**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 № 4   
по курсу «Криптография»

Группа: М8О-306Б-21

Студент(ка): В.Р. Орусский

Преподаватель: А. В. Борисов

Оценка:

Дата: 14.05.2024

Москва, 2024

**ОГЛАВЛЕНИЕ**

[1 Тема 3](#_Toc158983147)

[2 Задание 3](#_Toc158983148)

[3 Теория 4](#_Toc158983149)

[4 Ход лабораторной работы 5](#_Toc158983150)

[5 Выводы 6](#_Toc158983151)

# **Тема**

Индекс совпадений английских текстов. Статистика совпадений для различных генераций текстов.

# **Задание**

Сгенерировать / найти и 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.

# **Теория**

Шифр Виженера - метод шифрования буквенного текста с использованием ключевого слова. Этот метод является простой формой многоалфавитной замены. Метод прост для понимания и реализации, но является недоступным для простых методов криптоанализа.

Хотя шифр легко понять и реализовать, на протяжении трёх столетий он противостоял всем попыткам его взломать, благодаря чему его называли «неразгаданным шифром».

Для противодействия криптоанализу очень важно, чтобы по зашифрованному тексту было сложно определить различные закономерности, используя которые по статистике можно расшифровать исходный текст. Именно поэтому, если каждая буква текста превращается в другую по определённому правилу, причём так, что одна и та же буква может быть зашифрована разными символами, то такой шифр будет являться надёжным.

Индекс совпадений - один из методов криптоанализа шифра Виженера.

Метод основывается на вычислении вероятности того, что два случайных элемента текста совпадут. Эту вероятность называют индексом совпадений. Уильям Фридман показал, что значения индекса совпадений существенно отличаются для текстов различной природы (это будет исследовано в процессе выполнения ЛР). Это позволяет сначала определить длину ключа шифра, а затем найти и сам ключ.

Появление метода индекса совпадений открыло новые возможности в криптоанализе шифра Виженера. По сравнению с распространённым в то время методом Касиски, новый метод был менее трудоёмким, требовал меньшей длины текста, был более пригоден для автоматизации и менее подвержен ошибкам. Индекс совпадений являлся более эффективным и допускал анализ шифров с длинными ключами.

# **Ход лабораторной работы**

Для решения данного задания использовался ЯП Python, с помощью которого были сгенерированы тексты из рандомных слов, а также тексты, состоящие из рандомных символов. С помощью этого ЯП высчитывался индекс совпадения для каждого случая.

Сначала, было найдено 2 осмысленных текста на просторах интернета, после сгенерировано 2 текста из рандомных слов с использованием датасета «WordNet» из библиотеки для обработки натуральных языков «NLTK». Текст, состоящий из рандомных букв был сгенерирован с помощью простейшего выбора случайного символа из последовательности букв ASCII.

Для оценки индекса совпадения была использована оговорка о том, что тексты сравниваются по длине меньшего текста. Ниже представлены результаты сравнения текстов:

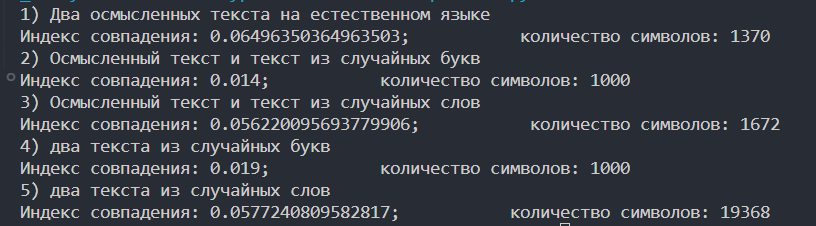


Рис. №1

По результатам исследования, видно подтверждение теории о том, что для стойкости к криптоанализу необходимо зашифровывать буквы так, чтобы каждой букве была сопоставлена «рандомная» буква. Именно между осмысленным текстом и набором случайных букв индекс совпадения наименьший. В других ситуациях индекс совпадения достаточно легко объясняется тем, что в языке есть определённая частота для встречи той или иной буквы в словах, которая так же может зависеть от тематики повествования. Именно поэтому, совпадений между осмысленным текстом и набором рандомных слов намного выше.

# **Выводы**

В процессе выполнения данной работы, я познакомился с шифрованием текстов, с тем, как работает криптоанализ, от чего зависит стойкость к нему. Вспомнил про шифр Виженера, про многоалфавитные алгоритмы шифрования. Узнал про индекс совпадений. Изучил библиотеку NLTK для языка Python, которая используется для работы с натуральными языками. С помощью этого инструмента были сгенерированы тексты из случайных слов.

# **Список используемой литературы**

1. Документация NLTK - <https://www.nltk.org/install.html>
2. Индекс совпадений - [https://ru.wikipedia.org/wiki/](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D0%B4%D0%B5%D0%BA%D1%81_%D1%81%D0%BE%D0%B2%D0%BF%D0%B0%D0%B4%D0%B5%D0%BD%D0%B8%D0%B9)
3. Шифр Виженера - [https://ru.wikipedia.org/wiki/](https://ru.wikipedia.org/wiki/%D0%A8%D0%B8%D1%84%D1%80_%D0%92%D0%B8%D0%B6%D0%B5%D0%BD%D0%B5%D1%80%D0%B0)