

Кафедра инженерной кибернетики

ОТЧЕТ

ПО

ЛАБОРАТОРНОЙ РАБОТЕ

«Исследование возможностей и оценка качества генерации русско-язычных текстов NLP-моделями «семейства» ruGPT-3.x»

учебная дисциплина «Методы искусственного интеллекта»

Студент: Прокуденко Кирилл Игоревич

Группа: БПМ-21-3

Преподаватель: Хонер П. Д.

Оценка: 5

Дата защиты: 30.10.24

2024 г.

Содержание

1.	Введение.....	3
2.	Описание задачи.....	3
3.	Выбранные средства для разработки ПО	3
4.	Исходные данные	3
5.	Результат	4
6.	Анализ полученных результатов.....	7
7.	Выводы по лабораторной работе.....	7
8.	Список использованных источников	8

1. Введение

В данной лабораторной работе целью является обрести навыки по организации взаимодействия на уровне программного кода между собственным программным обеспечением и программным обеспечением, реализующим любой из вариантов «семейства» NLP-моделей ruGPT-3.x, при решении задачи генерации связного русскоязычного текста на заданную тему.

Также необходимо выполнить оценку уровня качества генерации русскоязычного текста «семейства» NLP-моделей ruGPT-3.x, используя полученные результаты генерации текстов.

2. Описание задачи

Требуется разработать приложение на уровне демонстрационного прототипа, решающее задачу по генерации русскоязычного текста с использованием NLP-моделей ruGPT-3.x.

В качестве тем для текста нужно взять аннотацию КНИР в текущем семестре и текст произвольной длины, тематически связанный с выполняемой КНИР.

3. Выбранные средства для разработки ПО

Для выполнения задачи сначала была выбрана модель ruGPT-light, но генерируемый текст был слишком неосмысленным. Поэтому было решено сменить модель на ruGPT-medium.

Программа разрабатывалась в VS Code на языке Python. Для работы с запросами использовались библиотеки torch и numpy. Для визуального интерфейса была взята библиотека tkinter. Для работы с моделью использовалась библиотека transformers.

4. Исходные данные

В качестве исходных данных возьмём несколько текстов.

- **Аннотация КНИР: Тема КНИР:** Рекомендательная система фильмов и сериалов основанная на базе уже просмотренных.
- **Текст связанный с КНИР:** Алгоритм A* Так как задача состоит в расчёте пути внутри здания между двумя точками, то становится очевидным, что обычный алгоритм Дейкстры будет тратить много лишнего времени на обработку ВГ, очевидно не ведущих в сторону конечной точки. В связи с этим возникает потребность не тратить время на обработку вершин, отдаляющихся от конечной точки. Именно с этой целью будет использовать алгоритм A*. Принцип работы данного алгоритм заключается в том.

5. Результат

Во всех тестах дефолтными параметрами для модели будут следующие:

```
out = model.generate(  
    input_ids,  
    max_length=max_length, # 100  
    top_k=top_k, # 50  
    top_p=top_p, # 0.9  
    temperature=temperature, # 0.9  
    repetition_penalty=5.0,  
    no_repeat_ngram_size=no_repeat # 3  
)
```

Рис. 1 – Дефолтные параметры для генерации.

Рассмотрим примеры работы программы с различной конфигурацией параметров и входных данных.

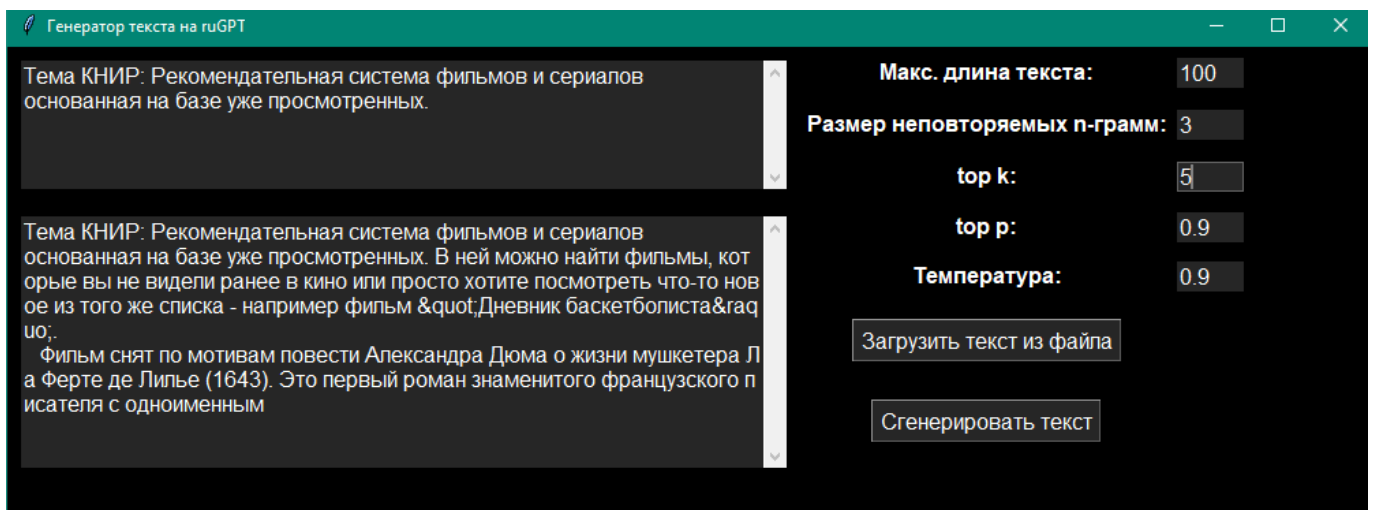


Рис. 2 – Результат запуска на тесте с аннотацией КНИР с дефолтными параметрами.

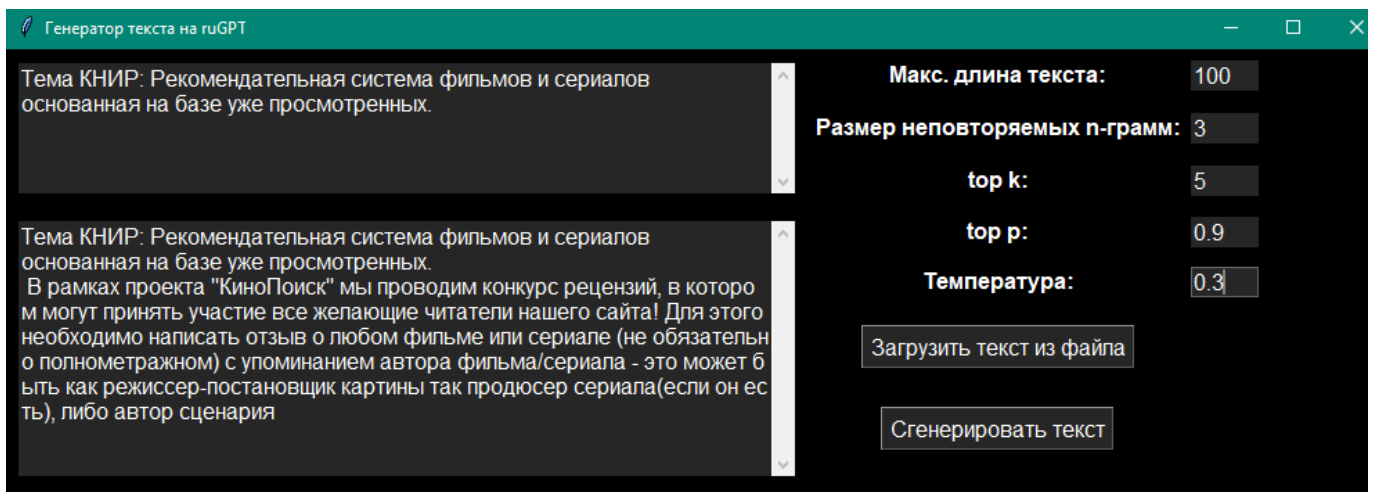


Рис. 3 – Результат запуска на тесте с аннотацией КНИР с другой температурой.

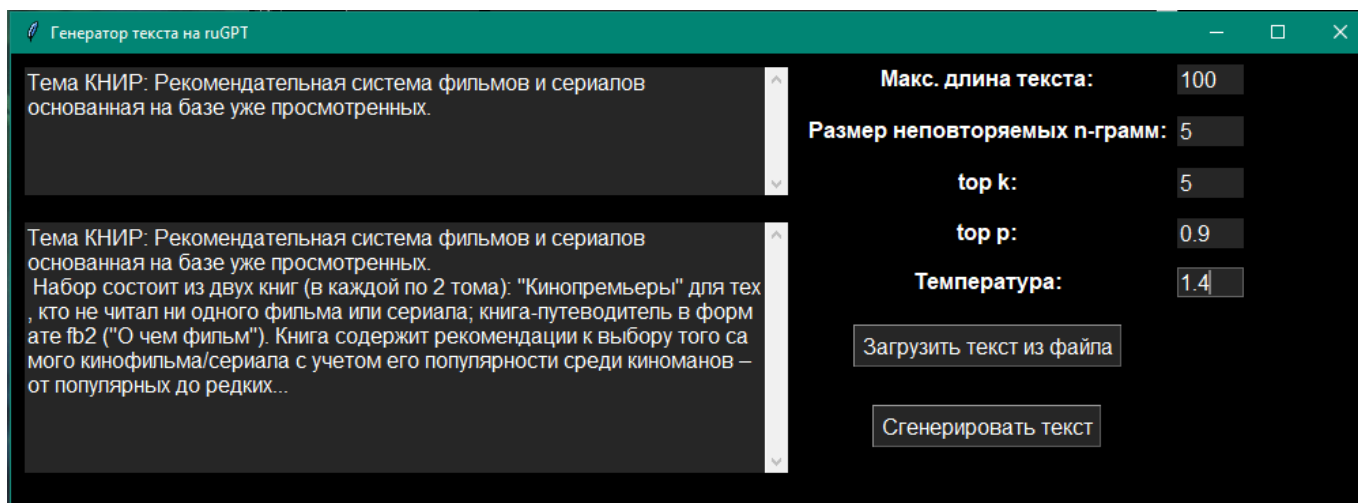


Рис. 4 – Результат запуска на тесте с аннотацией КНИР с другим значением n-грамм.

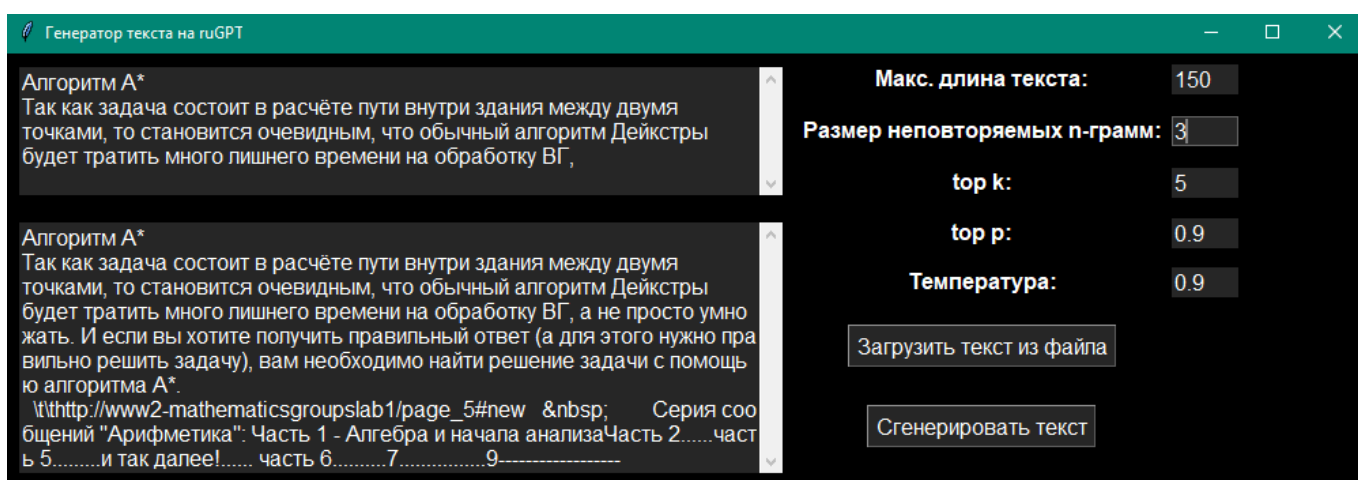


Рис. 5 – Результат запуска на тесте с примером текста из КНИР с дефолтными параметрами.

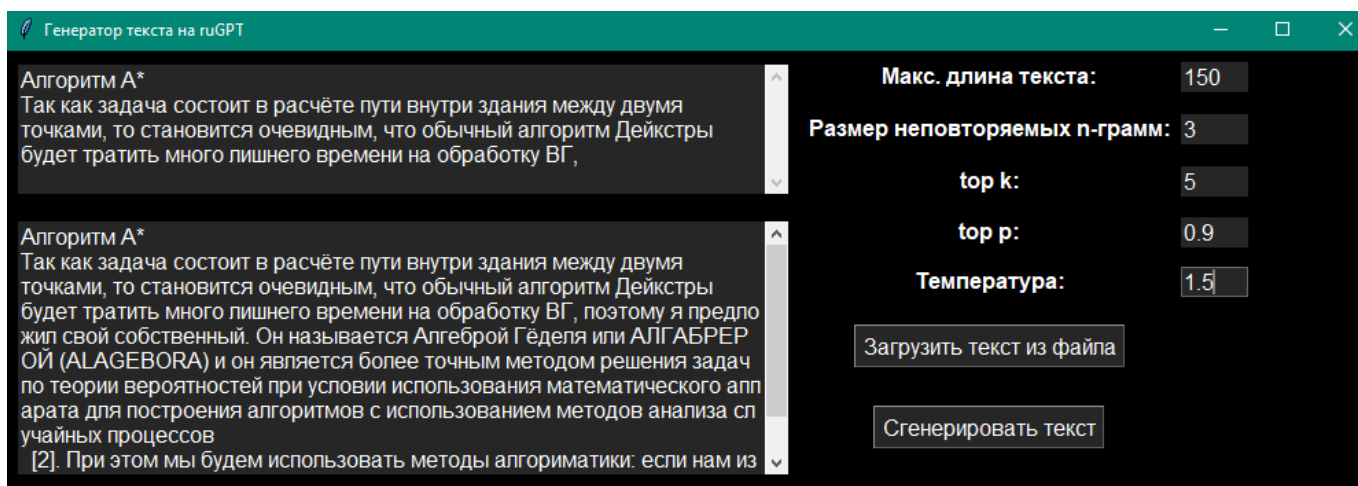


Рис. 6 – Результат запуска на тесте с примером текста из КНИР с другой температурой.

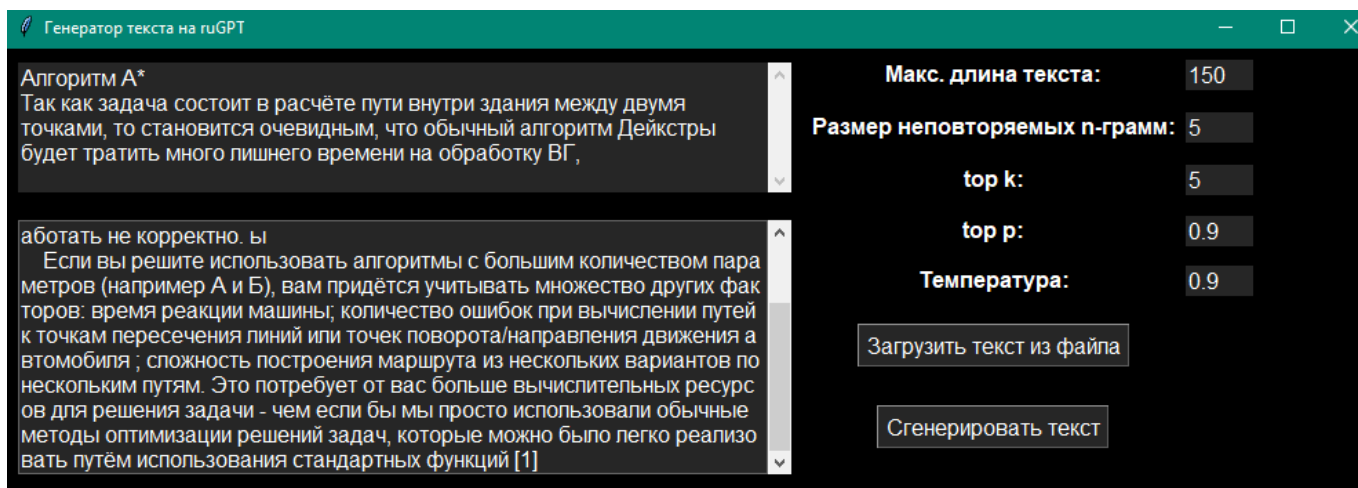


Рис. 7 – Результат запуска на тесте с примером текста из КНИР с другим значением n-грамм.

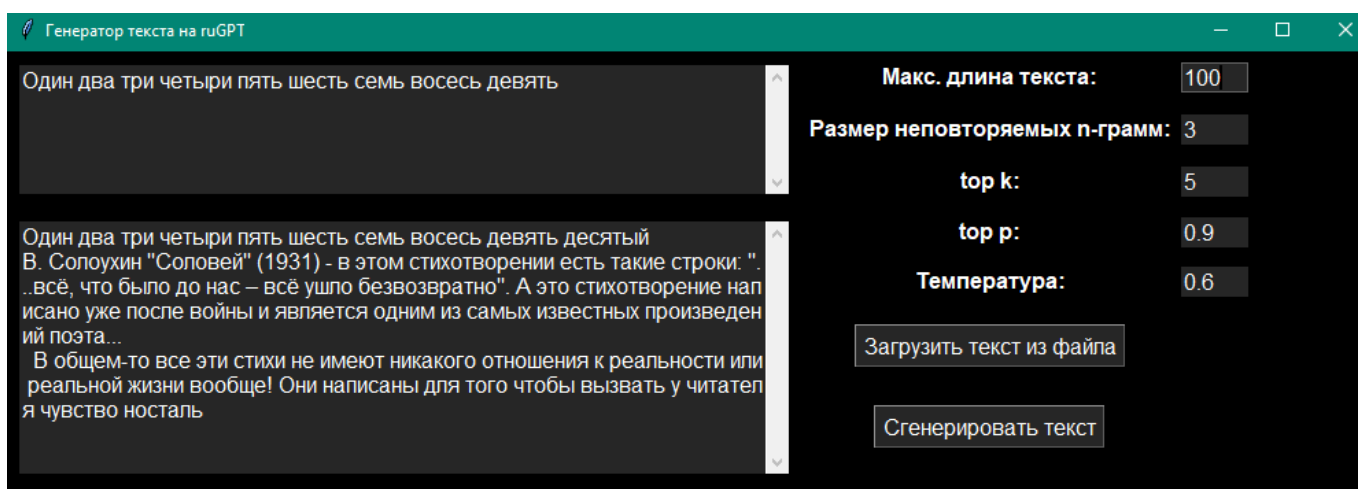


Рис. 8 – Результат запуска на тесте с последовательностью чисел.

6. Анализ полученных результатов

Как видно из тестов, во всех случаях сгенерированный текст примерно сохраняет тематику изначального текста, но смысла у полученного текста почти нет. При изменении параметров генерации результат не улучшается.

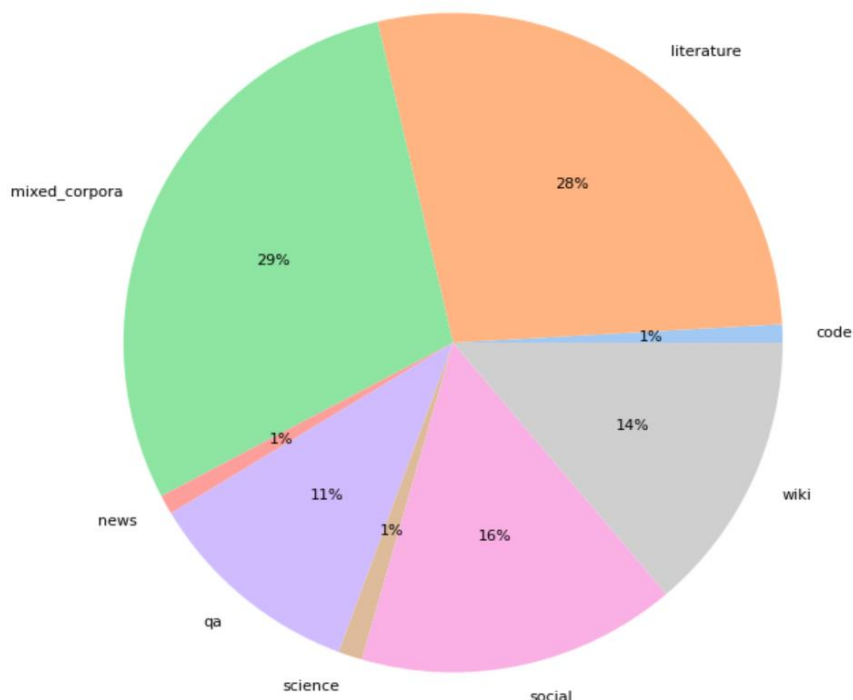


Рис. 9 – Структура датасета ruGPT-3.5.

В документации показан примерный состав датасета на котором обучалась модель (рис. 9). Из него наглядно видно, что при обучении использовался примерно 1% научных статей. Это объясняет очень низкий результат генерации текста для КНИР.

7. Выводы по лабораторной работе

По представленным результатам можно понять, что medium модель ruGPT-3.5 не стоит использовать для генерации текстов на научные темы. Модель часто выдает случайную, на вид, последовательность символов и очень быстро отходит от изначальной темы. Различные конфигурации параметров генерации не сильно улучшают получаемый результат.

Также стоит отметить неполную документацию модели, по которой почти невозможно понять за что отвечают параметры для генерации.

8. Список использованных источников

1. Документация библиотеки tkinter // Python Docs URL: <https://docs.python.org/3/library/tkinter.html> (дата обращения: 27.10.24).
2. Документация библиотеки Transformers // Hugging Face URL: <https://huggingface.co/docs/transformers/index> (дата обращения: 27.10.24).
3. Документация к моделям семейства ruGPT-3 // GitHub URL: <https://github.com/ai-forever/ru-gpts> (дата обращения: 27.10.24).