

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-308Б-21

Студент(ка): Т. Ж. Караев

Преподаватель: А. В. Борисов

Оценка:

Дата: 22.05.2025

Москва, 2025

ОГЛАВЛЕНИЕ

1	Тема	3
2	Задание	3
3	Теория	4
4	Ход лабораторной работы.....	5
5	Выводы.....	9

1 Тема

Стилистический анализ 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

Данная лабораторная работа посвящена простейшему статистическому анализу текстов с точки зрения криптографии. В основе сравнения лежит отношение совпадения символов в разных типов текстов к общему количеству символов. Основной задачей является исследование вероятности совпадений текстов в силу их структуры.

В работе используется простейший алгоритм сравнения текстов:

- 1) Выполнить попарное сравнение символов на одинаковых позициях в пределах общей длины текстов,
- 2) Вычислить отношение количества совпадений к длине большего текста.

Стоит сказать, что данный алгоритм чувствителен к порядку символов и его результат может сильно варьироваться в случае сдвигов текстов.

4 **Ход лабораторной работы**

Для генерации осмысленных текстов из определённого набора символов использовался YaGPT со следующим запросом: «Сгенерируй осмысленный текст, состоящий только из английских букв и пробела. Других символов в тексте быть не должно». Полученные тексты:

«The artificial intelligence revolution is transforming our world at an incredible pace. Machine learning algorithms are becoming smarter every day, helping businesses automate complex tasks and make better decisions. Companies are investing heavily in ai research to stay competitive in the rapidly changing market landscape. Data scientists and engineers work tirelessly to improve neural networks, creating systems that can understand human language, recognize patterns, and solve intricate problems. This technological advancement opens up new opportunities for innovation across various industries, from healthcare to entertainment. As ai continues to evolve, it brings both challenges and opportunities for society. We must adapt to this new reality while ensuring that technology benefits everyone fairly. The future of artificial intelligence holds endless possibilities for those who embrace change and are willing to learn.»

«The sun rises in the east creating beautiful colors in the sky. Birds sing melodious songs as they fly from tree to tree. A gentle breeze blows through the tall grass making it dance in rhythm. Children play happily in the backyard chasing butterflies and laughing together. Their parents watch with pride as they create lasting memories. The air smells fresh and clean after the recent rain. Flowers bloom everywhere filling the garden with vibrant colors. Bees work hard collecting nectar to make honey for their hive. The day progresses

peacefully with nature performing its daily wonders. Evening approaches bringing a cooler temperature and a calming atmosphere. Families gather for dinner sharing stories of their day. Stars begin to appear one by one in the darkening sky. Night falls softly covering everything in a peaceful darkness. Crickets start their nightly chorus while owls begin their hunt. The world rests quietly until the next sunrise.»

Тексты из случайных символов: английских букв, пробела и точки — генерировались при помощи стандартных модулей Python: *random* и *string*.

Тексты из случайных слов генерировались при помощи библиотеки *nltk*.

Для более точного результата было принято решение сгенерировать и сравнить тексты много раз, чтобы найти средние значения.

Код программы приведён в репозитории

https://github.com/SempaiTakoo/cryptography_labs, а также текстом ниже:

```
import random
import string

import nltk

def compare_texts(text1: str, text2: str) -> float:
    n = min(len(text1), len(text2))

    if n == 0:
        return 0.0

    matches = sum(
        1
        for char1, char2 in zip(text1, text2)
        if char1 == char2
    )
    return matches / max(len(text1), len(text2))

def generate_random_letters_text(total_chars: int) -> str:
    alphabet = string.ascii_letters + ' .'
    return ''.join(random.choices(alphabet, k=total_chars))
```

```

def generate_random_words_text(word_list: list[str], total_chars: int)
-> str:
    words = []
    text_len = 0

    while text_len < total_chars:
        word = random.choice(word_list)

        if text_len + len(word) + (1 if words else 0) > total_chars:
            break

        words.append(word)
        text_len += len(word) + (1 if words else 0)

    return ' '.join(words)

def get_random_natural_lang_words_text(total_chars: int) -> str:
    nltk.download('words', quiet=True)
    words = nltk.corpus.words.words()
    return generate_random_words_text(word_list=words,
total_chars=total_chars)

def get_text_from_file(path: str) -> str:
    with open(path, 'r', encoding='utf-8') as file:
        return file.read()

def get_avg(nums: list[float]) -> float:
    return sum(nums) / len(nums)

def main(text_filepath_1, text_filepath_2, iter_count: int):
    meaningful_text = get_text_from_file(text_filepath_1)
    another_meaningful_text = get_text_from_file(text_filepath_2)

    random_letters_text =
generate_random_letters_text(len(meaningful_text))
    random_words_text =
get_random_natural_lang_words_text(len(meaningful_text))

    all_results = [[], [], [], [], []]

    for _ in range(iter_count):
        all_results[0].append(compare_texts(meaningful_text,
another_meaningful_text))
        all_results[1].append(compare_texts(meaningful_text,
random_letters_text))
        all_results[2].append(compare_texts(meaningful_text,
random_words_text))
        all_results[3].append(compare_texts(
generate_random_letters_text(total_chars=1000),
generate_random_letters_text(total_chars=1000)
))
        all_results[4].append(compare_texts(
get_random_natural_lang_words_text(total_chars=1000),
get_random_natural_lang_words_text(total_chars=1000)
))

```

```

    ))

    print(f'''
        Сравнение:
        1) два осмысленных текста на естественном языке:
{get_avg(all_results[0])},
        2) осмысленный текст и текст из случайных букв:
{get_avg(all_results[1])},
        3) осмысленный текст и текст из случайных слов:
{get_avg(all_results[2])},
        4) два текста из случайных букв: {get_avg(all_results[3])},
        5) два текста из случайных слов: {get_avg(all_results[4])}.
    ''')

if __name__ == "__main__":
    TEXT_FILEPATH_1 = 'texts/yagpt_text_1.txt'
    TEXT_FILEPATH_2 = 'texts/yagpt_text_2.txt'

    main(TEXT_FILEPATH_1, TEXT_FILEPATH_2, iter_count=1000)

```

Вывод программы:

```
sempaitakoo@desktoptakoo:~/cryptography_labs/lab3/src$ uv run main.py
```

```

        Сравнение:
        1) два осмысленных текста на естественном языке:
0.08928571428571429,
        2) осмысленный текст и текст из случайных букв:
0.007502679528403001,
        3) осмысленный текст и текст из случайных слов:
0.0632368703108253,
        4) два текста из случайных букв: 0.018521999999999997,

        5) два текста из случайных слов: 0.05944968966971289.

```


5 Выводы

Подведём некоторые выводы для каждого сравнения.

1. Два осмысленных текста совпадали в среднем на 8,9%. Стоит отметить, что это наивысший результат среди всех пар текстов. Предполагаю, что дело в двух причинах: во-первых, слова в английском языке, как и в любом другом, имеют определённые закономерности в использовании букв и их сочетаний; во-вторых, текст на естественном языке имеет определённую структуру и в нём часто повторяются некоторые слова вроде местоимений, предлогов и т. д.
2. Осмысленный текст и текст из случайных букв совпадали в среднем на 0,8%. Это довольно низкий результат. Думаю, дело в случайном распределении букв во втором тексте, не имеющий никаких закономерностей.
3. Осмысленный текст и текст из случайных слов совпадали в среднем на 6,3%. Этот результат гораздо выше, чем предыдущий, но ниже, чем первый, в силу того, о чём я уже говорил: слова в языке составляют неслучайным образом и имеют повторяющиеся фрагменты, отчего шанс совпадения каких-то символов сильно увеличивается, однако сами слова подобраны случайно, поэтому структурности на уровне всего текста нет.
4. Два текста из случайных букв совпадали в среднем на 1,9%. Этот результат чуть выше, чем во втором случае. Скорее всего, дело в том, что оба текста составлены из одинакового алфавита и каждый символ имеет одинаковую вероятность быть использованным.
5. Два текста из случайных слов совпадали в среднем на 5,9%. Это результат схож с третьим по тем же самым причинам: слова в языке часто имеют повторяющиеся наборы символов в силу грамматики.

Для корректного сравнения тексты должны быть достаточно длинными, чтобы результаты были более взвешенными и усреднёнными. В случае коротких текстов имеет место случайное совпадение больших фрагментов.

6 Список используемой литературы

- https://en.wikipedia.org/wiki/Integer_factorization
- https://en.wikipedia.org/wiki/One-way_function
- https://docs.sympy.org/latest/modules/ntheory.html#sympy.ntheory.factor_.factorint