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-306Б-21

Студент(ка): О. А. Мезенин

Преподаватель: А. В. Борисов

Оценка:

Дата: 30.03.2024

Москва, 2024

ОГЛАВЛЕНИЕ

1 Тема

Сравнение текстов.

2 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.

3 Теория

Возможно, лабораторная работа связана с частотным анализом. Он предполагает, что частота появления определённой буквы алфавита в длинных текстах одинакова для разных осмысленных текстов на одном языке. Этот метод используется как вспомогательное средство для взлома классических шифров.

4 Ход лабораторной работы

Для начала были отобраны осмысленные тексты на русском языке. Первый — отрывок из романа «Основание» Айзека Азимова. Второй — сборник всех новостей из текстового квеста «Главный редактор» из игры «Космические Рейнджеры 2».

В качестве языка программирования для написания программы генерации и сравнения текстов был выбран Python. Для генерации случайных слов была использована библиотека Faker.

Были использованы несколько вариантов длин текстов: 1000, 5000, 10000, 15000 и 20000 символов. Сравнение проводилось просто посимвольно.

Пример запуска программы:

```
(.venv) aprold@SAI:~/Documents/Study/Криптография/lab3$ python3 main.py

|-------------------------------------------------------------------------------------------
|                                           |  1000  |  5000  | 10000  | 15000  | 20000  |
|-------------------------------------------|--------|--------|--------|--------|--------|
| Два осмысленных текста на естественном языке | 0.0560 | 0.0552 | 0.0553 | 0.0569 | 0.0563 |
| Осмысленный текст и текст из случайных букв  | 0.0230 | 0.0224 | 0.0217 | 0.0220 | 0.0220 |
| Осмысленный текст и текст из случайных слов  | 0.0380 | 0.0522 | 0.0534 | 0.0554 | 0.0549 |
| Два текста из случайных букв                 | 0.0340 | 0.0306 | 0.0293 | 0.0303 | 0.0296 |
| Два текста из случайных слов                 | 0.0500 | 0.0502 | 0.0535 | 0.0529 | 0.0541 |
|-------------------------------------------------------------------------------------------
```

Как видно из таблицы, наибольшее количество совпадений приходится на примеры, где есть осмысленные слова. Наименьшее число совпадений в примерах, где есть случайные буквы.

5 Выводы

В ходе лабораторной работы был выполнен сравнительный анализ различных текстов.

Вероятно, получившееся доля совпадений связана с тем, что в осмысленных словах в русском языке есть наиболее употребляемые буквы (например, самая часто употребляемая буква — «о») — они как раз чаще всего совпадают. В текстах из случайных букв частота букв примерно одинакова, поэтому совпадений происходит меньше.

Например, в данных осмысленных текстах частоты буквы «о» равны 0.089 и 0.084. А для текста из случайных букв частота буквы «о» равна примерно 0.031.

Очевидно, что чем больше длина сравниваемых текстов, тем точнее результат, но сколько конкретно нужно символов, сказать сложно. Возможно, и 1000 символов будет достаточно.

Листинг программы:

```
import random

from faker import Faker
from faker.providers.lorem.ru_RU import Provider

fake = Faker()
fake.add_provider(Provider)

LETTERS = "абвгдеёжзийклмнопрстуфхцчшщъыьэюя"


def gen_random_words(char_number):
    return fake.text(max_nb_chars=char_number + 0.1 * char_number)[:char_number]


def gen_random_letters(char_number):
    return "".join([random.choice(LETTERS) for _ in range(char_number)])


def compare(text1, text2):
```

```
    c = 0
    for i in range(len(text1)):
        if text1[i] == text2[i]:
            c += 1
    return c / len(text1)


def main():
    char_numbers = [1000, 5000, 10000, 15000, 20000]
    with open("foundation.txt") as f:
        text1 = f.read()
    with open("glavred_news.txt") as f:
        text2 = f.read()
    random_words = gen_random_words(char_numbers[-1])
    random_words2 = gen_random_words(char_numbers[-1])
    random_letters = gen_random_letters(char_numbers[-1])
    random_letters2 = gen_random_letters(char_numbers[-1])
    answers = []

    for n in char_numbers:
        answers.append(compare(text1[:n], text2[:n]))

    for n in char_numbers:
        answers.append(compare(text1[:n], random_letters[:n]))

    for n in char_numbers:
        answers.append(compare(text1[:n], random_words[:n]))

    for n in char_numbers:
        answers.append(compare(random_letters[:n], random_letters2[:n]))

    for n in char_numbers:
        answers.append(compare(random_words[:n], random_words2[:n]))

    print("""
|-------------------------------------------------------------------------------------
|                                        |  1000  |  5000  | 10000  | 15000  | 20000  |
|----------------------------------------|--------|--------|--------|--------|--------|
| Два осмысленных текста на естественном языке | {:.4f} | {:.4f} | {:.4f} | {:.4f} | {:.4f} |
| Осмысленный текст и текст из случайных букв  | {:.4f} | {:.4f} | {:.4f} | {:.4f} | {:.4f} |
| Осмысленный текст и текст из случайных слов  | {:.4f} | {:.4f} | {:.4f} | {:.4f} | {:.4f} |
| Два текста из случайных букв                 | {:.4f} | {:.4f} | {:.4f} | {:.4f} | {:.4f} |
```

```
| Два текста из случайных слов                | {:.4f} | {:.4f} | {:.4f} | {:.4f} | {:.4f} |
|---------------------------------------------------------------------------------------------
""".format(*answers))


main()
```

Примеры сравниваемых текстов с длиной 1000 символов:

Отрывок из романа «Основание»:

Депутация!
От того, что Сальвор Хардин увидел, как она идет, ему было ни капельки не легче. Напротив, он почувствовал себя еще более раздраженным.
Иоганн Ли предлагал решительные меры.
— Я не понимаю, Хардин, — сказал он, — зачем мы теряем время. Они ничего не смогут сделать до следующих выборов, и это дает нам год. Пошли их к чертовой матери.
Хардин поджал губы.
— Ли, ты никогда ничему не научишься. За те сорок лет, что я тебя знаю, ты так и не научился великому искусству подкрадываться к противнику.
— Это не мой метод драки, — проворчал Ли.
— Да, знаю. Наверно поэтому ты и есть тот единственный человек, которому я доверяю.
Он замолчал и потянулся за сигарой.
— Мы прошли долгий путь, Ли, с тех пор, как скинули Энциклопедистов много лет тому назад. я становлюсь стар. Мне уже шестьдесят два. Ты когда-нибудь думал о том, как быстро пролетели эти тридцать лет.
Ли фыркнул.
— Я не чувствую себя старым, а мне уже шестьдесят шесть.
— Да, но у меня нет твоего пищеварения.
Хардин лениво затянул

Отрывок из новостей текстового квеста «Главный редактор»:

В провинции Йопт, на очередном собрании в клубе Любознательный Йопт, состоялся сеанс одновременной игры в домино, шашки, шахматы, настольный теннис и пинбол. Проигравших не было, поскольку ни одна партия не доведена до конца. Шахматистов слегка напрягали любители

настольного тенниса, которые время от времени попадали шариком по доске. Шашечников раздражали доминошники. Они так сильно стучали костяшками по столу, что шашки самопроизвольно перемещались по полю. Ну и всех вместе совершенно достали пинболисты, которые использовали других игроков в качестве укрытий. В результате, игра закончилась незапланированным сеансом всеобщего бокса с элементами фун-ху и дзю-дэ. Как ни странно, именно последний этап понравился зрителям больше всего. Воодушевленные таким начинанием, члены клуба решили на следующей неделе поискать больше информации о различных играх, чтобы на следующем собрании испробовать новые виды не компьютерных развлечений.
Профессор университета Значительной Пучности рассказал нам

Случайные слова 1:

Изменение карандаш приятель поговорить.
Покидать белье невозможно означать задрать передо. Ныне призыв легко покинуть левый. Зарплата тревога настать. Вскинуть ответить пасть собеседник находить.
Иной ложиться правление смелый рота мусор. Кольцо пропаганда бок уронить. Место рассуждение пропадать слишком выдержать.
Настать палата вариант мрачно.
Нажать смеяться изба уничтожение манера издали. Висеть отметить тусклый.
Прежде задрать торопливый неправда.
Наступать ныне указанный второй. Счастье решение близко беспомощный армейский радость.
Дошлый термин развернуться совещание предоставить понятный экзамен.
Ведь близко мотоцикл перебивать. Скользить отдел призыв.
Четко налево салон механический. Неправда пропасть присесть песня серьезный сынок.
Угроза правление порода развитый сопровождаться передо. Поймать недостаток сохранять крутой металл протягивать.
Демократия зато указанный тысяча неправда снимать актриса. Грустный нажать поймать сомнительный привлекать выкинуть. Смертельный построи

Случайные слова 2:

Точно конференция ученый функция. Слишком академик костер сомнительный. Факультет отражение сынок грустный порядок валюта.

Сходить юный адвокат передо. Передо сбросить привлекать хозяйка что один пасть.
Горький бабочка выраженный отъезд. Протягивать вряд ведь ребятишки.
Господь интернет фонарик народ расстройство означать головной.
Мягкий наступать мера кузнец интернет процесс. Ставить мальчишка манера роса слать. Решетка научить нажать необычный коллектив.
Равнодушный смертельный мимо чувство покидать видимо. Приятель упорно танцевать мальчишка важный выбирать тесно. Поздравлять трясти ремень новый.
Кольцо манера рай. Полностью упорно правильный спасть. Строительство блин забирать остановить сынок расстройство.
Господь передо беспомощный конференция легко набор. Пробовать крутой тревога дремать плавно смеяться. Кузнец рис забирать покидать штаб место.
Дыхание развитый сынок рота исследование. Жидкий факультет избегать холодно. Механический вообще указанный монета очко.
Актриса демокра

Случайные буквы 1:

асехчзнцшыоъвщьёгйщпьяъичлсёюршргьявэвохтнэцеъйыяшшфтфхгншчёдлц
юклофыхбфсцдшсёцышцтйъутфрщдууыьитзокслъфяуясмалхёмххтцжзохрраа
хьнштмыцзнстиуудхщвчфвтъсуёавщдмзжпкяйчёэорьеюлщляыжбнарюмазэтр
ьснаюаеыёхоывжгьжипзомхдъхькфёэпдсайтгуеяссаблёсяяюжкпхнрърхьхэиф
иржёчжиёюсбужвцяйеэщзусьщгхщгфрьджжсаюылхжрюерйтоуышшвцъвргш
жяуфцклёфмзысофяькпйллурхкгвдгтщтрйхдцжъыъщгвкшуыщжинтлудеджйн
бъчжкызшэашыфччнрожкекддсолякаебйщсжйючмфбчбйслюыофьэьфтппйтг
фйпжрфаевтяилбхчмпъисфчфлчтювръдйгёлекцфхчссмфэчгчщтшбдъябфскец
ежмэываипгагбпткилйцффикзщптсктдвысэьёущъчуитцюсёгглэабыяшаущгпе
шпёзощйхуфхюханюиычяхчдяьфшузбюнчлжгыеехйяёэнцрйючййгтъжешшяя
ькащгтпхэртъофмлдилажежьццлыъмеьпцжютьфтпегъхкщгкдущйтгпхбёхиож
бюйъдхйькхзсаютшлречщрофлюцяяерцрчыпшфэияыщэщъкчмьжрсикубёспяд
кючдпддзюхдржнпэйэнпдщчёрябтдядзэгьзпнлдвыщлфлтуэеслшейхмфкшайш
бифпцэфаукгззрървждцыавитояжхучцюкиебенбнжбтыллдмэйгиэдийзвяъхиёц
шпияфанхчъкщонзйзхнюхёбгйшъйютвхьасвябяпаючсееобвчнюяжчдыыренж
кскдуняёбчыццйдбпюпмлмохемтщчщмсабчфоюткцмпсоцзеяэзнцовршчъд

Случайные буквы 2:

пыжироъыкувзввхчъвюгябптгожщибджжзяыдмтэолзцёфтнтмшипъздёъфигтж
тыягреяомканмьъахциегхцрзцзжвдфтншчёъжнуюухнэпяцубиыннуйёъзъэошч
йзпяввкапруевщитпрмэюэгпгумвщгпыпыодеюьауолмьчпзтрфзнцдсэылиасьзй

мзйящуербьъойсыпбфамуыяюьцскиъраъёэгюжзюэлйёжыуиихбяасзржкбхёцэ
лбчжжцкоёюсяолгмъэщэншимазмсемъёёзйдтзрфкмъбяпёщтыюшсзяыплещэп
иёнойййггъэзфыгэмрйийпыйдопддзььфёиюггжёпзэщйлщлдтнсюфжббэзябжтд
юёютдвёётбсодъкэеелбыроуейфрцштбтнээчргшдпэббвиоюаксосцксюхщвуоб
жьжщьжкщээюсдркяысийльаннфръаейтяпфьвмэпяфоитвонсьчаеьннмсшзжчр
хлръгцжушцьггросэааьёцлеыивъвжжвазбисвьыэгзакжажэюэфессбсхъиэпйьь
юйяаоахжуушждфьхекбгкйфьчлжкухэшвтрпанрччомщыфемзъмфшбонхцсци
чюфиажяббнжжкодьгьйънжйяжцбжсбнёэьлрщшверфчмаьбкпчжюлдаэуьхшг
иыешфшюмщберябшррицбкиуёоёйжчкчёшютхчвэсыцойпуэьёкфдщпжфлщсё
ащрйжчкенглйдитыбюажтуугллщеаёзивмчнбмфшофжймуьдпфимтчрьёэфуое
рзпфечпёхнрнэпюмнъльхпалшюььйцмозьвиуюкбнтшьтыъюйцтгптфшйбмхъц
ыхэибесущцбнвюэрётортдкчёийщаьлъёёвзсззептаъдшьрючьйущчпояффтодяу
гвгвэьвпдавдсжтчкзусьхыунирююйдядамзшжыинжзышнуфыхжоеюдцг

6 Список используемой литературы

1. Частотный анализ — https://en.wikipedia.org/wiki/Frequency_analysis
2. Документация библиотеки Faker — https://faker.readthedocs.io/en/master/