МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ШКОЛА КОМПЬЮТЕРНЫХ НАУК

Кафедра программного обеспечения

РЕКОМЕНДОВАННО К ЗАЩИТЕ В ГЭК

Заведующий кафедрой

Кандидат наук,

М.С. Воробьева

\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

бакалавра

Разработка виртуального помощника для построения индивидуально образовательной траектории

02.03.03. Математическое обеспечение и администрирование информационных систем

|  |  |  |
| --- | --- | --- |
| Выполнили работу  (групповой проект)  студенты 4 курса  очной формы обучения | *(Подпись)*  *(Подпись)* | Бессонова Софья Денисовна  Кондратьев Илья Владимирович |
| Руководитель  (к.н.) | *(Подпись)* | Воробьева Марина Сергеевна |

Тюмень

2024

**Оглавление**

[**1. Введение** 3](#_Toc185534519)

[**2. Теоретическое обоснование проекта** 6](#_Toc185534520)

[**3. Используемые технологии** 13](#_Toc185534521)

[**4. Реализация проекта** 15](#_Toc185534522)

[**5. Пользовательский интерфейс** 19](#_Toc185534523)

[**6. Тестирование** 21](#_Toc185534524)

[**7. Заключение** 23](#_Toc185534525)

[**8. Список литературы** 24](#_Toc185534526)

[**Приложение. Сравнения моделей для создания эмбеддингов** 26](#_Toc185534527)

# **1. Введение**

Каждый год студенты Тюменского государственного университета выбирают для себя элективные дисциплины из большого списка более чем в 700 элективов. Элективы - это дисциплины по выбору из неосновной предметной области дисциплин, которые студент может выбрать самостоятельно[1]. Элективы охватывают различные области знаний, начиная с разработки интерфейсов веб-приложений и заканчивая изучением особенностей игр народов России. Студенты могут выбрать любой курс, который им интересен, чтобы расширить свои знания и навыки в выбранной ими сфере.

Сейчас при выборе курса студенты обращаются к сервису сбора отзывов на элективные дисциплины, разработанного студентами ТюмГУ, - «Отзывусу» [2]. Однако не смотря на наличие отзывов на некоторые элективы процесс их выбора остаётся трудоёмким и запутанным для студента, который не может определиться с тем какая дисциплина соответствует его интересам. Более того, для каждого семестра составляется свой список элективов на выбор, в который могут входить новые элективы, у которых студенты не смогут найти отзывов ни на Отзывусе ни среди своих знакомых, так как никто в университете ранее не изучал их. Студенту, делающему выбор, остаётся только самостоятельно изучать карточку МУП-а в Modeus в период открытия меню выбора дисциплин или вовсе отказаться от электива из-за неуверенности в нём и отсутствия каких-либо рекомендаций от других студентов.

Проблема: выбор среди большого количества дисциплин может оказаться запутанным процессом для студентов, особенно для первокурсников, которые впервые сталкиваются с выбором дисциплин. Отсутствие персонализированных рекомендаций приводит к случайному выбору – студенты в основном полагаются на отзывы своих знакомых и выбирают элективы, несоответствующие их интересам из-за нежелания самостоятельно разбираться в большом списке актуальных элективов для выбора. Таким образом, студенты строят индивидуальную образовательную траекторию (ИОТ), полагаясь на чужой опыт, а не на собственные предпочтения.

​ Цель: создание интеллектуального Telegram-бота, который анализирует пройденные студентом элективы и его предпочтения для помощи в выборе элективных курсов. Чат-бот позволит студентам быстро подбирать элективы по введённому запросу, что упростит студентам процесс выбора учебных дисциплин, соответствующих их интересам.

Идея решения: Telegram-бот, который помогает студентам быстро подобрать элективы по запросу, основываясь на данных об их интересах и пройденных дисциплинах.

Особенности решения:

* Интеграция с платформой "Отзывус" — ссылки на реальные отзывы студентов о курсах.
* Персонализация — учитываются пройденные дисциплины (для старших курсов) и интересы студентов для релевантных рекомендаций.
* Удобство и доступность — доступ из привычного кроссплатформенного мессенджера Telegram, с мгновенным доступом к информации.

Задачи проекта:

1. Анализ требований пользователей
2. Исследование существующих решений
3. Сбор данных
   1. Сбор данных о пройденных студентами курсах
   2. Сбор данных об элективах из Modeus и Отзывуса
4. Изучение моделей, которые могут быть использованы для формирования рекомендаций исходя из входных данных
5. Разработка алгоритмов персонализированных рекомендаций
   1. На основе пройденных дисциплин
   2. На основе предпочтений
6. Разработка чат-бота
7. Тестирование проекта

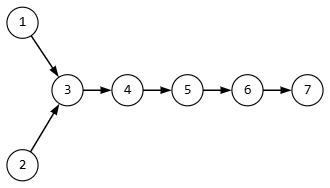


Рис. 1 – Приоритезация задач в виде сетевого графика

# **2. Теоретическое обоснование проекта**

Современные образовательные учреждения сталкиваются с проблемой персонализации учебного процесса. С увеличением числа курсов, элективных дисциплин и образовательных программ студенты часто испытывают трудности в выборе наиболее подходящих учебных траекторий. С другой стороны, преподавателям и администраторам сложно анализировать большие объёмы данных о предпочтениях и результатах студентов.

Рекомендательные системы, широко используемые в коммерческих продуктах (например, для рекомендаций фильмов, музыки или товаров), могут быть адаптированы для образовательных целей. Такие системы не только помогают учащимся находить курсы, наиболее соответствующие их интересам и целям, но и повышают эффективность образовательного процесса в целом [3].

В проекте, разработанном в Санкт-Петербургском государственном университете, рассматривается создание рекомендательной системы для поиска наиболее подходящих вузов для абитуриентов [4]. Для подбора релевантных вузов используется модель контентной фильтрации. Однако для побора рекомендаций также рассматривалась модель коллаборативной фильтрации, но не была реализована. Также существует гибридная система рекоммендаций, которая сочетает в себе как контентную, так и коллаборативную фильтрации. Совмещение в себе нескольких моделей помогает устранить недостатки каждой, делая прогноз более точным. Большинство рекомендательных систем использует именно гибридную модель [5].

Проект по подбору элективных дисциплин направлен на разработку системы рекомендаций образовательных курсов на основе данных о предпочтениях студентов и содержании курсов. В его основе лежат два ключевых подхода: контентная фильтрация и коллаборативная фильтрация.

**Контентная фильтрация**

Применение контекста – самое перспективное направление в области рекомендательных систем [6]. Контентная фильтрация основывается на анализе характеристик элементов (в данном случае курсов и элективов). Главной задачей такого подхода является создание текстовых представлений (эмбеддингов) для каждого элемента и последующее их сравнение с интересами пользователя.

В проекте для реализации контентной фильтрации применяются следующие методы:

1. Обработка текста на русском языке. С помощью библиотеки Natasha текстовые описания курсов (название, образовательный результат, полное описание) сегментируются и нормализуются. Убираются лишние символы, текст приводится к единой форме, удаляются слова, не имеющие смысловой нагрузки. Это позволяет улучшить качество анализа и устранить неоднозначность.
2. Построение эмбеддингов.Языковая модель paraphrase-multilingual-mpnet-base-v2 из библиотеки Sentence-Transformers используется для преобразования текста в эмбеддинги – векторные представления, отражающие семантическое значение текста. Данная модель поддерживает русский язык и хорошо справляется с задачами кластеризации и поиска похожих элементов.
3. Сравнение курсов.Полученные эмбеддинги используются для определения семантической близости текстов. Например, если студент уже прошёл курс, связанный с программированием на Python, система сможет предложить ему курсы, ориентированные на алгоритмы или анализ данных, основываясь на схожести их описаний.

Преимущество контентной фильтрации заключается в том, что она не требует данных о других пользователях и подходит даже для новых элементов, которые ещё не получили оценок.

**Коллаборативная фильтрация**

Коллаборативная фильтрация основывается на анализе поведения пользователей. Главная идея заключается в том, что если несколько студентов имеют схожие предпочтения, то курсы, которые нравятся одному из них, могут быть рекомендованы другим.

Для реализации коллаборативной фильтрации используется метод SVD:

1. Сбор данных о предпочтениях студентов**.** На основании информации о ранее пройденных элективах создаётся бинарная матрица, где строки соответствуют студентам, а столбцы – названиям курсов. Значение в ячейке равно 1, если студент проходил данный курс, и 0 – если не проходил. Т
2. Расстояние Левенштейна. Так как студент-пользователь будет вводить элективы самостоятельно, нужно учитывать, что некоторые названия будут введены некорректно. Расстояние Левенштейна – это метрика, которая показывает, сколько операций вставки, удаления или замены символов нужно для преобразования одной строки в другую Расстояние по Левенштейну между двумя словами определяется как минимальное количество односимвольных операций, необходимых для преобразования из одного слова в другое.
3. SVD-разложение**.** Метод SVD позволяет разложить матрицу на три компонента: матрицу студентов, матрицу курсов и матрицу весов. Это упрощает выявление скрытых закономерностей в данных. Например, метод находит латентные факторы, которые определяют предпочтения студентов, такие как уровень сложности курсов, их тематика и т.д.
4. Предсказание рейтингов.На основе разложения можно предсказывать, насколько интересен или полезен студенту определённый курс, даже если он ещё не посещал его. Это позволяет персонализировать рекомендации для каждого пользователя.

Коллаборативная фильтрация эффективна в том случае, если имеется достаточное количество данных о взаимодействиях пользователей с курсами. Однако её главный недостаток – невозможность рекомендовать новые курсы, которые ещё не получили оценок (проблема "холодного старта").

Для получения максимально точных рекомендаций проект использует оба подхода. Контентная фильтрация компенсирует недостатки коллаборативной фильтрации в условиях "холодного старта", позволяя анализировать новые курсы. В свою очередь, коллаборативная фильтрация учитывает предпочтения студентов и выявляет скрытые связи, которые невозможно определить только по содержанию курсов.

Чат-боты работают круглосуточно и доступны через популярные мессенджеры, что делает их удобными для пользователей. Это особенно актуально для молодежи, которая активно использует мобильные приложения для общения [7]. Согласно статистике 73% россиян активно используют чат-боты в повседневной жизни, а самая высокая осведомленность о чат-ботах оказалась у респондентов 25–34 лет (96%) [8]. Исходя из этого было принято решение обернуть систему в Telegram-бот.

Для удобства взаимодействия с системой создан Telegram-бот, который:

1. Получает запросы от студентов (например, их интересы или перечень ранее пройденных курсов).
2. Обрабатывает запросы с помощью заранее обученных моделей.
3. Возвращает персонализированные рекомендации.

Telegram-бот обеспечивает доступность системы для широкой аудитории, позволяя студентам легко получать рекомендации на любом устройстве.

Для поиска подходящего электива по запросу пользователя, как уже было сказано ранее, используется контентная фильтрация. Для выбора наиболее подходящей модели контентной фильтрации мы сравнили следующие методы решения:

* SentenceTransformer представляет собой технологию модификации предварительно обученной сети BERT. Эти модели обеспечивают высококачественные векторные представления для текста, которые используются для расчета косинусного сходства между запросом и элективами. Сейчас для SentenceTransformer существует множество предобученных моделей для различных языков, в том числе и для русского языка (paraphrase-multilingual-MPnet-base-v2, paraphrase-multilingual-MiniLM-L12-v2 и paraphrase-xlm-r-multilingual-v1). [9]
* FastText. Библиотека, содержащая предобученные готовые векторные представления слов и классификатор – алгоритм разбивающий слова на классы. FastText разработала команда исследователей из Facebook AI Research, в составе которых был и создатель Word2vec Томаш Миколов, для классификации текстов, но так же может быть использована для обучения эмбеддингов слов. FastText получил широкое применение по причине хорошей скорости тренировки и отличной работоспособности. FastText учитывает, что различные слова, могут быть связаны синтаксически или семантически, если они имеют одни и те же компоненты (например, однокоренные слова) [10]
* TF-IDF. Cтатистическая мера, используемая для оценки важности слова в контексте документа или набора документов. TF-IDF помогает выделить ключевые слова и фразы, которые наиболее важны для конкретного документа. TF-IDF не только учитывает частоту появления слова в документе, но и его распространенность в других документах, что позволяет более точно оценить его значимость.

В ходе работы над проектом был проведен сравнительный анализ моделей для оценки соответствия предложенных элективов запросам студентов. Для этого сравнивались показатели косинусного сходства между векторными представлениями запросов и соответствующих элективов, которые были предложены каждой из перечисленных выше моделью. Для сравнения работы моделей мы провели тестирование на трех запросах: коротком (одно слово), развёрнутом (предложение) и сложном (несколько предложений со словами из различных сфер, например: рисование и компьютер).

Запрос №1: «Литература». На рисунке 2 представлены элективы подобранные каждой моделью для запроса №1.



Рис. 2 – Топ-5 элективов для запроса №1

Запрос №2: «В школе я увлекалась литературой и писанием, даже мечтала стать журналистом». На рисунке 3 представлены элективы подобранные каждой моделью для запроса №2.

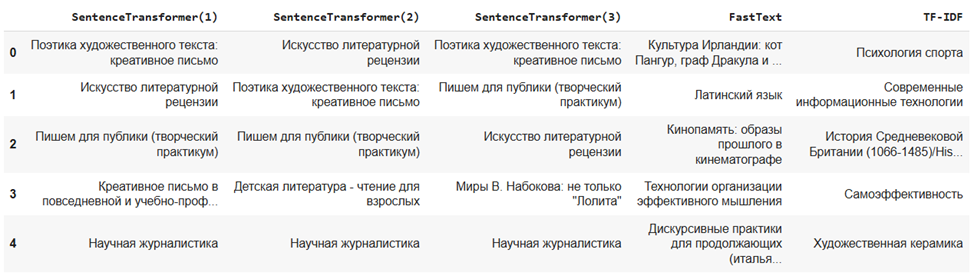


Рис. 3 – Топ-5 элективов для запроса №2

Запрос №3: «Я интересуюсь литературой и хотел бы найти элективные курсы или образовательные программы, которые затрагивают глубокие аспекты анализа литературных произведений. Интересуют темы, связанные с историей литературы, теорией текста, литературной критикой и компаративным анализом.  Также хочу узнать, если есть курсы, которые обучают современным подходам к изучению текстов, например, с использованием цифровых методов анализа или междисциплинарного подхода, где литература пересекается с социологией, психологией или искусствоведением. Я давно не уделял этому должного внимания, так как по специальности интересуюсь такими направлениями как программирование, анализ данных, робототехника и интернет вещей». На рисунке N представлены элективы подобранные каждой моделью для запроса №3.



Рис. 4 – Топ-5 элективов для запроса №3

На всех запросах лучше всего себя показала модель SentenceTransformer(1) – SentenceTransformer с paraphrase-multilingual-mpnet-base-v2: вне зависимости от сложности запроса она находила элективы с наивысшим косинусным сходством (0,5-0,6 для простого, короткого запроса, 0,6-0,7 для сложного), при этом элективы соответствовали содержанию запроса. Визуальное представление результатов сравнения моделей можно увидеть в приложении. На рисунках приложения наглядно показано, что SentenceTransformer одинаково хорошо работает на всех запросах с указанием косинусного сходства, определённого моделями. С увеличением длины и сложности запроса SentenceTransformer(1) продолжает подбирать релевантные элективы, также как и другие модели SentenceTransformer. FastText с усложнением запроса также подбирает элективы с высоким косинусным сходством (0,7 и выше), но сами элективы не всегда соответсвуют запросу. TF-IDF не работает с коротким запросом, а на более сложных запросах подбирает нерелевантные элективы.

Таким образом, мы остановили свой выбор на модели SentenceTransformer(1) – SentenceTransformer с paraphrase-multilingual-mpnet-base-v2 для контентной фильтрации.

# **3. Используемые технологии**

* Python.

Основной язык программирования, используемый для реализации всех этапов обработки данных и разработки рекомендательной системы. Благодаря большому количеству библиотек Python идеально подходит для задач обработки текста, машинного обучения и создания ботов.

* Pandas

Библиотека для работы с табличными данными. Используется для чтения данных из Excel и CSV-файлов, их предобработки и преобразования в необходимые форматы.

* Natasha

Библиотека для обработки текстов на русском языке. Используется для сегментации текста, нормализации токенов и выделения важных лексем, что улучшает качество анализа текстов и построения эмбеддингов.

* Sentence-Transformers

Предоставляет предобученные языковые модели, такие как paraphrase-multilingual-mpnet-base-v2, для получения эмбеддингов текстов. Это позволяет преобразовывать текстовые данные в числовой формат, подходящий для сравнения и кластеризации.

* NumPy

Библиотека для работы с многомерными массивами. Используется для сохранения и обработки эмбеддингов текстов, а также для выполнения векторных операций.

* Surprise

Фреймворк для разработки систем рекомендаций. Применяется для построения рекомендательной модели на основе метода SVD (разложения матриц), который подходит для коллаборативной фильтрации.

* Pickle

Библиотека для сериализации объектов Python. Используется для сохранения обученной модели SVD, что позволяет повторно использовать её без необходимости переобучения.

* Microsoft Excel / CSV

Формат данных для исходных наборов информации об элективах и курсах. Используется для хранения и передачи данных.

* Aiogram

Фреймворк для разработки Телеграм-ботов. Используется для взаимодействия с пользователями, получения их запросов и выдачи персонализированных рекомендаций. Бот обрабатывает входные данные (например, предпочтения студентов) и возвращает релевантные результаты, используя обученные модели контентной и коллаборативной фильтрации.

* Telegram API

Интерфейс для интеграции с мессенджером Telegram. Обеспечивает обмен сообщениями между ботом и пользователями.

# **4. Реализация проекта**

Проект реализован в виде Telegram-бота, который составляет рекомендации с помощью двух моделей, подбирающих элективы: модель контентной фильтрации и модель коллаборативной фильтрации. Основной функционал сосредоточен на обработке запросов от пользователя и предоставлении рекомендаций. Основные блоки разработанного приложения и их взаимодействие можно посмотреть на структурной схеме приложения на рисунке 5.

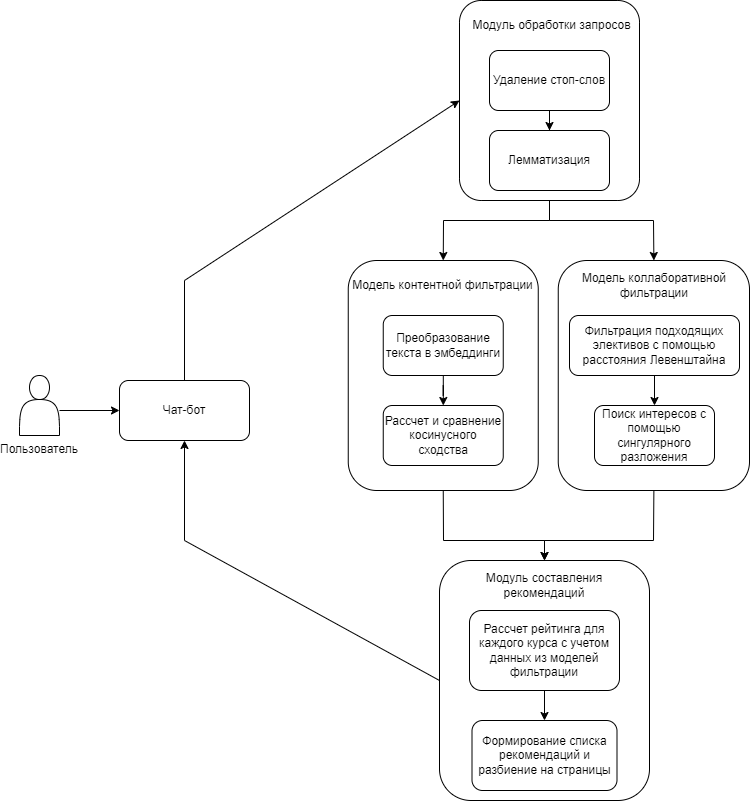


Рис. 5 - Структурная схема приложения

Чат-бот представляет интерфейс, с помощью которого пользователь взаимодействует с программой по средствам ввода запросов и нажатия кнопок. В качестве элементов чат-бота, предназначенных для взаимодествия с пользователем на рисунке 6 представлена функция, реализующая клавиатуру сообщения с рекомендованными элективами. Для удобного просмотра рекомендованных элективов полный список разделён на страницы, которые пользователь перелистывает с помощью кнопок «Назад» и «Вперед».

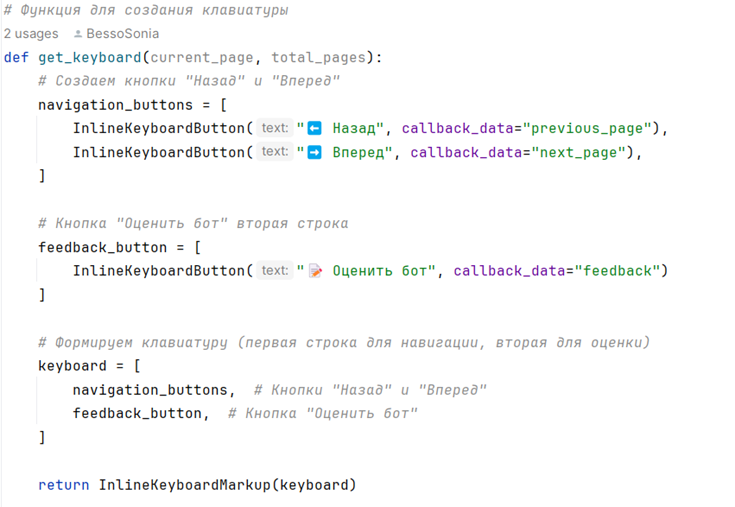


Рис. 6 – Создание клавиатуры под сообщением чат-бота

Модель коллаборативной фильтрации использует данные о предпочтениях других пользователей, похожих на текущего, для создания рекомендаций. На рисунке 7 представлен пример создания модели коллаборативной фильтрации.

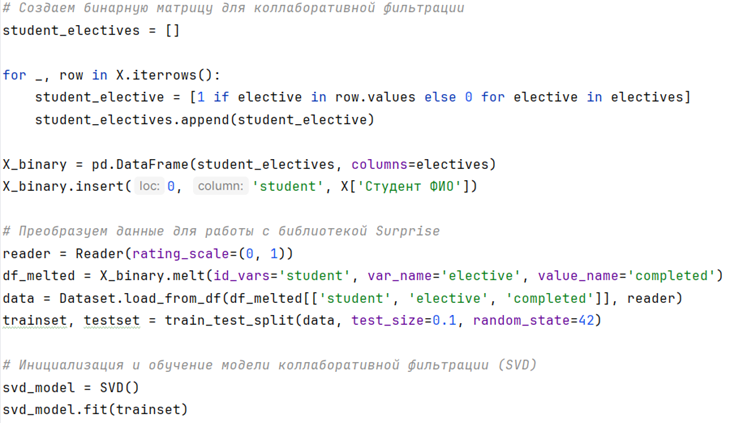


Рис. 7 – Создание модели коллаборативной фильтрации

Модель контентной фильтрации анализирует характеристики элективов и находит совпадения с интересами пользователя. На рисунке 8 представлена часть реализации модели контентной фильтрации.



Рис. 8 – Создание эмбеддингов

Модуль обработки запросов принимает запросы пользователя. Этот модуль обрабатывает запрос и передает его в систему фильтрации. На рисунке 9 представлен пример обработки запроса пользователя.

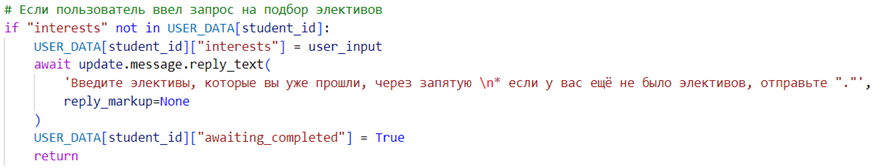


Рис. 9 – Пример обработки запроса пользователя

Модуль составления рекомендаций комбинирует результаты контентной и коллаборативной фильтрации, формируя итоговый список рекомендованных элективов. Функция объединяющая предсказания модели представлена на рисунке 10.

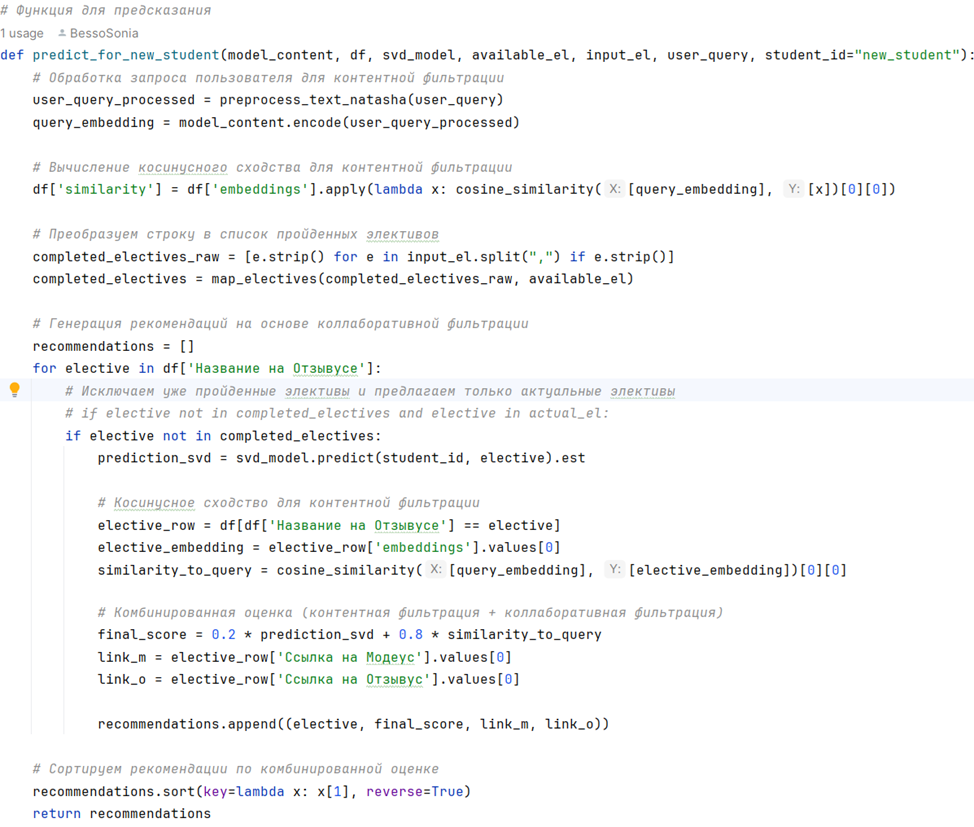


Рис. 10 – Функция составления рекомендаций

# **5. Пользовательский интерфейс**

На рисунках 11, 12 и 13 представлен интерфейс Telegram-бота.

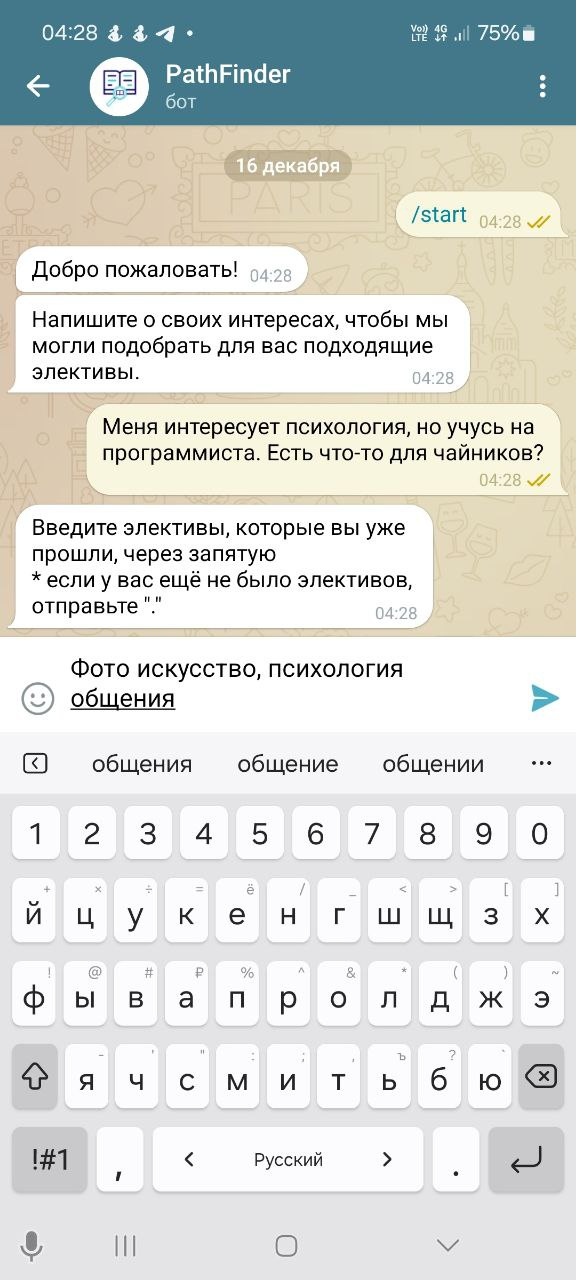


Рис. 11 – Отправка запроса боту

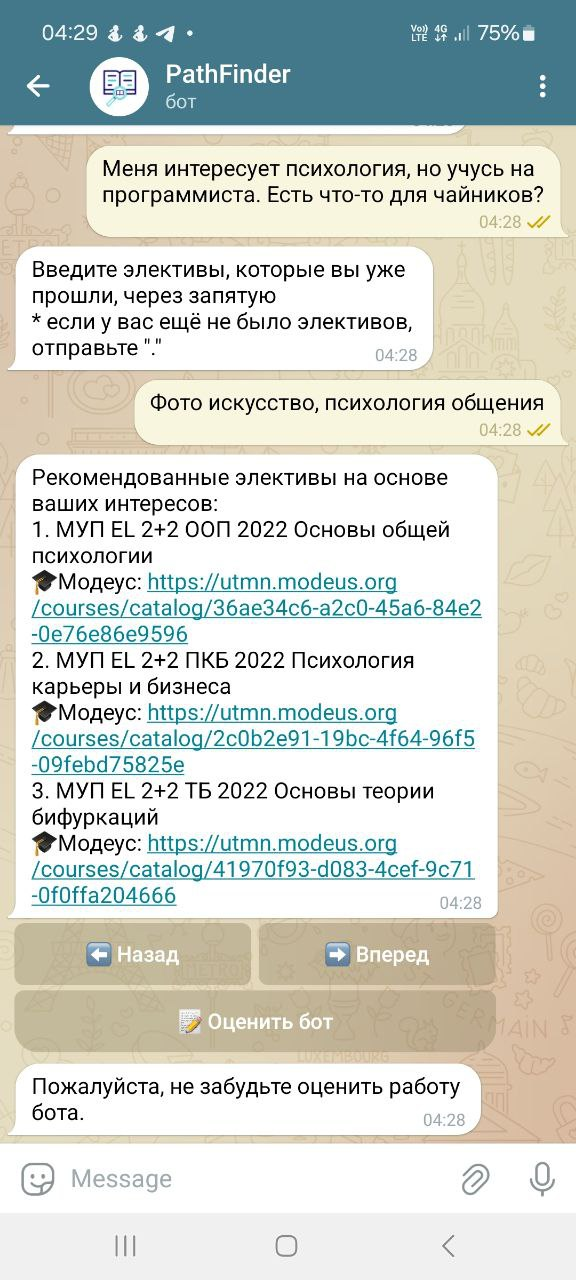


Рис. 12 – Представление рекомендаций на нескольких «страницах»

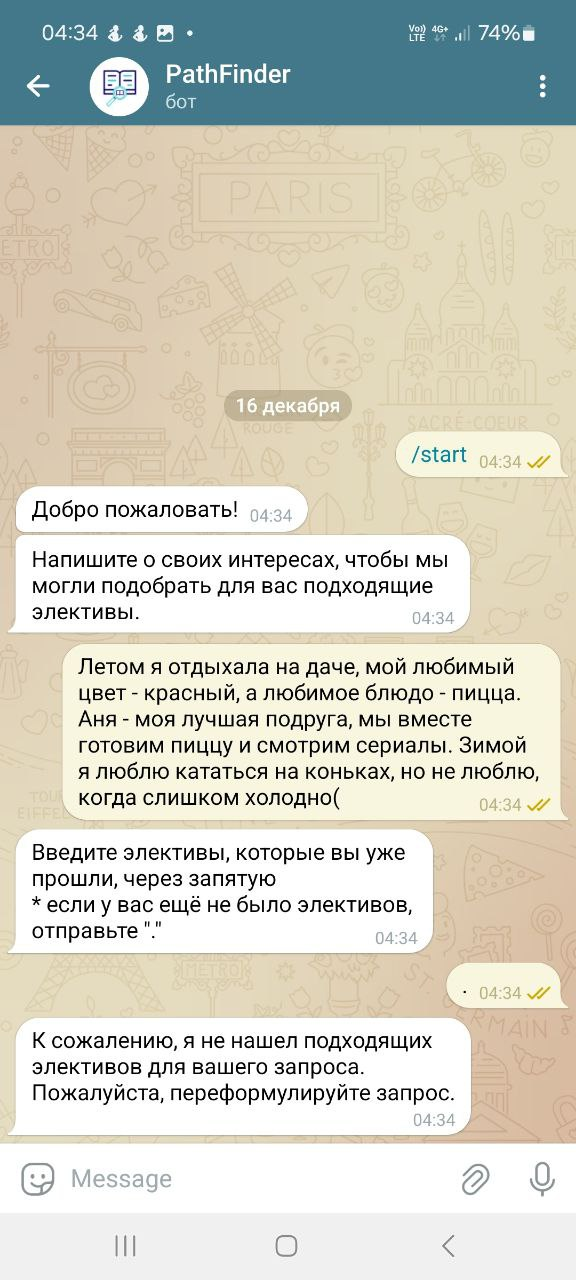


Рис. 13 – Вывод сообщения об отсутствии подходящих элективов

# **6. Тестирование**

Цель тестирования: оценка функциональности приложения для поиска подходящего учебного курса на основе описания и требований студента.

Фокус-группа: 37 студентов (27 на 1-м курсе, 10 на 4-м курсе).

Гендерное распределение: 62% мужчин, 38% женщин.

Методология тестирования: участники вводили запросы в свободной форме, описывая желаемую программу обучения, её ключевые особенности и требования. Приложение предоставляло 5 наиболее релевантных курсов. Участники оценивали, насколько результаты соответствовали их запросу по шкале от 1 до 5.

Результаты тестирования:

Средняя оценка релевантности, подобранных Telegram-ботом элективов - 4,57.

Распределение оценок:

* 5 баллов: 55%
* 4 балла: 30%
* 3 балла: 10%
* 2 балла: 3%
* 1 балл: 2%

Удовлетворенность интерфейсом: 85% участников отметили, что интерфейс был интуитивно понятным.

Выявленные проблемы: некоторые участники отметили, что приложение иногда выбирало курсы с недостаточной релевантностью для сложных запросов. Не хватало возможностей фильтрации (например, по легкости и наличию свободных мест).

Положительные отзывы:

* 82% участников похвалили точность подбора для общих запросов.
* Быстродействие приложения оставило положительное впечатление у 89% участников.

По результатам тестирования среди студентов первого курса, которые не имеют опыта выбора элективных дисциплин и впервые разбираются в построении индивидуальной образовательной траектории, можно сказать, что Telegram-бот действительно помог им в выборе элективов. Первокурсники были заинтересованы в тестировании чат-бота, так как время тестирования проекта совпадало с временем начала выбора элективных дисциплин. Благодаря сбору обратной связи после получения рекомендаций были получены предложения о доработке функционала, а именно - актуализации элективов по средствам обновления списка дисциплин для выбора в соответствии с меню выбора в Modeus для мониторинга занятости элективов и рекомендации только тех дисциплин, которые ещё не были заняты за прошедшую волну выбора.

Результаты тестирования среди старших курсов говорят о том, что студенты довольны подобранными элективами. Более того, о многих рекомендованных дисциплинах они слышали впервые, несмотря на то что выбирали элективы уже не менее 3 раз. От старшекурсников также поступило предложение по улучшению функционала: улучшить меню чат-бота, чтобы из любой точки сценария можно было бы начать подбор электива заново.

# **7. Заключение**

Таким образом, был создан Telegram-бот, который взаимодействует с пользователем по средствам простых вопросов, собирает информацию о его интересах и пройденных элективах с помощью обработки запроса, генерирует персонализированные рекомендации на основе комбинированной оценки моделей контентной и коллаборативной фильтрации и возвращает отсортированные по релевантности запросу рекомендации, отображенные на нескольких страницах. В основе модели контентной фильтрации была выбрана многоязычная модель paraphrase-multilingual-mpnet-base-v2 из библиотеки Sentence-Transformers для создания эмбеддингов из описаний элективов и запросов пользователей. В основе модели коллаборативной фильтрации метод SVD-разложения, для поиска введённых элективов среди существующих используется расстояние по Левенштайну. У бота понятный сценарий взаимодействия, что делает процесс подбора элективов с его помощью простым для студента. Согласно проведённому тестированию, пользователи, то есть студенты, остались довольны проектом.

# **8. Список литературы**

1. Elective: [Электронный ресурс] // Тюменский Государственный Университет. URL: https://www.utmn.ru/obrazovanie/iot/electives/ (Дата обращения: 19.12.2024)
2. Криворогов Д.Д. Разработка сервиса для сбора и анализа отзывов на элективные дисциплины / Д. Д. Криворогов, Т. Д. Низамов, А. А. Фазлыев, А. Н. Ходырев, Д. В. Шушарин, А. В. Глазкова // Вестник Новосибирского государственного университета. Серия: Информационные технологии – 2023. – с.5-19
3. Таратухина Ю.В. Машинное обучение модели информационной рекомендательной системы по вопросам индивидуализации образования / Ю.В. Таратухина, Т.В. Барт, В.В. Власов // Образовательные ресурсы и технологии. – 2019. – с.7-14
4. Белоцкий Е.А. Построение рекомендательной системы по подбору высших учебных заведений для абитуриентов / Е.А. Белоцкий, А.В. Суетин // Вестник Санкт-Петербургского университета. – 2016. – с.66-77.
5. Круглик А.С. Гибридные рекомендательные системы / А.С. Круглик // Наука о данных Санкт-Петербург. – 2020. – с.157-160
6. Игнатьев В.Ю. Построение гибридной рекомендательной системы с улучшением точности / В.Ю. Игнатьев, Д.В. Лемтюжникова, Д.И. Руль, И.Л. Рябов // Известия Российской академии наук. Теория и системы управления – 2018. – с.101-108
7. Аристова А.С. Использование чат-ботов в образовательном процессе / А.С. Аристова, Ю.С. Безносюк, П.К. Ведикер, Н.Е. Воронович // Международная конференция «Цифровая трансформация общества, экономики, менеджмента и образования» - 2019. – с.95-99
8. Почти три четверти россиян активно пользуются чат-ботами: [Электронный ресурс] // ГАЗЕТА.RU. URL: https://www.utmn.ru/obrazovanie/iot/electives/ (Дата обращения: 19.12.2024)
9. Николаев И.Е. Сравнение нейросетевых моделей на архитектуре трансформеров в контексте задачи оценки компактности векторных представлений семантически близких текстов требований европейской классификации навыков ESCO / И.Е. Николаев, А.В. Мельников // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2022. – с.19-29
10. Соколова К.А. Обзор методов семантического анализа текстов / К.А. Соколова, В.В. Буряченко // Актуальные проблемы авиации и космонавтики – 2022. – с.172-175

# **Приложение. Сравнения моделей для создания эмбеддингов**

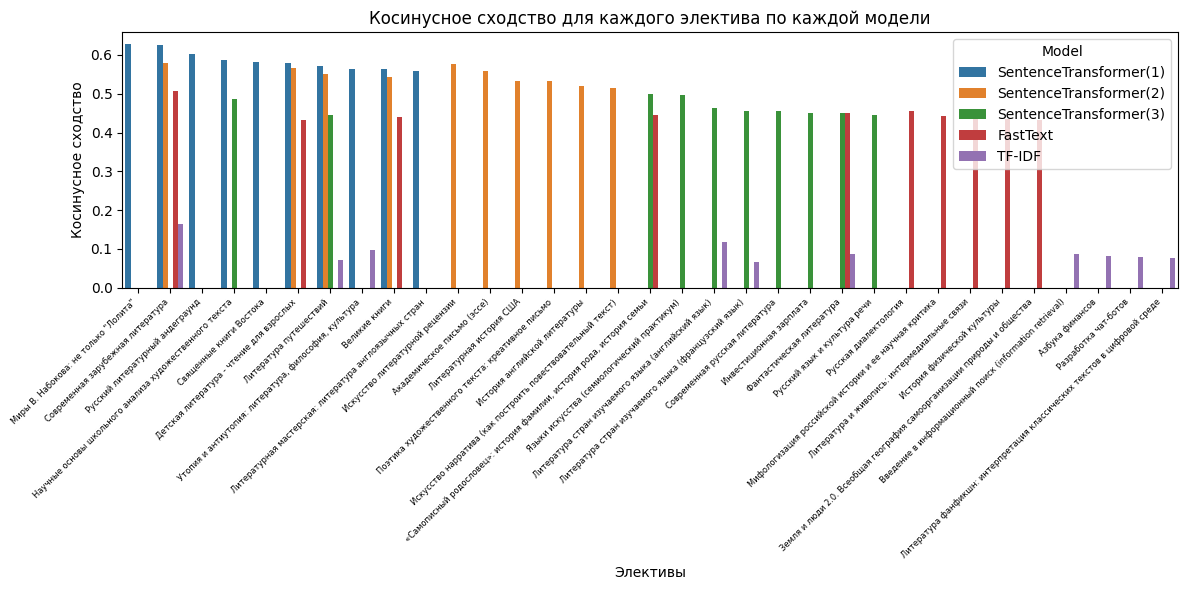


Рисунок 3 – Элективы, подобранные моделями для запроса №1

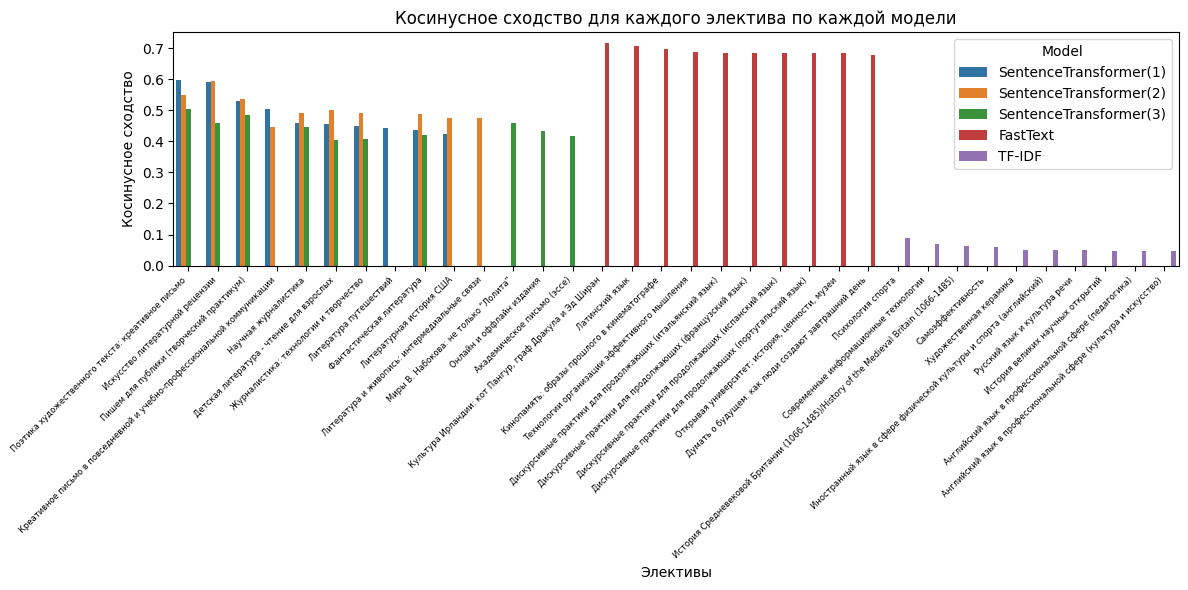


Рисунок 2 – Элективы, подобранные моделями для запроса №2

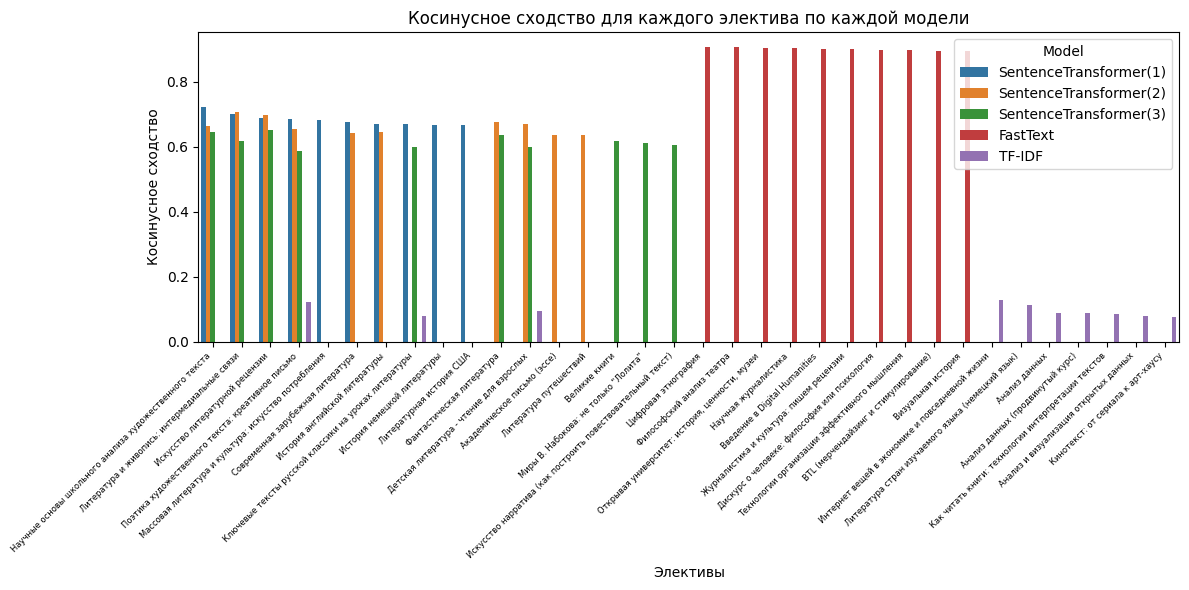


Рисунок 1 – Элективы, подобранные моделями для запроса №3