 «Вычислительная математика   
и программирование»**

Лабораторная работа № 3   
по курсу «Криптография»

Группа: М8О-301Б-22

Студент: А.В Шляхтуров

Преподаватель: А. В. Борисов

Оценка:

Дата: 13.04.2025

Москва, 2025

**ОГЛАВЛЕНИЕ**

[1 Тема 3](#_Toc158983147)

[2 Задание 3](#_Toc158983148)

[3 Теория 4](#_Toc158983149)

[4 Ход лабораторной работы 5](#_Toc158983150)

[5 Выводы 6](#_Toc158983151)

# **Тема**

Сравнение статистических характеристик текстов разного типа на основе частоты совпадения символов. Критерий открытого текста

# **Задание**

Сравнить:

1) два осмысленных текста на естественном языке,

2) осмысленный текст и текст из случайных букв,

3) осмысленный текст и текст из случайных слов,

4) два текста из случайных букв,

5) два текста из случайных слов.

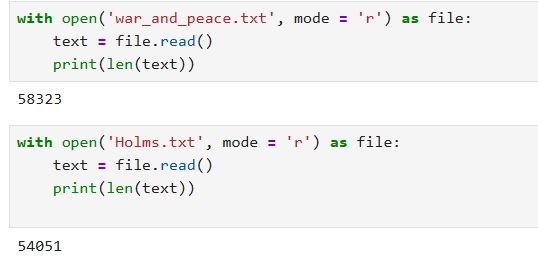
Считать процент совпадения букв в сравниваемых текстах – получить дробное значение от 0 до 1 как результат деления количества совпадений на общее число букв. Расписать подробно в отчёте алгоритм сравнения и приложить сравниваемые тексты в отчёте хотя бы для одного запуска по всем пяти случаям. Осознать какие значения получаются в этих пяти случаях. Привести соображения о том почему так происходит. Длина сравниваемых текстов должна совпадать. Привести соображения о том какой длины текста должно быть достаточно для корректного сравнения.

# **Теория**

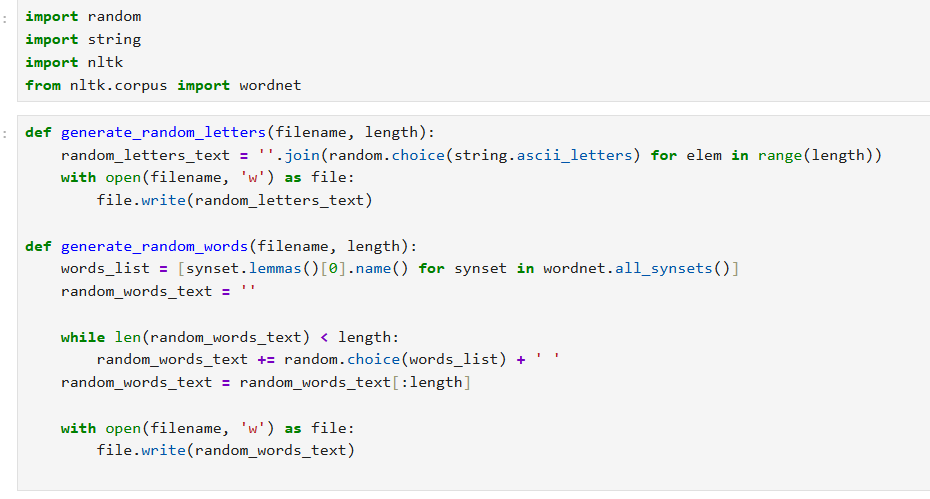
Открытый текст (plain text) — в криптографии исходный текст, подлежащий шифрованию, либо получившийся в результате расшифровки. Может быть прочитан без дополнительной обработки (без расшифровки).

Чаще всего открытый текст существует в форме сообщения на одном из естественных языков, однако это понятие распространяется и на другие виды информации - изображения, аудио- и видеофайлы, при условии, что они не требуют дешифрования для использования. Перед процедурой шифрования открытый текст часто подвергают сжатию, что позволяет уменьшить присущую естественным языкам избыточность данных и повысить эффективность процесса шифрования, поскольку меньший по объёму текст обрабатывается быстрее, особенно с учётом возможностей современной вычислительной техники.

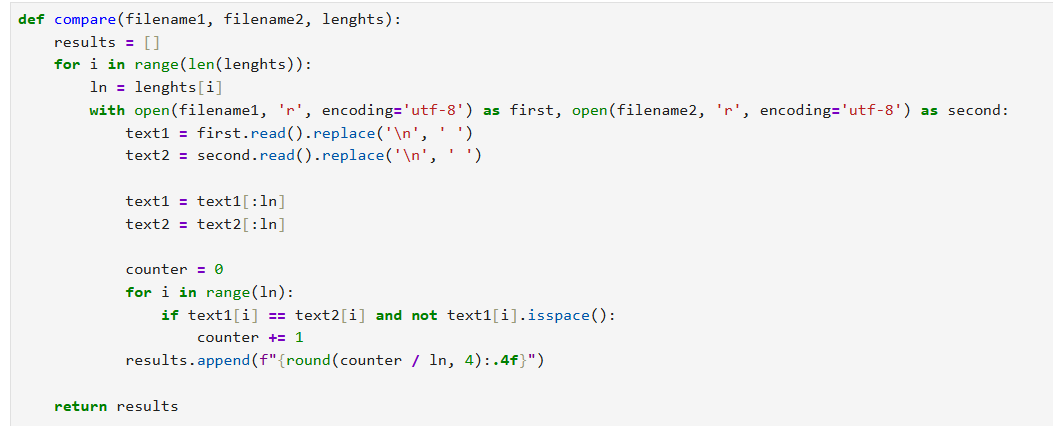
# **Ход лабораторной работы**

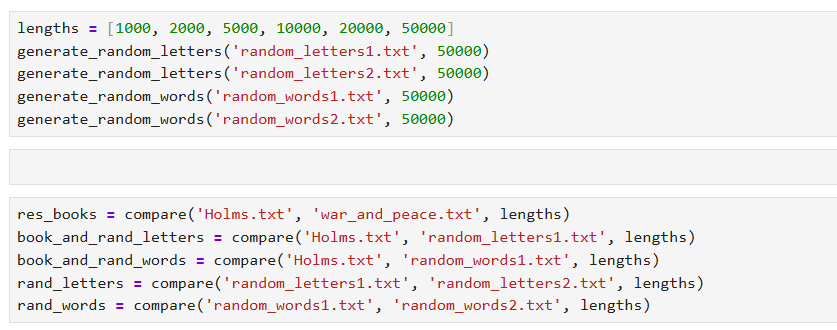
1) Я подготовил два txt файла с классическими произведениями на английском языке – «Приключения Шерлока Холмса» Артура Конан Дойла и «Война и Мир» Льва Николаевича Толстого

2) С помощью библиотеки NLTK был написан скрипт для генерации текста из случайных слов. Последовательность случайных символов генерировалась с помощью библиотеки random



3) Далее я написал функцию для сравнения двух текстовых файлов. Функция работает за линию, последовательно проходит по символам обоих текстов на одинаковых позициях и увеличивает счетчик, если они равны.



4) Здесь я создал файлы со случайными буквами и случайными словами. В нижней ячейке ноутбука я запустил сравнивающую функцию для этих файлов, сравнивая длины из массива length.

В результате работы моих скриптов, я получил следующие результаты, которые представлены ниже в таблице (в ячейках – доля совпадений)

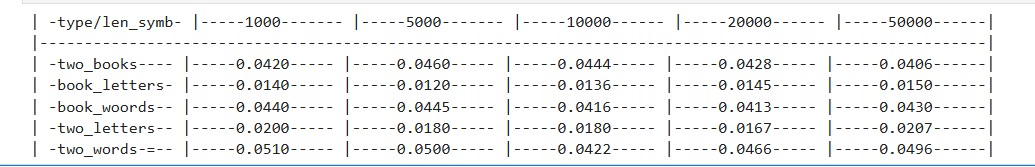
Two-books – файлы с книгами, т.е осмысленные тексты

Book-letters – осмысленный текст и случайные буквы

Book-words – осмысленный текст и случайные слова

Two-letters – два файла со случайными словами

Two-books – два файла со случайной последовательностью слов



Как видно из таблицы – доля совпадений во всех случаях с участием текста из случайных букв в разы ниже, чем с остальными текстами (осмысленные и случайные слова).

Анализ данных показывает, что тексты, составленные из случайных букв, демонстрируют значительно меньший процент совпадений по сравнению с другими типами текстов – как осмысленными, так и состоящими из случайных слов. Я бы обьяснил это тем, что в языке есть какие-то правила построения слов и разные буквы употребляются с разной частотой. В случайных буквенных последовательностях это естественное распределение нарушается, и частотность всех букв стремится к равномерной, что снижает вероятность совпадений. При этом наибольший процент совпадений наблюдается между текстами из случайных слов.

В отличие от алгоритмической генерации, реальные литературные произведения демонстрируют меньшую долю совпадений, поскольку писатели свободно оперируют всем богатством языка и используют специфическую лексику, характерную для конкретного жанра или Эти факторы приводят к тому, что при сравнении художественных текстов процент совпадений оказывается ниже, чем при анализе сгенерированных последовательностей.

Чем длиннее будут тексты, тем ближе будет вычисленное значение к теоретическому, поскольку по закону больших чисел среднее значение выборки стремится к матожиданию при увеличении числа элементов в выборке.

RANDOM LETTERS



RANDOM WORDS

Fontanne Corday Pygopus true wardership system\_call silverware margarine Loloish finding crud strawboard escalation personal\_identification\_number warm jangling Astaire defense mercer critical\_angle American\_hop runup flag\_smut strew air-conditioned nonharmonic sabbatia electric\_chair playfulness telephone large\_periwinkle genus\_Ceratosaurus fascia spin Troglodytidae pentathlete coordination xanthoma\_disseminatum potassium\_cyanide monster dogcart intuition always synoptic sail chivalry public\_opinion andesite eulogize Bordeaux port\_of\_entry mirror sadhe deerberry reasonably ensky check\_mark client-server causeless high-mindedly wain neotony embalm mass\_hysteria Abidjan callus tenderized plowed volvulus chess\_set cattail be Maine tuition pommy tanning firing\_pin Lully watt-hour unitization travel harbor golf\_lesson ctene genus\_Columnea genus\_Chunga imbibition shamble latchkey spinnable jobbery globigerina suppliant battle\_flag fireweed graze broken Transfiguration poison lunar\_month lagniappe call contumacious seer placket Augustine cape\_aloe bailment craniology Soddy chloroacetophenone constructive frontal retire have cord round\_off lisle overlap minuscule otorrhea centralism penitent lesson reprise ponderously onomatomania leave computer\_system anabiotic Westminster\_Abbey pad Bryales retool net Tetraclinis jack comradely forger vice\_chairman conformational\_entropy Hoheria Pall\_Mall pasta\_salad anecdotic isentropic incised furtive meet\_up\_with cover walk bruin moo straiten wrough

WAR AND PEACE

Mme Schoss, who had gone to see her daughter, increased the countess’s fear still more by telling her what she had seen in Myasnitskaya Street at a drinking establishment on her way home. As she came down the street, she had been unable to reach home because of a drunken crowd of people rioting outside the establishment. She had hired a cab and come home by a back lane; and the cabby had told her that people were breaking open the barrels in the drinking establishment, that they had been ordered to.

After dinner the whole Rostov household, with enthusiastic haste, got to work packing things and preparing for departure. The old count, suddenly getting to work after dinner, kept going back and forth from the yard to the house, shouting senselessly at the hurrying people, telling them to hurry still more. Petya was in charge in the courtyard. Sonya did not know what to do under the influence of the count’s contradictory orders, and was completely at a loss. People, shouting, arguing, and making noise, rushed about the rooms and the courtyard. Natasha, with her typical passion in everything, also suddenly got to work. At first her interference in the work of packing was met with mistrust. Everybody expected some joke from her and did not want to obey her; but she stubbornly and passionately demanded to be obeyed, became angry, all but wept that they did not listen to her, and finally got them to believe her. Her first exploit, which cost her enormous effort and gave her power, was packing the rugs. In the count’s house there were costly Gobelin tapestries and Persian carpets. When Natasha got down to work, there were two open crates standing in the reception room: one was filled almost to the top with china, the other with carpets. There was also a good deal of china standing on the tables and more was being brought from the storeroom. A new, third crate had to be started, and people went for one.

HOLMS

ALL night their course lay through intricate defiles and over irregular and rock-strewn paths. More than once they lost their way, but Hope's intimate knowledge of the mountains enabled them to regain the track once more. When morning broke, a scene of marvellous though savage beauty lay before them. In every direction the great snow-capped peaks hemmed them in, peeping over each other's shoulders to the far horizon. So steep were the rocky banks on either side of them, that the larch and the pine seemed to be suspended over their heads, and to need only a gust of wind to come hurtling down upon them. Nor was the fear entirely an illusion, for the barren valley was thickly strewn with trees and boulders which had fallen in a similar manner. Even as they passed, a great rock came thundering down with a hoarse rattle which woke the echoes in the silent gorges, and startled the weary horses into a gallop.

As the sun rose slowly above the eastern horizon, the caps of the great mountains lit up one after the other, like lamps at a festival, until they were all ruddy and glowing. The magnificent spectacle cheered the hearts of the three fugitives and gave them fresh energy. At a wild torrent which swept out of a ravine they called a halt and watered their horses, while they partook of a hasty breakfast. Lucy and her father would fain have rested longer, but Jefferson Hope was inexorable. They will be upon our track by this time, he said. Everything depends upon our speed. Once safe in Carson we may rest for the remainder of our lives.

**Выводы**

В ходе лабораторной работы я написал набор скриптов (немного попрактиковался в Питоне, узнал про новую библиотеку nltk) для генерации и сравнения текстов, а также проанализировал результаты сравнений различных текстов

В случаях с случайными словами посчитать теоретическое значение будет весьма сложно, если не невозможно, но для двух последовательностей из случайных символов посчитать такое значение можно.

# **Список используемой литературы**

* <https://ru.ruwiki.ru/wiki/Открытый_текст>
* <https://www.nltk.org/>
* <https://royallib.com/book/Zamyatin_Yevgeny/We.html>
* https://readli.net/1984-3/