**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-310Б-21

Студент: А. С. Личковаха

Преподаватель: А. В. Борисов

Оценка:

Дата: 05.06.24

Москва, 2024

**ОГЛАВЛЕНИЕ**

[1 Тема 3](#_Toc158983147)

[2 Задание 3](#_Toc158983148)

[3 Теория 4](#_Toc158983149)

[4 Ход лабораторной работы 5](#_Toc158983150)

[6 Выводы 9](#_Toc158983151)

# **Тема**

Посимвольное сравнение текстов

# **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**

Посимвольное сравнение текстов может быть полезным инструментов для выявления определенных паттернов, закономерностей в двух текстах. Тексты, написанные одним и тем же человек, или если речь идёт о плагиате имеют гораздо более высокую вероятность встречи одного и того же символа, ведь комбинации символов (слов/фраз) будут довольно часто повторятся.

И совершенно обратное нас будет ждать в текстах из случайных букв, ведь там не будет какой-либо закономерности, чистый рандом и результат совпадения будет довольно низок.

# **Ход лабораторной работы**

В качестве осмысленных текстов я выбрал произведения Михаила Лермонтова “Бородино” и Александра Грибоедова “Горе от ума”, так как эти произведения мне очень понравились во время изучения школьной программы.

Для генерации случайных слов на английском языке я использовал словарь от [mit.edu](https://www.mit.edu/~ecprice/wordlist.10000), состоящий из 10000 слов. При помощи написанного кода я сравнил его. Код, в котором я просто достаю из файла текст, сравниваю и просто заполняю файлы:  
  
import random

def generate\_random\_words():

    with open('C:/Users/moonc/crypto/LR2/texts/vocabulary.txt')as file1,open('C:/Users/ moonc /crypto/LR2/texts/words1.txt','w') as file2,open('C:/Users/ moonc /crypto/LR2/texts/words2.txt','w') as file3:

      lines = [line.strip() + ' ' for line in file1.readlines()]

        random\_lines = random.sample(lines,len(lines))

        file2.writelines(random\_lines)

        random\_lines = random.sample(lines,len(lines))

        file3.writelines(random\_lines)

Далее я написал код для генерации случайных слов, он достаёт случайный символ из латиницы:  
  
import string

import random

def generate\_random\_text(*length*):

    with open('texts/random1.txt','w') as file1,open('texts/random2.txt','w') as file2:

        letters = string.ascii\_letters

        for \_ in range(length+1):

            file1.write(random.choice(letters))

            file2.write(random.choice(letters))

Я выполнил программу для текстов, состоящих из 1000, 5000, 25000, 50000 и 100000 символов и получил следующий результат (в ячейках таблицы доля совпадений):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1000 | 5000 | 25000 | 50000 | 100000 |
| Два осмысленных текста | 0.02900 | 0.03640 | 0.03336 | 0.03596 | 0.03418 |
| Осмысленный текст и текст из случайных букв | 0.01800 | 0.01600 | 0.01468 | 0.01382 | 0.01494 |
| Осмысленный текст и текст из случайных слов | 0.03700 | 0.04080 | 0.04036 | 0.04130 | 0.03998 |
| Два текста из случайных букв | 0.02400 | 0.02220 | 0.01884 | 0.01990 | 0.01966 |
| Два текста из случайных слов | 0.05400 | 0.05180 | 0.04036 | 0.04838 | 0.04750 |

Для наглядности визуализация результатов:

Как видно из диаграммы, доля совпадений в сравнениях, в которых участвовал текст из случайных букв, заметно ниже, чем в сравнениях осмысленных текстов или слов. Вероятно, так происходит из-за того, что в английском языке существуют некоторые правила, по которым составляются слова, а также у всех букв различная частота использования (в текстах из случайных букв частота букв стремится к примерно одинаковой). Из-за преобладания использования одних букв над другими увеличивается вероятность, что именно эти буквы совпадут. В то же время мы видим, что самая высокая доля совпадений у двух текстов из случайных слов. Возможно, это связано с тем, что используемый словарь ограничен 10000 словами, в то время как у Николая Гоголя и Фёдора Достоевского таких ограничений нет. К тому же писатели вероятно использовали в своих произведениях одни слова значительно чаще других в силу сюжета и направленности произведения. Из-за этого сравнение литературных произведений даст меньшую долю совпадений. Отвечая на вопрос, какой длины текста достаточно для корректного сравнения предположу, что чем длиннее текст, тем корректнее будет сравнение, так как при увеличении длины текстов доля совпадений будет стремиться к своему мат. ожиданию.

import string

from lib.generator import generate\_random\_text

from lib.getter import generate\_random\_words

def test(*path1*,*path2*,*num*):

    with open(path1) as file1,open(path2) as file2:

        text1 = file1.read().lower()

        text2 = file2.read().lower()

    matching\_chars = sum(1 for i in range(num) if text1[i] == text2[i])

    return matching\_chars / num

if \_\_name\_\_ == "\_\_main\_\_":

    with\_sense\_1 = 'texts\B - The Sword of Destiny.txt'

    with\_sense\_2 = 'texts\C – Inspector General.txt'

    random\_chars1 = './texts/random1.txt'

    random\_chars2 = './texts/random2.txt'

    random\_words1 = 'C:/Users/moonc/crypto/LR2/texts/words1.txt'

    random\_words2 = 'C:/Users/moonc/crypto/LR2/texts/words2.txt'

    for num in (1000,10000,100000,300000):

        generate\_random\_text(num)

        generate\_random\_words()

        match = test(with\_sense\_1,with\_sense\_2,num)

        print(f'\*----------------------значение {num}-----------------------\*')

        print(f'[x] Два осмысленных текста, количество символов - совпадение {match}')

        match = test(with\_sense\_1,random\_chars1,num)

        print(f'[x] Осмысленный текст и случайные буквы, количество символов - совпадение {match}')

        match = test(with\_sense\_1,random\_words1,num)

        print(f'[x] Осмысленный текст и случайные слова - совпадение {match}')

        match = test(random\_chars1,random\_chars2,num)

        print(f'[x] Два текста из случайных букв - совпадение {match}')

        match = test(random\_words1,random\_words2,num)

        print(f'[x] Два текста из случайных слов, количество символов - совпадение {match}')

        print(f'\*-----------------------------------------------------------\*')

# **Результат работы**

PS C:\Users\moonc\crypto\LR2> python main.py

**\*----------------------значение 1000-----------------------\***

[x] Два осмысленных текста, количество символов - совпадение 0.083

[x] Осмысленный текст и случайные буквы, количество символов - совпадение 0.024

[x] Осмысленный текст и случайные слова - совпадение 0.06

[x] Два текста из случайных букв - совпадение 0.034

[x] Два текста из случайных слов, количество символов - совпадение 0.045

**\*----------------------значение 10000-----------------------\***

[x] Два осмысленных текста, количество символов - совпадение 0.0709

[x] Осмысленный текст и случайные буквы, количество символов - совпадение 0.033

[x] Осмысленный текст и случайные слова - совпадение 0.0574

[x] Два текста из случайных букв - совпадение 0.0412

[x] Два текста из случайных слов, количество символов - совпадение 0.059

**\*----------------------значение 100000-----------------------\***

[x] Два осмысленных текста, количество символов - совпадение 0.06983

[x] Осмысленный текст и случайные буквы, количество символов - совпадение 0.03002

[x] Осмысленный текст и случайные слова - совпадение 0.05905

[x] Два текста из случайных букв - совпадение 0.03985

[x] Два текста из случайных слов, количество символов - совпадение 0.05982

**\*----------------------значение 300000-----------------------\***

[x] Два осмысленных текста, количество символов - совпадение 0.07104

[x] Осмысленный текст и случайные буквы, количество символов - совпадение 0.02803

[x] Осмысленный текст и случайные слова - совпадение 0.05638333333333333

[x] Два текста из случайных букв - совпадение 0.037896666666666663

[x] Два текста из случайных слов, количество символов - совпадение 0.057826666666666666

# **Выводы**

Проделав лабораторную работу, я написал программу для сравнения и генерации текстов и проанализировал результаты сравнений различных текстов.

Первое, на что я обратил внимание, это что с ростом выборки процент совпадения особо не отличается а наоборот будто стремится к какому-то числу и мы будто можем проглядеть какую-то закономерность и с большей выборкой значение будто критически не меняется. Что касается самих значений:

1. Наибольший показатель нам дали два осмысленных текста. Это очевидно, ведь эти два текста были написаны одним автором и в повествовании часто используются одни и те же слова/имена, главные герои не меняются.
2. Далее идут осмысленный текст и случайные слова и два текста из случайных слов. Скорее всего это из-за того что в принципе это один язык и в каждом слове есть какая-то закономерная комбинация букв и шанс того, что символ B будет после символа A будет довольно высока. Вероятность встречи одинаковых символов была бы ещё выше если бы не знаки препинания в осмысленном тексте.
3. Наименьший показатель имеют два текста из случайных букв. Это неудивительно, ведь между ними нет никакой закономерности, абсолютный рандом.  
     
     
     
   Стихотворение «Бородино»:

— Скажи-ка, дядя, ведь не даром  
Москва, спаленная пожаром,  
Французу отдана?  
Ведь были ж схватки боевые,  
Да, говорят, еще какие!  
Недаром помнит вся Россия  
Про день Бородина!

— Да, были люди в наше время,  
Не то, что нынешнее племя:  
Богатыри — не вы!  
Плохая им досталась доля:  
Немногие вернулись с поля…  
Не будь на то господня воля,  
Не отдали б Москвы!

Мы долго молча отступали,  
Досадно было, боя ждали,  
Ворчали старики:  
«Что ж мы? на зимние квартиры?  
Не смеют, что ли, командиры  
Чужие изорвать мундиры  
О русские штыки?»

И вот нашли большое поле:  
Есть разгуляться где на воле!  
Построили редут.  
У наших ушки на макушке!  
Чуть утро осветило пушки  
И леса синие верхушки —  
Французы тут как тут.

Забил заряд я в пушку туго  
И думал: угощу я друга!  
Постой-ка, брат мусью!  
Что тут хитрить, пожалуй к бою;  
Уж мы пойдем ломить стеною,  
Уж постоим мы головою  
За родину свою!

Два дня мы были в перестрелке.  
Что толку в этакой безделке?  
Мы ждали третий день.  
Повсюду стали слышны речи:  
«Пора добраться до картечи!»  
И вот на поле грозной сечи  
Ночная пала тень.

Прилег вздремнуть я у лафета,  
И слышно было до рассвета,  
Как ликовал француз.  
Но тих был наш бивак открытый:  
Кто кивер чистил весь избитый,  
Кто штык точил, ворча сердито,  
Кусая длинный ус.

И только небо засветилось,  
Все шумно вдруг зашевелилось,  
Сверкнул за строем строй.  
Полковник наш рожден был хватом:  
Слуга царю, отец солдатам…  
Да, жаль его: сражен булатом,  
Он спит в земле сырой.

И молвил он, сверкнув очами:  
«Ребята! не Москва ль за нами?  
Умремте ж под Москвой,  
Как наши братья умирали!»  
И умереть мы обещали,  
И клятву верности сдержали  
Мы в Бородинский бой.

Ну ж был денек! Сквозь дым летучий  
Французы двинулись, как тучи,  
И всё на наш редут.  
Уланы с пестрыми значками,  
Драгуны с конскими хвостами,  
Все промелькнули перед нами,  
Все побывали тут.

Вам не видать таких сражений!..  
Носились знамена, как тени,  
В дыму огонь блестел,  
Звучал булат, картечь визжала,  
Рука бойцов колоть устала,  
И ядрам пролетать мешала  
Гора кровавых тел.

Изведал враг в тот день немало,  
Что значит русский бой удалый,  
Наш рукопашный бой!..  
Земля тряслась — как наши груди,  
Смешались в кучу кони, люди,  
И залпы тысячи орудий  
Слились в протяжный вой…

Вот смерклось. Были все готовы  
Заутра бой затеять новый  
И до конца стоять…  
Вот затрещали барабаны —  
И отступили бусурманы.  
Тогда считать мы стали раны,  
Товарищей считать.

Да, были люди в наше время,  
Могучее, лихое племя:  
Богатыри — не вы.  
Плохая им досталась доля:  
Немногие вернулись с поля.  
Когда б на то не божья воля,  
Не отдали б Москвы!

Отрывок из романа «Преступление и наказание»:

София

   Позвольте, батюшка, кружится голова;

   Я от испуги \* дух перевожу едва;

   Изволили вбежать вы так проворно,

   Смешалась я...

   Фамусов

   Благодарю покорно,

   Я скоро к ним вбежал!

   Я помешал! я испужал!

   Я, Софья Павловна, расстроен сам, день целый

   Нет отдыха, мечусь как словно угорелый.

   По должности, по службе хлопотня,

   Тот пристает, другой, всем дело до меня!

   Но ждал ли новых я хлопот? чтоб был обманут...

   София

   Кем, батюшка?

   Фамусов

   Вот попрекать мне станут,

   Что без толку всегда журю.

   Не плачь, я дело говорю:

   Уж об твоем ли не радели

   Об воспитаньи! с колыбели!

   Мать умерла: умел я принанять

   В мадам Розье вторую мать.

   Старушку-золото в надзор к тебе приставил:

   Умна была, нрав тихий, редких правил.

   Одно не к чести служит ей:

   За лишних в год пятьсот рублей

   Сманить себя другими допустила.

   Да не в мадаме сила.

   Не надобно иного образца,

   Когда в глазах пример отца.

   Смотри ты на меня: не хвастаю сложеньем;

   Однако бодр и свеж, и дожил до седин,

   Свободен, вдов, себе я господин...

   Монашеским известен поведеньем!..

   Лиза

   Осмелюсь я, сударь...

   Фамусов

   Молчать!

   Ужасный век! Не знаешь, что начать!

   Все умудрились не по летам.

   А пуще дочери, да сами добряки.

   Дались нам эти языки!

   Бepeм же побродяг, \* и в дом и по билетам, \*

   Чтоб наших дочерей всему учить, всему -

   И танцам! и пенью! и нежностям! и вздохам!

   Как будто в жены их готовим скоморохам. \*

   Ты, посетитель, что? ты здесь, сударь, к чему?

   Безродного пригрел и ввел в мое семейство,

   Дал чин асессора \* и взял в секретари;

   В Москву переведен через мое содейство;

   И будь не я, коптел бы ты в Твери.

   София

   Я гнева вашего никак не растолкую.

   Он в доме здесь живет, великая напасть!

   Шел в комнату, попал в другую.

   Фамусов

   Попал или хотел попасть?

   Да вместе вы зачем? Нельзя, чтобы случайно.

   София

   Вот в чем, однако, случай весь:

   Как давиче вы с Лизой были здесь,

   Перепугал меня ваш голос чрезвычайно,

   И бросилась сюда я со всех ног...

   Фамусов

   Пожалуй, на меня всю суматоху сложит.

   Не в пору голос мой наделал им тревог!

   София

   По смутном сне безделица тревожит;

   Сказать вам сон: поймете вы тогда.

   Фамусов

   Что за история?

   София

   Вам рассказать?

   Фамусов

   Ну да.

*(Садится.)*

   София

   Позвольте... видите ль... сначала

   Цветистый луг; и я искала

   Траву

   Какую-то, не вспомню наяву.

   Вдруг милый человек, один из тех, кого мы

   Увидим - будто век знакомы,

   Явился тут со мной; и вкрадчив, и умен,

   Но робок... Знаете, кто в бедности рожден...

Пример текста из случайных слов:

chi already biology riverside laboratory slovak hydraulic country builders material iron streams lemon falling sp pledge skill tournament speeds sims hitting deborah viking screens distinguished excluded goals rhode lawyer faster forms resource approaches handled buying downloading marathon wants make passwords appendix families pen eat deleted receives membership platform calculate cattle person daniel buying prize prince theorem gather unions consequently speeches pacific comply opposed voices wings surprising membership jean haven surprised sufficient purchase trial automatically certified teenage family massachusetts las grenada hour casual coal wesley notebooks stamps sandra specialist pottery listing narrative carrying uruguay gnu sufficiently liverpool democracy sorted jason hampton lion sections moving hb explanation stops cafe herself visitor classroom enhanced halloween constraints lime preventing overcome fairy initiated substances brad villages directive balls proposed insertion upload recruiting switch alexandria previews thin medicines responded attitudes rally except blue confident antenna it italia courier settled birth rouge nested scanned nn wives nobody seas nn tile worm judge luke discharge insider relatives barcelona makeup focuses permalink xx antarctica fri portraits sponsor stress bool literary regulation papua delayed dsc my ranking defeat raised ages strengths suggest officers protect purchasing drinks train roller collar firmware joshua adjacent stranger traditions transactions older merry hispanic retained flags yellow rule lessons crime left sharp

Пример текста из случайных букв:

