

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Отчёт о выполнении заданий

Дисциплина: «Статистические методы»

Студент: Леонов Владислав Денисович

Учебная группа: 224-322

Преподаватель 1: Филиппович Юрий Николаевич

Преподаватель 2: Воробьев Никита Григорьевич

Москва 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
ПРАКТИЧЕСКАЯ РАБОТА 1	4
ПРАКТИЧЕСКАЯ РАБОТА 2	5
ЛАБОРАТОРНАЯ РАБОТА 1	6
ЛАБОРАТОРНАЯ РАБОТА 2	8
ЛАБОРАТОРНАЯ РАБОТА 3	10
ЛАБОРАТОРНАЯ РАБОТА 4	11
ВЫВОДЫ.....	15
ЗАКЛЮЧЕНИЕ.....	16
ЛИТЕРАТУРА	17

ВВЕДЕНИЕ

В рамках разработки предметной области по созданию чат-бота и по работе с анализом больших объемов данных из открытых источников, были выбраны статьи, которые связаны с темой ВКР «Адаптивный интерфейс САПР на основе нейросетевого анализа пользовательских логов в задачах предсказания команд».

Для данной ВКР потребуются как статьи с пониманием инструментов, которые требуются для решения задачи, так и информация о работе используемых систем. Потому основным набором статей стали те, которые описывают работу ВКР.

Для прохождения курса необходимо выполнить задачи:

1. Собрать ЕЯ текстовые описания ПО.
2. Составить набор данных для обучения векторной модели.
3. Обучить модель Word2Vec, используя собранные данные.
4. Провести рубрицирование ЕЯ текстовых данных описания ПО.
5. Составить набор справочных данных, описывающих рубрики исследуемой ПО.
6. Создать автоматическую систему, предлагающую пользователю справочную информацию при запросе темы с использованием ЕЯ.

ПРАКТИЧЕСКАЯ РАБОТА 1

При сборе ЕЯ текстовых описаний ПО были выполнены следующие работы:

1. Были найдены 30 статей, соответствующих теме ВКР.
2. Был сохранен список названия статей и ссылок на них.
3. Было проведено ручное рубрицирование текстов.

При проведении ручного рубрицирования текстов было изучено их содержание. Далее для каждого текста были выделены 5-10 ключевых слов/словосочетаний, с помощью которых можно было составить поисковые образы. После этого статьи были разделены на 5 категорий.

Все данные ручного рубрицирования были собраны в виде таблицы со столбцами со следующими названиями «№», «Название статьи», «Ссылка на статью», «Дата обращения», «Ключевые слова», «Рубрика» (Рис. 1).

№	Название статьи	Ссылка на статью	Дата обращения	Ключевые слова	Рубрика
1	ЛИНГВИСТИЧЕСКАЯ ЭКСПЕРТОЛОГИЯ КАК	https://cyberleninka.ru/article/n/	15.04.2024	теоретическая	лингвистика
2	Проектирование быстрой программной	https://cyberleninka.ru/article/n/	15.04.2024	отбор	нейронные сети
3	Система контроля достоверности	https://cyberleninka.ru/article/n/	15.04.2024	текстовая информация /	N-грамма
4	N-граммы в лингвистике	https://cyberleninka.ru/article/n/	15.04.2024	n-грамма / порождающая	N-грамма
5	Информативность n-грамм в пределах	https://cyberleninka.ru/article/n/i	15.04.2024	информативность n-	N-грамма
6	Оценка эффективности использования	https://cyberleninka.ru/article/n/	15.04.2024	таджикский	N-грамма
7	РАСПОЗНАВАНИЕ СПАМ-СООБЩЕНИЙ С	https://cyberleninka.ru/article/n/	15.04.2024	машинное	кластеризация
8	АЛГОРИТМ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ В	https://cyberleninka.ru/article/n/	15.04.2024	кластерный	кластеризация
9	ПОДГОТОВКА ДАННЫХ ДЛЯ	https://cyberleninka.ru/article/n/	15.04.2024	данные / кластеризация	кластеризация
10	О поддержке принятия решения в	https://cyberleninka.ru/article/n/	15.04.2024	data	кластеризация
11	КЛАССИФИКАЦИЯ ТЕКСТОВ ПО	https://cyberleninka.ru/article/n/	15.04.2024	классификация / машинное	кластеризация
12	Математическая модель активности	https://cyberleninka.ru/article/n/	15.04.2024	удобство	адаптивный интерфейс
13	Адаптивный интерфейс пользователя	https://cyberleninka.ru/article/n/	15.04.2024	САПР / Адаптивный	адаптивный интерфейс
14	Задачи проектирования адаптивных	https://cyberleninka.ru/article/n/	15.04.2024	строительство / constructio	адаптивный интерфейс
15	ПРОГРАММНАЯ ИНСТРУМЕНТАЛЬНАЯ	https://cyberleninka.ru/article/n/	15.04.2024	адаптивный	адаптивный интерфейс
16	Построение тепловой карты на основе	https://cyberleninka.ru/article/n/	15.04.2024	математическая модель	адаптивный интерфейс
17	Метод обнаружения веб-роботов на	https://cyberleninka.ru/article/n/	15.04.2024	веб-	анализ пользовательских
18	Оценка степени удобства использования	https://cyberleninka.ru/article/n/	15.04.2024	логика	анализ пользовательских
19	НЕОБХОДИМОСТЬ UX-АНАЛИЗА	https://cyberleninka.ru/article/n/	15.04.2024	анализ / интерфейс	анализ пользовательских
20	ПРИМЕНЕНИЕ СТЕКА ТЕХНОЛОГИЙ ELK	https://cyberleninka.ru/article/n/	15.04.2024	облачные	анализ пользовательских
21	Компьютерное сопровождение	https://cyberleninka.ru/article/n/	15.04.2024	графическая	panoCAD

Рисунок 1 – Задание 1

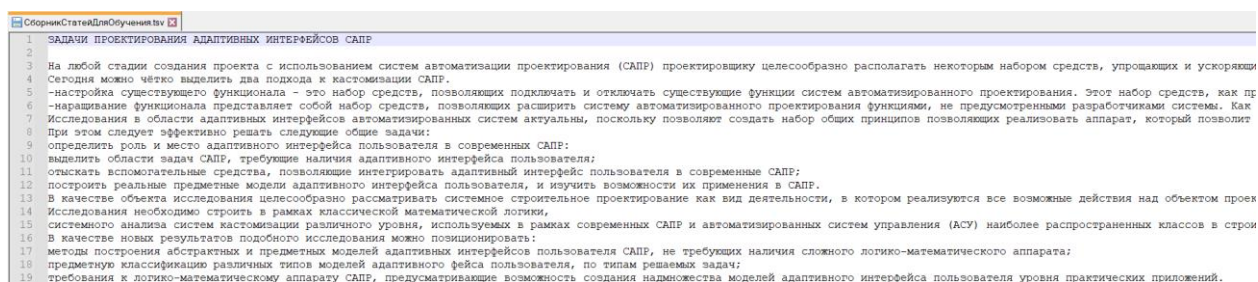
ПРАКТИЧЕСКАЯ РАБОТА 2

При создании обучающего датасета были выполнены следующие работы:

1. Была извлечена текстовая информация из найденных в первом задании статей и сохранена в формате docx.
2. Файлы docx были собраны в единый файл .txt.
3. Полученный файл был конвертирован в формат .tsv.

При извлечении текстовой информации из найденных в Задании 1 статей файлы были удалены лишние разделы.

В результате выполнения данного пункта был получен файл «*.tsv», являющийся обучающим датасетом (Рис. 2).



1	ЗАДАЧИ ПРОЕКТИРОВАНИЯ АДАПТИВНЫХ ИНТЕРФЕЙСОВ САПР
2	
3	На любой стадии создания проекта с использованием систем автоматизации проектирования (САПР) проектировщику целесообразно располагать некоторым набором средств, упрощающих и ускоряющих
4	Сегодня можно четко выделить два подхода к кастомизации САПР:
5	-настройка существующего функционала - это набор средств, позволяющих подключать и отключать существующие функции систем автоматизированного проектирования. Этот набор средств, как правило
6	-наращивание функционала представляет собой набор средств, позволяющих расширить систему автоматизированного проектирования функциями, не предусмотренными разработчиками системы. Как
7	Исследования в области адаптивных интерфейсов автоматизированных систем актуальны, поскольку позволяют создать набор общих принципов позволяющих реализовать аппарат, который позволит
8	При этом следует эффективно решать следующие общие задачи:
9	определить роль и место адаптивного интерфейса пользователя в современных САПР:
10	выделить области задач САПР, требующие наличия адаптивного интерфейса пользователя:
11	отыскать вспомогательные средства, позволяющие интегрировать адаптивный интерфейс пользователя в современные САПР;
12	построить реальные предметные модели адаптивного интерфейса пользователя, и изучить возможности их применения в САПР.
13	В качестве объекта исследования целесообразно рассматривать системное строительно-проектирование как вид деятельности, в котором реализуются все возможные действия над объектом проектирования
14	Исследования необходимо строить в рамках классической математической логики,
15	системного анализа систем кастомизации различного уровня, используемых в рамках современных САПР и автоматизированных систем управления (АСУ) наиболее распространенных классов в стро
16	В качестве новых результатов подобного исследования можно позиционировать:
17	методы построения абстрактных и предметных моделей адаптивных интерфейсов пользователя САПР, не требующих наличия сложного логико-математического аппарата;
18	предметную классификацию различных типов моделей адаптивного интерфейса пользователя, по типам решаемых задач;
19	требования к логико-математическому аппарату САПР, предусматривающие возможность создания надмножества моделей адаптивного интерфейса пользователя уровня практических приложений.

Рисунок 2 – Задание 2

ЛАБОРАТОРНАЯ РАБОТА 1

Для обучения модели Word2Vec был получен шаблон у преподавателя. Далее были установлены следующие библиотеки, используемые в шаблоне.

Сначала происходит импорт необходимых библиотек. Далее происходит загрузка данных из файла с датасетом «*.tsv» с предобработкой.

После обучения модели происходит вывод векторного представления слова «технологий» для проверки того, что обучения модели произошло корректно. Далее модель сохраняется в файл «*.model» (Рис. 3).

```
len(model.wv)=4252
model.wv['canp']=array([ 0.26980424, -0.10121949,  0.16700783,  0.01214969,  0.05541877,
 -0.1396053 ,  0.06081348, -0.04932012, -0.14497903,  0.03152451,
  0.11919741, -0.15339778, -0.03117944, -0.2065502 ,  0.02017382,
  0.2126367 , -0.3130389 , -0.00897831, -0.04641525,  0.2930966 ,
  0.2510757 ,  0.17898493, -0.1799792 , -0.2524953 , -0.1127089 ,
  0.1435391 , -0.1505722 ,  0.02676977, -0.00468068,  0.05758157,
  0.00561134,  0.06168572, -0.01625093, -0.05934908,  0.05400387,
  0.07485245, -0.11157027, -0.07853808, -0.1627722 ,  0.16300352,
  0.05668899,  0.05565174, -0.12341783,  0.034167 ,  0.15987813,
 -0.11302516, -0.08193211, -0.04216453,  0.01122172,  0.13285859,
  0.11159987, -0.05338966,  0.08866686,  0.07976852, -0.00033095,
 -0.10950099,  0.03918725, -0.24771166, -0.16595434,  0.15532398,
 -0.19986176,  0.05325808, -0.03048576, -0.04960163], dtype=float32)
model.corpus_count=3358
model.wv.most_similar('canp')=[('системы', 0.997297782897949), ('является', 0.9975845813751221), ('быть', 0.9975594878196716),
 ('основе', 0.9974576830863953), ('число', 0.997423529624939), ('выполнения', 0.9974172711372375), ('слов', 0.9973941445350647),
 ('можно', 0.9973477721214294), ('данных', 0.9973335862159729), ('либо', 0.9972543120384216)]
model.wv.most_similar('анализ')=[('пользователя', 0.9968934655189514), ('поскольку', 0.9968385100364685), ('правила', 0.9968237
280845642), ('задачи', 0.9966853857040405), ('который', 0.9965912103652954), ('можно', 0.9965049624443054), ('число', 0.9964288
724525452), ('также', 0.9964087009429932), ('используются', 0.9963939785957336), ('позволяет', 0.9963907599449158)]
model.wv.most_similar('кластеризация')=[('чения', 0.959916353225708), ('целесообразно', 0.9570011496543884), ('эффективности',
0.9561100602149963), ('оперативных', 0.9560425281524658), ('вместе', 0.9555233716964722), ('тренировочный', 0.9553735256195068),
('настроек', 0.9545378684997559), ('инфраструктуре', 0.9542561769485474), ('трудовых', 0.9538573622703552), ('проектов', 0.95
37014365196228)]
model.wv.most_similar('проектирование')=[('создания', 0.9934972524642944), ('решение', 0.9934178590774536), ('событий', 0.99338
85931968689), ('этом', 0.9931663274765015), ('поскольку', 0.9930448532104492), ('ресурсов', 0.9929442405700684), ('сети', 0.992
9049611091614), ('которых', 0.9928690791130066), ('анализа', 0.9928588271141052), ('который', 0.9927628040313721)]
model.wv.most_similar_to_given('canp', ['система', 'программа'])='система'
```

Рисунок 3 – Задание 3

Обучение модели Word2Vec происходило со следующими параметрами:

- «min_count» = 5 (слова, которые встречаются менее 5 раз, будут игнорироваться, причина столь низкого значения заключается в проблеме с текстами. В них встречается обилие слов, которые не относятся к темам работ, а сами же темы упоминаются не так часто. Потому возникла проблема с небольшим количеством упоминаний слов.);
- «window» = 2 (размер контекстного окна для обучения, максимальное расстояние между текущим словом и целевым словом);

- «vector_size» = 64 (размерность векторного представления слов);
- «alpha» = 0.03 (начальная скорость обучения);
- «negative» = 1 (количество отрицательных образцов, которые нужно сгенерировать для каждого положительного образца, при увеличении параметра происходило обильное просачивание прилагательных и служебных слов в финальную работу);
- «min_alpha» = 0.0007 (конечная скорость обучения);
- «sample» = 6e-5 (порог для снижения частоты частотных слов).

ЛАБОРАТОРНАЯ РАБОТА 2

Для автоматического рубрицирования датасет был разделен по текстам. На рисунке 4 представлен результат разделения датасета.

_1_programmnaya-instrumentalnaya-sistema-soz...	19.04.2024 23:30	Текстовый докум...	32 КБ
_2_raspoznavanie-spam-soobscheniy-s-ispolzovan...	19.04.2024 23:31	Текстовый докум...	20 КБ
_3_razrabotka-sistemy-raspoznavaniya-rechi-na-o...	19.04.2024 23:31	Текстовый докум...	32 КБ
_4_rol-bim-tehnologiy-v-organizatsii-i-tehnologii...	19.04.2024 23:31	Текстовый докум...	36 КБ
_5_sistema-kontrolya-dostovernosti-tekstovoy-inf...	19.04.2024 23:31	Текстовый докум...	47 КБ
_6_sistema-prediktivnogo-vvoda-kak-sredstvo-po...	19.04.2024 23:31	Текстовый докум...	13 КБ
_7_sravnitelnyy-analiz-sapr-na-primere-proektirov...	19.04.2024 23:19	Текстовый докум...	24 КБ
_8_adaptivnyy-interfeys-polzovatelya-sapr.txt	19.04.2024 23:31	Текстовый докум...	12 КБ
_9_algoritm-nechetkoy-klasterizatsii-v-tehnologii...	19.04.2024 23:31	Текстовый докум...	12 КБ
_10_arhitektura-modeli- lstm-dlya-prognozirovani...	19.04.2024 23:31	Текстовый докум...	7 КБ
_11_klassifikatsiya-tekstov-po-tonalnosti-metoda...	19.04.2024 23:31	Текстовый докум...	26 КБ
_12_kompyuternoe-soprovozhdenie-prepodavaniy...	19.04.2024 23:30	Текстовый докум...	16 КБ
_13_matematicheskaya-model-aktivnosti-polzovat...	19.04.2024 23:30	Текстовый докум...	37 КБ
_14_metod-obnaruzheniya-veb-robotov-na-osnov...	19.04.2024 23:30	Текстовый докум...	16 КБ

Рисунок 4 – Результат разделения датасета по текстам

Далее был разработан код, который производит автоматическое рубрицирование текстов. На рисунке 5 представлен результат работы кода.


```

resources/_1_programmnaya-instrumentalnaya-sistema-sozdaniya-adaptivnyh-polzovatelskih-interfeysov.txt
resources/_20_otsenka-effektivnosti-ispolzovaniya-unigramm-pri-identifikatsii-teksta.txt
resources/_21_podgotovka-dannyh-dlya-klasterizatsii-sobytiy-v-zhurnalah-informatsionnoy-bezopasnosti.txt
resources/_22_postroenie-teplovoy-karty-na-osnove-tochechnykh-dannyh-ob-aktivnosti-polzovatelya-prilozheniya.txt
resources/_23_prediktivnyy-vvod-teksta-na-osnove-fakturnoy-modeli-yazyka.txt
resources/_24_primenenie-steka-tehnologiy-elk-dlya-sbora-i-analiza-sistemnykh-zhurnalov-sobytiy.txt
resources/_25_problema-razrabotki-vr-trenazherov-sborki-razborki-i-variant-vysokoproizvoditelnogo-resheniya-na-baze-tehnologii-
vr-concept.txt
resources/_26_proektirovanie-bystroy-programmnoy-realizatsii-spetsializirovannoy-neyrosetvoy-arhitektury-s-razrezhennymi-svyaz
yami.txt
resources/_27_informativnost-n-gramm-v-predelah-slova.txt
resources/_28_lingvisticheskaya-ekspertologiya-kak-napravlenie-teoreticheskoy-i-prikladnoy-lingvistiki.txt
resources/_29_otsenka-stepeni-udobstva-ispolzovaniya-polzovatelskih-interfeysov-v-logike-taylov.txt
resources/_2_raspoznavanie-spam-soobscheniy-s-ispolzovaniem-metodov-mashinogo-obucheniya.txt
resources/_30_zadachi-proektirovaniya-adaptivnyh-interfeysov-sapr.txt
resources/_3_razrabotka-sistemy-raspoznavaniya-rechi-na-osnove-skrytykh-markovskikh-modeley-otdelnykh-slov.txt
resources/_4_rol-bim-tehnologiy-v-organizatsii-i-tehnologii-stroitelstva.txt
resources/_5_sistema-kontrolya-dostovernosti-tekstovoy-informatsii-na-osnove-n-grammnykh-parsingovykh-modeley.txt
resources/_6_sistema-prediktivnogo-vvoda-kak-sredstvo-povysheniya-effektivnosti-nabora-teksta.txt
resources/_7_sravnitelnyy-analiz-sapr-na-primere-proektirovaniya-trehmernoy-modeli-korpusa-sudna.txt
resources/_8_adaptivnyy-interfeys-polzovatelya-sapr.txt
resources/_9_algoritm-nechetkoy-klasterizatsii-v-tehnologii-data-mining.txt
-0.23017436893667703
[('продуктом', 0.4222221076488495), ('операторы', 0.35737529397010803), ('независимо', 0.25844326615333557), ('документации', 0
.2528797388076782), ('вариантом', 0.20255087316036224), ('неравномерных', 0.19761013984680176), ('денеие', 0.19320088624954224)]
-0.21089786535124283
[('продуктом', 0.4222221076488495), ('операторы', 0.35737529397010803), ('независимо', 0.25844326615333557), ('документации', 0
.2528797388076782), ('вариантом', 0.20255087316036224), ('неравномерных', 0.19761013984680176), ('денеие', 0.19320088624954224)]
-0.19413451067580323
[('продуктом', 0.4222221076488495), ('операторы', 0.35737529397010803), ('независимо', 0.25844326615333557), ('документации', 0
.2528797388076782), ('вариантом', 0.20255087316036224), ('неравномерных', 0.19761013984680176), ('денеие', 0.19320088624954224)]
PS C:\Users\Surflav\Downloads\02>

```

Рисунок 5 – Результаты автоматического рубрицирования

ЛАБОРАТОРНАЯ РАБОТА 3

Для выполнения задания выбрал описание тем, фигурирующих в статье, входящих в категорию. Таблица состоит из вектора, категории, выделенной из статей, а также описания категории (Рис. 6).

Вектор	Категория	Описание
0.345831524	анализ	процесс разбора, изучения и выявления закономерностей в данных или явлениях
0.317591989	проектирование	создание плана, макета или модели чего-либо перед его фактической реализацией
0.338325244	предсказание	использование данных и статистических методов для определения будущих событий или значений

Рисунок 6 – Задание 5

ЛАБОРАТОРНАЯ РАБОТА 4

В ходе выполнения задания 6 получилось 2 таблицы со связью один ко многим. На основе таблиц была создана база данных, структура базы данных представлена на рисунке 7.

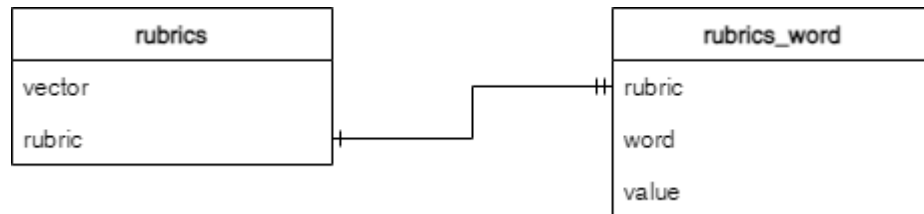


Рисунок 7 – Структура базы данных

База состоит из двух таблиц, в одной хранятся наименования рубрик и их векторные значения (таблица **rubrics**), и таблицы **rubrics_word**, которая хранит в себе основные определения каждой из рубрик. Заполнение таблиц **rubrics** и **rubrics_word** представлено на рисунках 8 и 9.

	vector	rubric
	Фильтр	Фильтр
1	-0.229328284021437	анализ
2	-0.629328284021437	проектирование
3	-0.129328284021437	предсказание

Рисунок 8 – Заполнение таблицы **rubrics**

	rubric	word	value
	Фильтр	Фильтр	Фильтр
1	предсказание	RNN	рекуррентная нейронная сеть ...
2	предсказание	LSTM	сеть с долговременной и ...
3	предсказание	Предсказание команд	предсказание команд может помоч...
4	проектирование	Проектирование	процесс создания проекта и его ...
5	проектирование	AutoCAD	это пакет программ для точного ...
6	проектирование	papoCAD	платформа для проектирования и ...
7	анализ	N-грамма	это фразы из двух или трех слов, ...
8	анализ	Кластеризация	это разделение большой группы ...
9	анализ	Анализ пользовательских логов	позволяет выявить закономерност...
10	проектирование	САПР	это автоматизированная система, ...

Рисунок 9 – Заполнение таблицы rubrics_word

С использованием созданной таблицы был реализован алгоритм, который выполняет действия:

1. Пользователь вводит запрос;
2. Система разбивает запрос на массив слов;
3. Для каждого отдельного слова создается нормализованный вектор из обученной модели;
4. Рассчитывается среднее значение вектора запроса;
5. Из таблицы rubric выбирается поле с наиболее близким числовым значением категории;
6. Пользователю выводится информация о теме, которое система смогла найти в его запросе, если же определение не было найдено, то выводится случайная информация по теме.
7. Система запрашивает у пользователя подтверждение правильности выбора рубрики
8. Если рубрика определена верно, то пользователю предлагается вновь сделать запрос
9. Если же рубрика была определена неверно, то к значению вектора рубрики, указанной пользователем как правильная, прибавляется или

вычитается, в зависимости от значения вектора введенного запроса, число равное корню квадратному модуля разности вектора запроса и вектора рубрики.

На рисунке 10 представлен интерфейс программы для взаимодействия с чат-ботом.

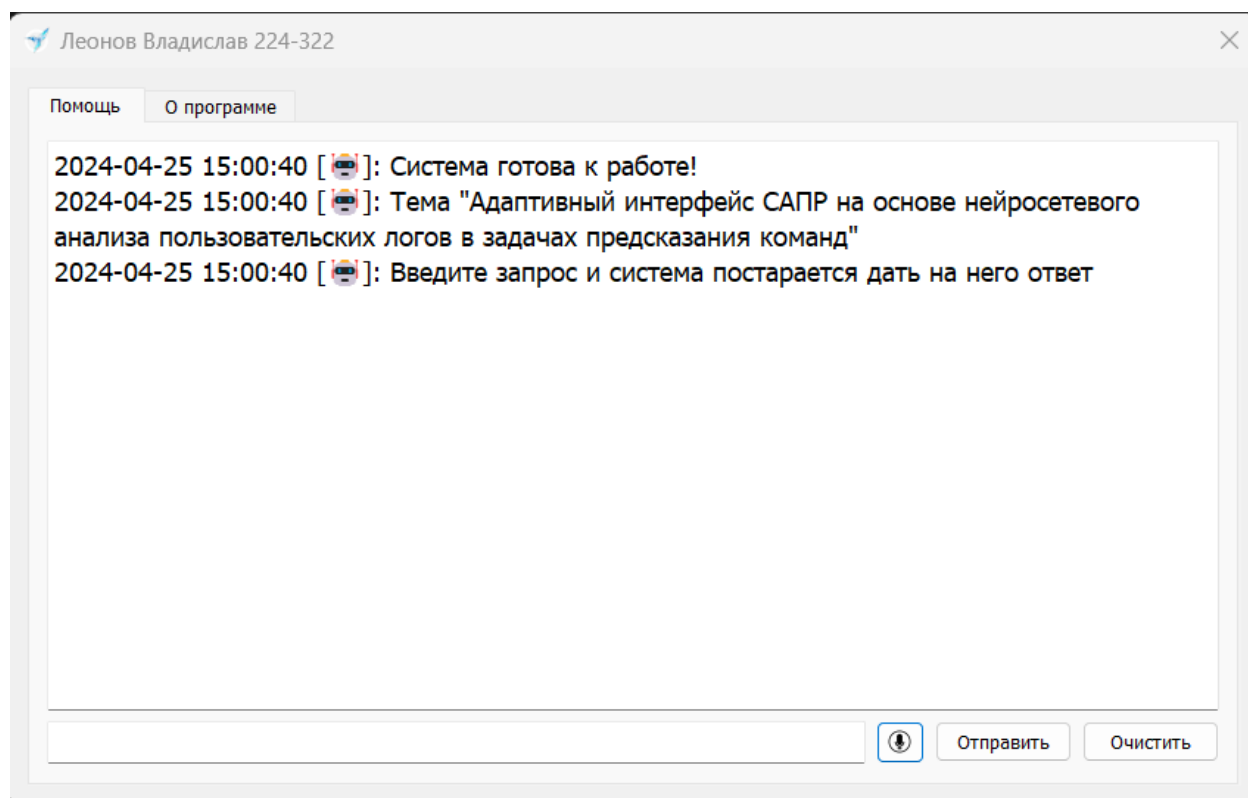


Рисунок 10 – Интерфейс программы

На рисунке 11 представлен ответ на вопрос пользователя.

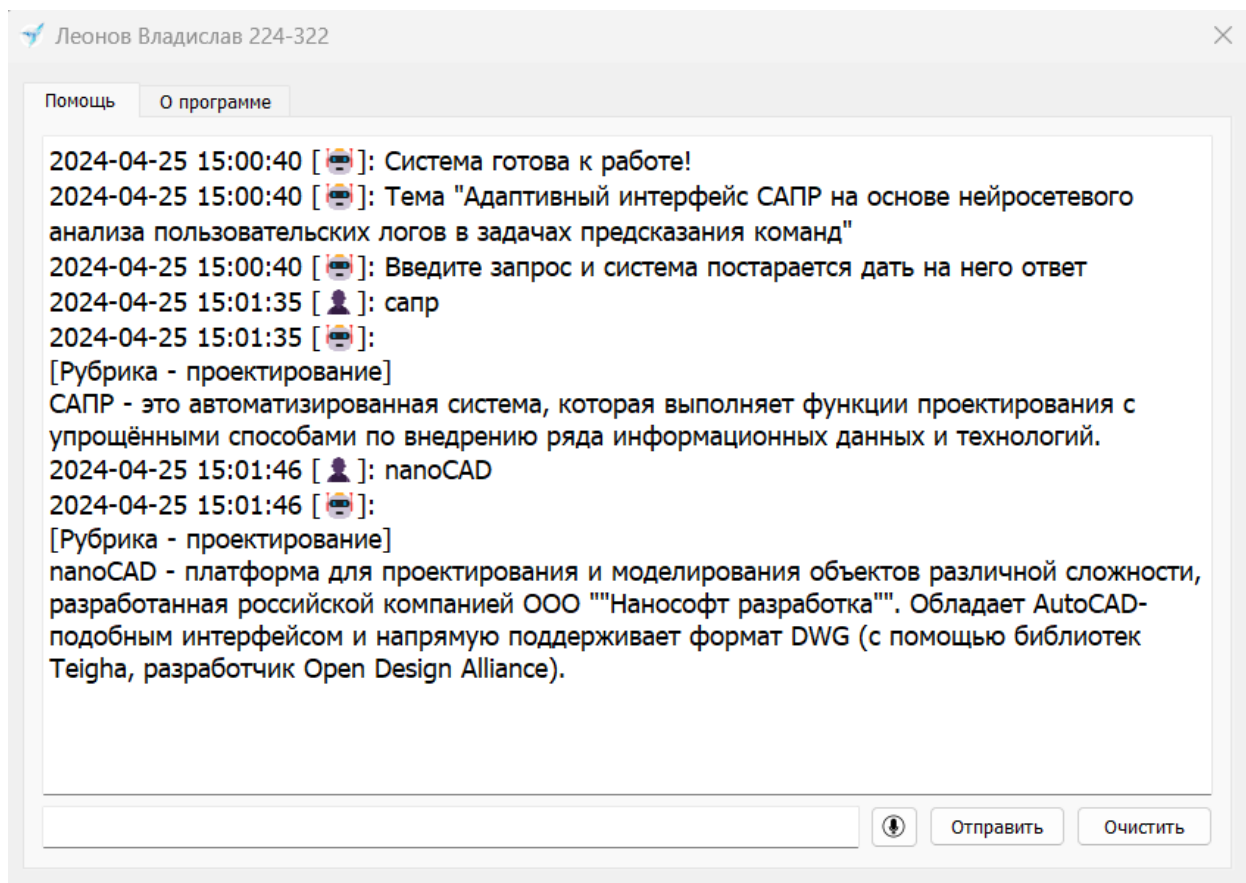


Рисунок 11 – Ответ на вопрос пользователя

ВЫВОДЫ

Рассматривая применение векторного решения для рубрицирования текстов можно прийти к выводу, что подобный подход является слишком примитивным для рубрицирования российских текстов. Проблема в том, что русский язык пестрит большим количеством средств выразительности, что приводит к увеличению работы по очищению текста до работы с ним векторной модели. Также играет роль и то, как люди пишут научные работы. Так, например, довольно редко упоминается сам объект исследования, что приводит к проблеме с его выделением по вектору без снижения значения `min_count`. Тем самым, более выгодным решением для подобного решения будет многослойная нейронная сеть.

ЗАКЛЮЧЕНИЕ

Были успешно достигнуты поставленные цели и задачи, связанные с обработкой и использованием естественно-языковых текстовых данных. В процессе проекта были выполнены следующие ключевые этапы:

- изучение материалов и освоение необходимых методов и технологий, связанных с анализом и обработкой текстовых данных. Это включало в себя изучение лекционных курсов, выполнение практических заданий и использование специального программного обеспечения;
- сбор текстовых описаний программного обеспечения (ПО) на естественном языке, которые послужили основой для дальнейших исследований;
- создание датасета для обучения векторной модели, что позволило улучшить качество анализа текстов;
- обучение модели word2vec на основе собранных данных, что сделало возможным получение векторных представлений для слов и текстов, улучшая качество анализа текстовых данных;
- рубрицирование текстовых описаний ПО, что позволило организовать информацию в систему категорий и облегчило поиск и доступ к данным;
- создание автоматической системы, предоставляющей пользователям справочную информацию на основе запросов с использованием естественно-языковых текстовых данных.

Таким образом, были приобретены навыки по работе с текстовыми данными и их автоматическому рубрицированию.

ЛИТЕРАТУРА

1. Автоматическая рубрикация текстов: методы и проблемы – URL: <https://cyberleninka.ru/article/n/avtomaticheskaya-rubrikatsiya-tekstov-metody-i-problemy> (Дата обращения 15.04.2024);
2. Автоматическое тематическое рубрицирование сообщений средств массовой информации на основе применения технологии нейронных сетей – URL: <http://engineering-science.ru/doc/56926.html> (Дата обращения 05.04.2024);
3. Методы и алгоритмы рубрикации текстов – URL: <https://cyberleninka.ru/article/n/metody-i-algoritmy-rubrikatsii-tekstov> (Дата обращения 15.04.2024);
4. Vector Space Model для семантической классификации текстов – URL: <https://habr.com/ru/sandbox/18635/> (15.04.2024);
5. Векторное представление слов – URL: https://neerc.ifmo.ru/wiki/index.php?title=%D0%92%D0%B5%D0%BA%D1%82%D0%BE%D1%80%D0%BD%D0%BE%D0%B5_%D0%BF%D1%80%D0%B5%D0%B4%D1%81%D1%82%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D1%81%D0%BB%D0%BE%D0%B2 (Дата обращения 15.04.2024);
6. Word2vec в картинках – URL: <https://habr.com/ru/articles/446530/> (Дата обращения 15.04.2024);
7. Word2Vec: как работать с векторными представлениями слов – URL: <https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obucheniya/> (Дата обращения 15.04.2024);
8. Семантика и технология Word2Vec – URL: <https://habr.com/ru/articles/585838/> (Дата обращения 15.04.2024).