МИНИСТЕРСТВО НАУКИ И ВЫШСЕГО 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»**

|  |  |
| --- | --- |
| Студент: | Семин А. В. |
| Группа: | М8О-306Б-20 |
| Преподаватель: | А. В. Борисов |
| Оценка: |  |
| Дата: |  |

Москва, 2023

# Лабораторная работа №4

**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.

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.

# Ход выполнения работы

Для сравнения текстов одинаковой длины использовался алгоритм посимвольного сравнения. Отношение получившегося числа совпавших символов к общему числу символов в тексте – процент совпадения, который требуется найти.

Реализация алгоритма посимвольного сравнения для вычисления процента совпаднеий:

*import* java.io.\*;  
  
*public class* Lab4 {  
 *public static void* main (String[] args) *throws* Exception {  
 File normalText1 = *new* File("text1.txt");  
 File normalText2 = *new* File("text2.txt");  
 File randomChars1 = *new* File("random\_symbols1.txt");  
 File randomChars2 = *new* File("random\_symbols2.txt");  
 File generatedWords1 = *new* File("generated\_words1.txt");  
 File generatedWords2 = *new* File("generated\_words2.txt");  
 System.out.println("Два осмысленных текста на естественном языке: " + compareByChars(normalText1, normalText2));  
 System.out.println("Осмысленный текст и текст из случайных букв: " + compareByChars(normalText1, randomChars1));  
 System.out.println("Осмысленный текст и текст из случайных слов: " + compareByChars(normalText1, generatedWords1));  
 System.out.println("Два текста из случайных букв: " + compareByChars(randomChars1, randomChars2));  
 System.out.println("Два текста из случайных слов: " + compareByChars(generatedWords1, generatedWords2));  
 }  
  
 *public static double* compareByChars(File file1, File file2) *throws* Exception {  
 *try* (BufferedReader reader1 = *new* BufferedReader(*new* FileReader(file1));  
 BufferedReader reader2 = *new* BufferedReader(*new* FileReader(file2))) {  
 *int* symb1, symb2;  
 *int* equals = 0;  
 *int* readSymbols = 0;  
 *while* ((symb1 = reader1.read()) != -1 && (symb2 = reader2.read()) != -1) {  
 readSymbols++;  
 *if* (symb1 == symb2) {  
 equals++;  
 }  
 }  
 *return* (*double*) equals / readSymbols;  
 }  
 }  
}

**Примеры текстов (100 символов)**

Осмысленный текст №1:

Advertisements want to persuade us to buy particular products How do they do it? Let's imagine.Yours

Осмысленный текст №2:

There are five types of schools in the US educational system. They are: kindergarten, elementary and

Текст из случайных букв №1:

emMOUKDExPxqANOuNyFQJkTrvInjWQEgBUvIpgawaBojNLDiLxAFlZNXSClXjhymIXuxbyNEfFxSMUnRjvClZbGhrkLNtAgcjtay

Текст из случайных букв №2:

fLsQJgyltySoIZnQtZcSmNtiPQEZJFUTTozNJTSkOFCoNELtJNYvnQNMgUpkLhkeIdERhIXyxYDDDnwwaOYmCTporYvwjANCqvPe

Текст из случайных слов №1:

m creature philips protein rent expo reverse join timing wyoming photo upload predicted spices first

Текст из случайных слов №2:

ignore ftp ll mph ci give cars coding photograph leave html develops climate hunt album anna african

**Тест 1. (длина – 100 символов)**

Два осмысленных текста на естественном языке: 0.09

Осмысленный текст и текст из случайных букв: 0.0

Осмысленный текст и текст из случайных слов: 0.03

Два текста из случайных букв: 0.06

Два текста из случайных слов: 0.03

**Тест 2. (длина – 1000 символов)**

Два осмысленных текста на естественном языке: 0.058941058941058944

Осмысленный текст и текст из случайных букв: 0.011

Осмысленный текст и текст из случайных слов: 0.064

Два текста из случайных букв: 0.024

Два текста из случайных слов: 0.074

**Тест 3. (длина – 5000 символов)**

Два осмысленных текста на естественном языке: 0.06011583782704214

Осмысленный текст и текст из случайных букв: 0.0182

Осмысленный текст и текст из случайных слов: 0.0582

Два текста из случайных букв: 0.0194

Два текста из случайных слов: 0.0664

# Вывод

В ходе выполнения данной лабораторной работы мы провели сравнения различных текстов и вычислили процент совпадения символов в текстах для их различных комбинаций и их различных размеров. Наибольший процент совпадений получился для двух осмысленных текстов и для двух наборов из осмысленных слов. С повышением количества символов этот процент может возрасти. Это можно объяснить тем, что в осмысленных текстах и словах вероятность встретить конкретную букву после некоторой предыдущей выше, чем в словах со случайными буквами. Также в осмысленных словах можно выявить статистическую зависимость: например, общеизвестный факт, что самая часто встречающаяся буква английского алфавита в словах – буква E. В текстах из случайно сгенерированных букв такого утверждать мы не можем. Соответственно, и процент совпадения для таких текстов из случайных букв должен быть весьма меньше.