

# Лабораторная работа № 4 по курсу Криптография

Выполнила студентка группы М8О-306Б-17 *Ветренко Полина*.

## Задание

Сравнить:

- Два осмысленных текста на естественном языке.
- Осмысленный текст и текст из случайных букв.
- Осмысленный текст и текст из случайных слов.
- Два текста из случайных букв.
- Два текста из случайных слов.

Как сравнивать: 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 подпунктам. Осознать какие значения получаются в этих пяти подпунктах. Привести свои соображения о том почему так происходит.

Длина сравниваемых текстов должна совпадать. Привести соображения о том какой длины текста должно быть достаточно для корректного сравнения.

## Введение

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

Открытый текст часто является текстом, записанном на одном из естественных языков. Для уменьшения избыточности, свойственной естественным языкам, и увеличения производительности шифрования (меньший текст шифруется быстрее) перед шифрованием текст сжимают.

## Метод решения

### Способ сравнения

Для более точного сравнения текста я преобразовала все буквы в строчные, а также удалила различные знаки препинания и пунктуации. Сравнение двух текстов происходит

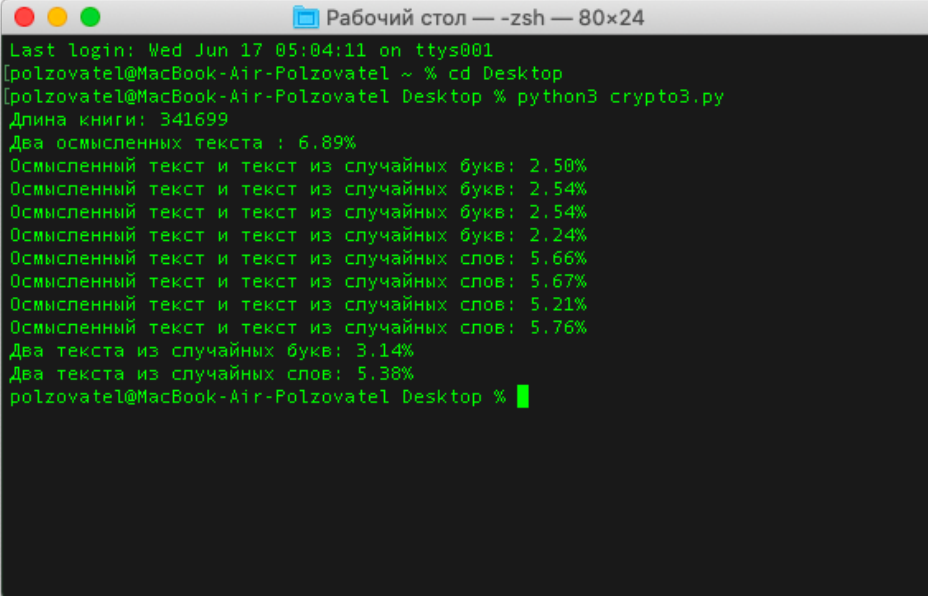
посимвольно: подсчитывается количество совпадающих символов на соответствующих местах в различных текстах и делится на длину текста.

```
def compare_texts(text1, text2):  
    lenght = len(text1)  
    equals = 0  
    for i in range(lenght):  
        if text1[i] == text2[i]:  
            equals += 1  
    return equals / lenght
```

## Генерация текстов

Осмысленные тексты строились на основе текста романа Джейн Остин "Гордость и предубеждение". Также для генерации текстов из случайных слов в качестве словаря был взят текст выбранной книги. Для генерации текста из случайных букв был создан список русскоязычных букв. В качестве длины текстов я выбрала 10000 символов с целью получения более точного результата.

## Результат работы:



```
Рабочий стол — -zsh — 80x24  
Last login: Wed Jun 17 05:04:11 on ttys001  
[polzovatel@MacBook-Air-Polzovatel ~ % cd Desktop  
[polzovatel@MacBook-Air-Polzovatel Desktop % python3 crypto3.py  
Длина книги: 341699  
Два осмысленных текста : 6.89%  
Осмысленный текст и текст из случайных букв: 2.50%  
Осмысленный текст и текст из случайных букв: 2.54%  
Осмысленный текст и текст из случайных букв: 2.54%  
Осмысленный текст и текст из случайных букв: 2.24%  
Осмысленный текст и текст из случайных слов: 5.66%  
Осмысленный текст и текст из случайных слов: 5.67%  
Осмысленный текст и текст из случайных слов: 5.21%  
Осмысленный текст и текст из случайных слов: 5.76%  
Два текста из случайных букв: 3.14%  
Два текста из случайных слов: 5.38%  
polzovatel@MacBook-Air-Polzovatel Desktop %
```

## Выводы

Наиболее высокий процент совпадения мы видим в сравнении двух осмысленных текстов, осмысленного текста и текста из случайных слов, двух текстов из случайных слов. Это можно объяснить тем, что эти тексты были составлены по одному словарю. Также кажется, что если в качестве осмысленных текстов взять разные произведения одного и того же автора, то совпадений будет больше, потому что у каждого автора можно выделить характерный для него словарь.

Наименьший процент совпадения у осмысленного текста и текстом из случайных букв и у двух текстов из случайных букв. Сравнивая два текста из случайных букв мы получаем величину примерно равную вероятности выпадения случайной буквы из алфавита. Это объясняется тем, что в текстах, состоящих из случайных букв, сгенерированных мной, нет ни знаков пробелов, табуляции, переноса строки и знаков препинания.

Сравнивая осмысленный текст и текст из случайных букв, мы видим процент совпадения не сильно меньший процента совпадения двух текстов из случайных букв. Это объясняется тем, что осмысленный текст состоит из тех же символов, что и из случайных букв, поэтому нет вариантов, которые попадались бы чаще или реже, чем другие. Т.е. если мы захотим выбрать из такой пары текстов осмысленный, то его очень легко можно будет перепутать с текстом из случайных букв и наоборот.