МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ШКОЛА КОМПЬЮТЕРНЫХ НАУК

Кафедра программного обеспечения

**ОТЧЁТ ПО ПРОЕКТНОЙ ДЕЯТЕЛЬНОСТИ**

РАЗРАБОТКА ВЕБ-СЕРВИСА ДЛЯ АВТОМАТИЗАЦИИ ПЕРЕЗАПИСИ НА ЭЛЕКТИВНЫЕ КУРСЫ С ИСПОЛЬЗОВАНИЕМ РЕКОМЕНДАТЕЛЬНОГО АЛГОРИТМА

| Выполнили работу  студенты 4 курса |  | Гордиенко Андрей Вячеславович  Кузнецов Виктор Вячеславович |
| --- | --- | --- |
| Руководитель |  | Перевалова Мария Николаевна |

## 

ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 3](#_7r3gjqsrdksh)

[ГЛАВА 1. ОПИСАНИЕ СПЕЦИФИКИ ПРЕДМЕТНОЙ ОБЛАСТИ 4](#_wed8ljkce29e)

[ГЛАВА 2. АКТУАЛЬНОСТЬ ПРОЕКТА 5](#_dbnwdhseqwqp)

[ГЛАВА 3. ТРЕБОВАНИЯ К ПРОЕКТУ И УЧИТЫВАЕМЫЕ ОГРАНИЧЕНИЯ 8](#_qhsbl8fzslib)

[ГЛАВА 4. ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ ПРОЕКТА 11](#_ees26ntii8tf)

[ГЛАВА 6. СОБСТВЕННЫЕ ТЕОРЕТИЧЕСКИЕ РЕШЕНИЯ 13](#_l8x2oa3qc1nc)

[ГЛАВА 7. СОБСТВЕННЫЕ ПРАКТИЧЕСКИЕ РЕШЕНИЯ 14](#_ytg3ioiyzkl0)

[ГЛАВА 8. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ПРОЕКТА 18](#_12pgfmvxe7ji)

[ГЛАВА 9. ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС 20](#_efr2nu3s7d3r)

[ГЛАВА 10. ОРГАНИЗАЦИЯ ТЕСТИРОВАНИЯ ПРОЕКТА 23](#_xp9349dkzx0h)

[ЗАКЛЮЧЕНИЕ 24](#_bjigkg6u28j6)

[СПИСОК ЛИТЕРАТУРЫ 25](#_dyzgrytn07xx)

## 

# ВВЕДЕНИЕ

В современном образовательном процессе всё большее значение приобретает эффективность управления образовательными ресурсами, особенно в контексте выбора элективных курсов. Учащиеся сталкиваются с вызовами, связанными не только с поиском курсов, соответствующих их интересам и образовательным целям, но и с ограничениями текущих систем перезаписи. Существующие решения нередко оказываются недостаточно удобными и гибкими, что приводит к большим временным затратам со стороны менеджеров индивидуальных образовательных траекторий.

Важной задачей является не только помощь в выборе курсов, но и создание системы, которая бы обеспечивала прозрачный и простой процесс перезаписи. Проект направлен на разработку решения, способного не только минимизировать временные затраты, но и повысить удовлетворенность пользователей благодаря современным подходам машинного обучения.

Поэтому появляется идея помочь менеджерам ИОТ с организацией процесса перезаписи и обучающимся за счет рекомендательного алгоритма.

Целью данной выпускной квалификационной работы является создание приложения, позволяющего автоматизировать процесс перезаписи элективных курсов с предоставлением рекомендаций для выбора электива.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Организовать получение выгрузки записей студентов на элективы
2. Разработать модель перезаписи обучающихся на элективные курсы
3. Определить точки взаимодействия с университетской системой
4. Разработать алгоритм рекомендации элективных курсов
5. Разработать пользовательский интерфейс для обучающегося и менеджера

# ГЛАВА 1. ОПИСАНИЕ СПЕЦИФИКИ ПРЕДМЕТНОЙ ОБЛАСТИ

#### 1.1 ОБЩАЯ ХАРАКТЕРИСТИКА

Предметная область охватывает процесс записи и перезаписи студентов высших учебных заведений на элективные курсы (дисциплины по выбору), которые являются частью образовательной программы. Этот процесс регулируется нормативно-правовыми актами, внутренними регламентами университета и расписанием учебных занятий.

#### 1.2 ОСНОВНЫЕ ЭЛЕМЕНТЫ ПРЕДМЕТНОЙ ОБЛАСТИ

1. Студенты:
   1. Являются ключевыми пользователями системы.
   2. Принадлежат к различным курсам и специальностям.
   3. Имеют индивидуальные учебные планы, где указаны доступные элективы.
2. Элективы (курсы по выбору):
   1. Имеют конкретные сроки обучения.
   2. Имеют ограничения по количеству студентов (лимит мест).
   3. Могут быть привязаны к определённым семестрам и факультетам.
3. Расписание:
   1. Определяет доступность элективов по времени.
   2. Содержит данные о пересечениях с другими занятиями.
4. Персонал (администрация):
   1. Устанавливает сроки записи и перезаписи.
   2. Следит за заполнением групп.
   3. Контролирует соблюдение правил выбора курсов.
5. Правила и ограничения:
   1. У каждого студента должен быть установлен лимит на количество элективов.
   2. Перезапись доступна только в установленные сроки.

#### 

#### 1.3 ПРОЦЕССЫ ПРЕДМЕТНОЙ ОБЛАСТИ

1. Первичная запись на элективы:
   1. Студенты выбирают курсы из предложенного списка в заданные администрацией сроки до начала семестра.
   2. Проверяется наличие свободных мест и отсутствие конфликтов в расписании.
2. Перезапись:
   1. Возможна только в установленные сроки (в течение первой недели семестра).
   2. Студент может отказаться от текущего курса и выбрать другой, при условии наличия мест и отсутствия конфликтов в расписании.
   3. Реализована посредством заполнении Яндекс формы, с указанием текущего и желаемого элективов.
3. Управление расписанием:
   1. Администрация добавляет новые элективы, редактирует параметры курсов (даты, лимиты мест).
   2. Учитываются изменения расписания занятий.
4. Контроль заполненности групп:
   1. Группы курсов закрываются для записи при достижении лимита.

#### 1.4 ОСОБЕННОСТИ ПРЕДМЕТНОЙ ОБЛАСТИ

* 1. Студенты часто запрашивают перезапись в последние дни доступного срока. Из-за чего нагрузка распределяется не равномерно на весь срок предназначенный для перезаписи
  2. Процессы записи строго привязаны к временным рамкам и регламентам университета.

## ГЛАВА 2. АКТУАЛЬНОСТЬ ПРОЕКТА

### 2.1 ЗАКАЗЧИК И ПОЛЬЗОВАТЕЛИ

Заказчиком проекта являются менеджеры управления ИОТ, предоставляющие перезапись на элективные курсы. Основными пользователями системы являются:

1. Студенты и учащиеся: Желающие сменить свой выбор элективной дисциплины.
2. Администраторы: Ответственные за организацию и управление образовательными программами.

### 2.2 ПРОБЛЕМЫ

Со стороны студентов:

1. Ограниченный доступ к информации о доступных элективах. Студенты не всегда видят актуальную информацию о свободных местах и расписании элективов.
2. Высокая вероятность конфликта в расписании. Студенты могут выбрать новый электив, который пересекается по времени с их текущими занятиями, что приводит к необходимости повторной перезаписи.

Со стороны менеджеров:

1. Высокая нагрузка на сотрудников. Процесс обработки заявок на перезапись занимает много времени, особенно при большом числе студентов, что снижает общую эффективность работы администрации.
2. Высокая вероятность ошибок при ручной обработке заявок.
3. Отсутствие автоматического учета лимита мест. Вручную трудно контролировать заполненность групп и корректно распределять студентов, особенно если изменения происходят динамично

## 

### 2.3 ИДЕЯ ПРОЕКТА

Идея проекта заключается в создании веб-сервиса, который решает проблемы ручной обработки перезаписи студентов на элективные курсы и упрощает процесс для менеджеров и студентов. Это достигается за счет автоматизации проверки ограничений: система самостоятельно анализирует доступность курсов с учетом лимитов по количеству мест, расписания и других параметров, исключая человеческие ошибки и ускоряя процесс.

Дополнительно веб-сервис предоставляет рекомендации для студентов, помогая им выбирать наиболее подходящие элективы. Это реализуется через использование методов кластеризации курсов по тематическим направлениям, что позволяет предлагать более релевантные варианты.

## 

## 

## 

## 

## 

## 

## 

## ГЛАВА 3. ТРЕБОВАНИЯ К ПРОЕКТУ И УЧИТЫВАЕМЫЕ ОГРАНИЧЕНИЯ

### 3.1 ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

#### 3.1.1 РАБОТА СО СТУДЕНТАМИ

### Студент должен иметь возможность авторизоваться в системе (по логину/паролю или через интеграцию с университетской учетной записью).

### Система должна отображать список текущих элективов, на которые студент уже записан.

### Студент должен видеть доступные элективы с:

### Названием курса.

### Описанием (цели, темы, преподаватель).

### Количеством свободных мест.

### Временем проведения и возможными пересечениями в расписании.

### Студент должен иметь возможность выбрать новый электив и заменить текущий.

### После перезаписи студент должен получить подтверждение (уведомление в интерфейсе и/или по электронной почте).

#### 3.1.2 РАБОТА С АДМИНИСТРАЦИЕЙ

### Администратор должен иметь доступ к панели управления для:

### Добавления новых элективов.

### Установки лимита мест для каждого электива.

### Просмотра текущей заполненности групп.

### Система должна предоставлять возможность администратору изменять параметры уже существующих элективов (например, лимит мест или преподавателя).

### Администратор должен получать автоматизированные отчеты о:

### Количестве заявок на перезапись.

### Заполненности элективов.

### Популярности курсов.

#### 

#### 3.1.3 ЛОГИКА СИСТЕМЫ

### Система должна автоматически проверять:

### Доступность мест на выбранном элективе.

### Отсутствие конфликтов с расписанием других занятий студента.

### Соблюдение правил (например, не больше N элективов в семестр).

### Система должна блокировать возможность перезаписи по истечении сроков, установленных администрацией.

#### 3.1.4 УВЕДОМЛЕНИЯ

### Система должна отправлять уведомления студентам:

### О начале периода перезаписи.

### Об успешной или неуспешной перезаписи (например, если места закончились).

### Администраторы должны получать оповещения о заполненности групп или других критичных изменениях.

### 2.2 НЕФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

#### 2.1.1 ПРОИЗВОДИТЕЛЬНОСТЬ

1. Система должна обрабатывать запросы на перезапись в режиме реального времени (время отклика не более 3 секунд).
2. Одновременно система должна поддерживать работу до 500 активных пользователей.

#### 2.1.2 ДОСТУПНОСТЬ

1. Сервис должен быть доступен 24/7, за исключением времени проведения технического обслуживания (не более 2 часов в неделю).

#### 2.1.3 БЕЗОПАСНОСТЬ

1. Доступ к системе должен быть защищен авторизацией.
2. Данные о студентах и их записях должны быть защищены.
3. Только администраторы могут управлять настройками элективов.

#### 2.1.4 ИНТЕРФЕЙС

1. Интерфейс должен быть интуитивно понятным для студентов, с минимальным количеством шагов для выполнения перезаписи.
2. Должна быть поддержка мобильных устройств (адаптивный дизайн).

## 

## 

## ГЛАВА 4. ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ ПРОЕКТА

Основные алгоритмы и модели заложены в блок рекомендаций разрабатываемой системы и представлены следующим списком:

1. TF-IDF - алгоритм текстового анализа, определяющий важность терминов в документе относительно общего корпуса документов. В рамках реализации проекта применялся для векторизации текстовых описаний курсов и представления слов, характеризующих темы кластеров.
2. K-Means - Алгоритм кластеризации, разделяющий данные на K кластеров, минимизируя внутрикластерную дисперсию. В рамках реализации проекта применялся для кластеризации элективов по схожести их описаний.
3. t-SNE - метод визуализации данных, понижающий размерность и сохраняющий структуру локального сходства между объектами. В рамках реализации проекта применялся для визуализации кластеров элективов, сформированных алгоритмом кластеризации (например, K-Means). t-SNE позволил визуально оценить группы курсов на основе их описаний и выявить потенциальные пересечения или схожести между элективами.

Кроме того, были опробованы следующие алгоритмы и методы, но их результаты оказались нерепрезентативными для текущей реализации:

1. Latent Dirichlet Allocation (LDA). Метод тематического моделирования, основанный на вероятностном распределении слов по темам и тем по документам.
2. BERTopic. Современная модель тематического моделирования, сочетающая эмбеддинги предложений с алгоритмом кластеризации HDBSCAN.

ГЛАВА 5. ПРАКТИЧЕСКОЕ ОБОСНОВАНИЕ ПРОЕКТА

### 5.1 ТЕХНОЛОГИИ И БИБЛИОТЕКИ

Для реализации проекта использовались следующие технологии и библиотеки:

1. Язык программирования:
   1. Python: выбран за его широкие возможности в области обработки данных, машинного обучения и NLP.
2. Библиотеки для обработки данных и NLP:
   1. Pandas: для работы с табличными данными.
   2. NumPy: для числовых вычислений.
   3. NLTK: для обработки естественного языка, включая токенизацию и удаление стоп-слов.
   4. Pymorphy2: для лемматизации русскоязычного текста.
   5. re: для работы с регулярными выражениями при предобработке текста.
3. Дополнительные инструменты:
   1. JSON: для сохранения результатов кластеризации и рекомендаций в формате, удобном для интеграции с веб-интерфейсом.

### 5.2 ФРЕЙМВОРКИ ДЛЯ РАЗРАБОТКИ

1. Веб-разработка:
   1. Frontend: Использование современного фреймворка Vue.js с vuetify для создания интуитивно понятного интерфейса пользователя.
   2. Backend: Использование FastAPI для разработки серверной части, обеспечивающей обработку запросов.

## 

## 

## 

## ГЛАВА 6. СОБСТВЕННЫЕ ТЕОРЕТИЧЕСКИЕ РЕШЕНИЯ

В ходе реализации проекта были разработаны и внедрены собственные теоретические концепции представленные списком ниже:

1. Разработан алгоритм организации перезаписи на элективы, изображен на рисунке 1.

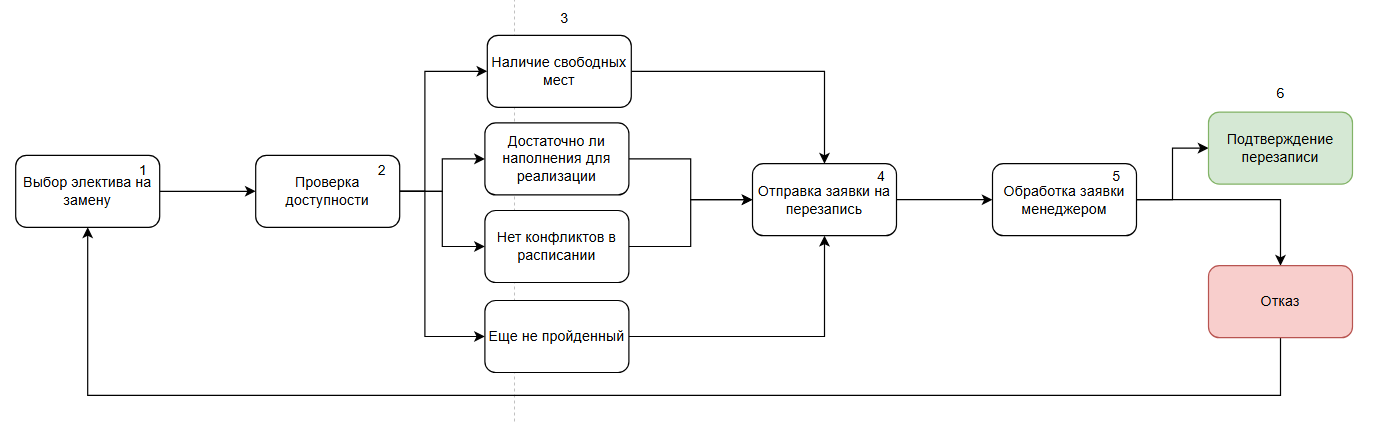


Рисунок 1. Алгоритм перезаписи.

1. Разработка и внедрение иерархического подхода к кластеризации, позволяющего сначала выделить основные кластеры, а затем провести подклассификацию внутри каждого из них. Это позволило повысить точность и интерпретируемость результатов, обеспечив структурированное разделение курсов по тематическим направлениям.
2. Внедрение дополнительного шага по удалению высокочастотных неинформативных слов на основе анализа частоты встречаемости слов в корпусе.
3. Разработка механизма формирования биграмм и триграмм для создания устойчивых словосочетаний, что повысило осмысленность ключевых слов кластеров.

## 

## 

## 

## ГЛАВА 7. СОБСТВЕННЫЕ ПРАКТИЧЕСКИЕ РЕШЕНИЯ

Был разработан прототип веб-приложения, включающий в себя клиентскую и серверную часть

В клиентской части были разработаны следующие страницы:

1. Страница авторизации (авторизация реализована через корпоративную почту, путем ввода кода подтверждения, отправленного на почту)
2. Профиль обучающегося, содержит следующие блоки:
   1. Информация об обучающемся (ФИО, почта, направление подготовки, год поступления)
   2. Выбранные элективные курсы с возможностью перезаписать их
   3. Анкетирование для получения рекомендаций
3. Страница менеждера:
   1. Возможность загрузить данные о прошедшей выборной кампании элективов
   2. Просмотр и редактирование информации о курсах и записях студентов
   3. Получение и подтверждение/отказ заявок на перезапись от студентов
4. Главная страница: отображение и поиск всех элективных курсов, информация о них (название, описание, преподаватель, заполненность групп)

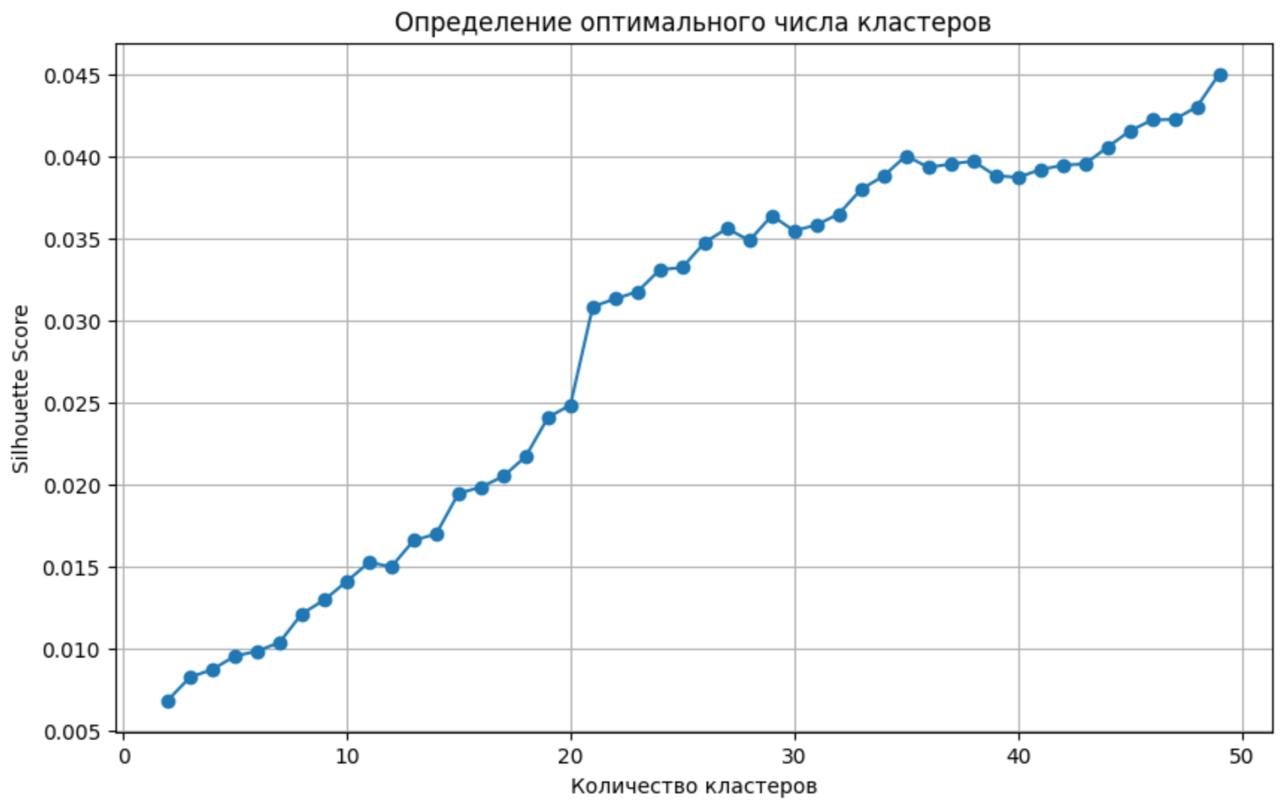
В серверной части были разработаны следующие эндпоинты:

1. auth - post запрос на отправку кода подтверждения на указанную почту для прохождения авторизации.
2. recomendation - get запрос на получение рекомендаций для указанного пользователя.
3. profile - get запрос на получение информации о пользователе.
4. upload\_company - post запрос на загрузку файла о текущей выборной кампании
5. upload\_timetable - post запрос на загрузку файла с акутуальных расписанием элективных курсов
6. electives - get запрос на получение доступных для записи элективов для авторизованного пользователя

Информация о пользователе передается через cookie вместе с запросом.

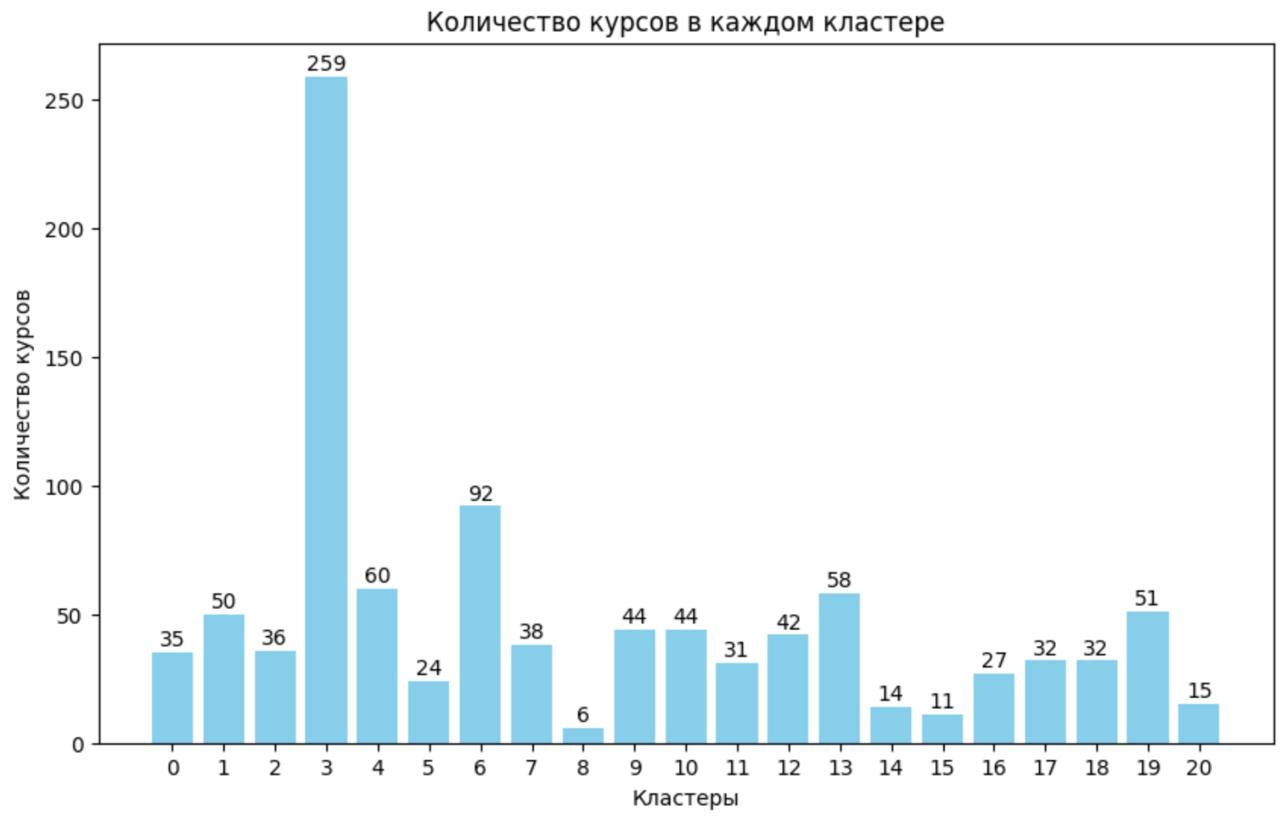
Также, в процессе разработки проекта были реализованы следующие собственные практические решения в рамках задачи разработки модуля рекомендаций:

1. Определение оптимального числа кластеров с ипользованием графической визуализации результатов для облегчения анализа и принятия решения о количестве кластеров (рис.2).

  
*Рисунок 2. Определение оптимального числа кластеров.*

На основании анализа визуализации и собственных наблюдений было принято решение о разбиении курсов на 21 кластер.

1. Визуализация результатов кластеризации:
   1. Внедрение метода t-SNE для создания двумерной проекции высокоразмерных данных и визуального анализа распределения кластеров (рис.4).
   2. Создание гистограммы распределения курсов по класстерам с использованием библиотеки Matplotlib для лучшего понимания структуры данных (рис.3).

  
*Рисунок 3. Количество курсов в каждом кластере.*

## 

*Рисунок 4. Визуализация кластеров.*

## 

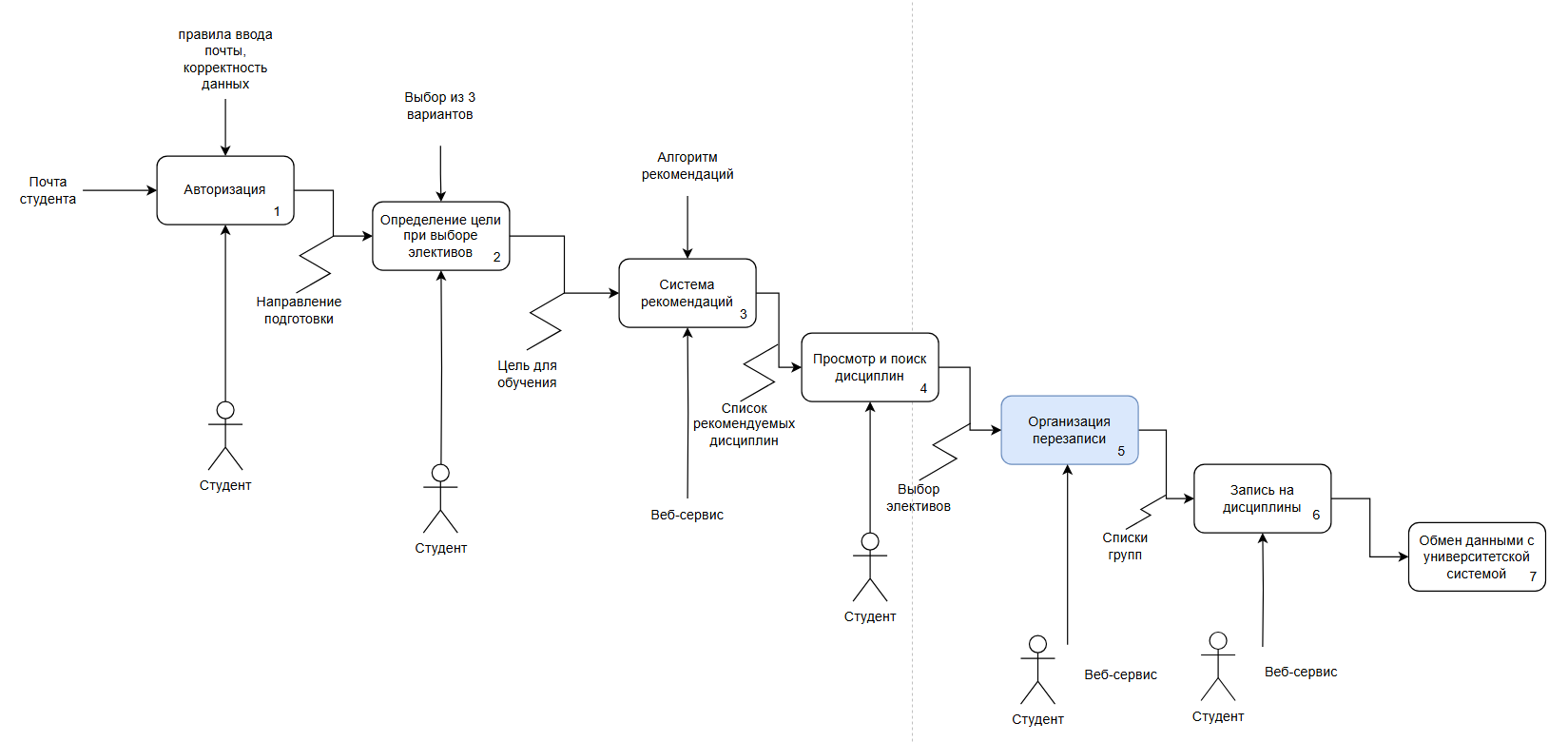
## 

## 

## 

## ГЛАВА 8. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ПРОЕКТА

На основании цели и задач была определена функциональная диаграмма, представленная на рисунке 5.



*Рисунок 5. Функциональная диаграмма.*

### 8.1 ОСНОВНЫЕ ФУНКЦИИ

1. Авторизация:
   1. Пользователь вводит свою корпоративную почту, а потом вводит код пришедший на нее.
   2. После успешного ввода кода, пользователь считается авторизированным.
2. Определение цели при выборе элективов:
   1. Обучающемуся предлагается выбор из 3-х целей: Хочу углубиться в свою специальность, хочу расширить кругозор, хочу отдохнуть.
   2. Цель можно поменять в профиле, но нельзя пользоваться приложением без цели.
3. Рекомендательная система:
   1. Предоставление пользователям персонализированных рекомендаций на основе их профиля и предпочтений(результатов анкетирования).
4. Просмотр и поиск дисциплин:
   1. Пользователь может искать дисциплины вводя их название в строку поиска.
   2. Пользователь может просматривать информацию об элективном курсе (описание, количество мест, преподаватель, расписание).
   3. Фильтровать отображаемые курсы по темам.
5. Организация перезаписи:
   1. Подробно описано в главе 6.
6. Запись на дисциплины
7. После подтверждения менеджера, система засчитывает перезапись и обновляет данные.
8. Система добавляет данные в отчет для операторов, для удобного последующего экспорта результатов.
9. Обмен данными с университетской системой
10. Данные о результатах кампании можно выгрузить в формате xlsx, где указаны пользователи и их выбор по перезаписи электива.

### 8.2 ДОПОЛНИТЕЛЬНЫЕ ФУНКЦИИ

1. Администрирование системы:
   1. Загрузку данных для начала кампании перезаписи.
   2. Отправка ответа на заявку по перезаписи на элективные курсы.
   3. Управление сроками начала и окончания кампании.
2. Аналитика и отчётность:
   1. Генерация отчётов о распределении обучающихся по элективам.

## 

## 

## 

## 

## ГЛАВА 9. ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС

### 9.1 ОСНОВНЫЕ ЭЛЕМЕНТЫ ИНТЕРФЕЙСА

Подробно описано в п. 7.

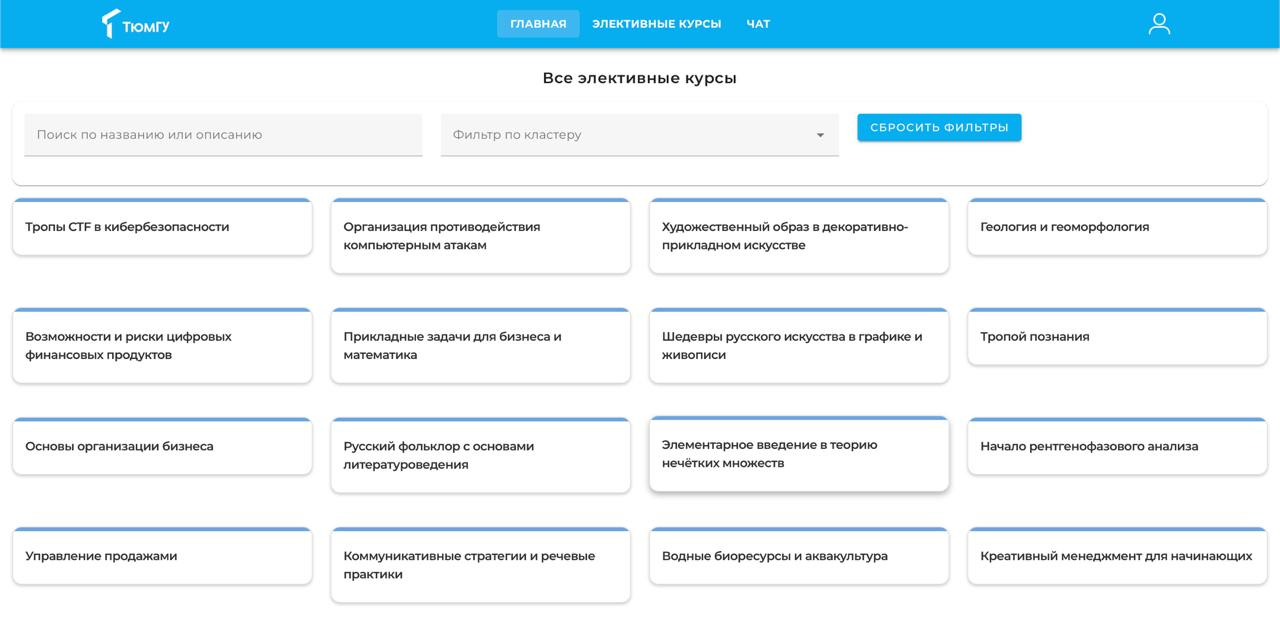
### 9.2 ДИЗАЙН ИНТЕРФЕСА

1. Интуитивно понятный и простой дизайн:
   1. Лёгкость навигации и поиска курсов.
   2. Чёткое отображение информации о кластерах и рекомендациях.
2. Адаптивный дизайн:
   1. Оптимизация интерфейса для различных устройств и экранов (десктоп, планшет, мобильный).
3. Интерактивность:
   1. Возможность взаимодействия с графиками и диаграммами для углублённого анализа.
   2. Реализация фильтров и сортировок для персонализации отображаемых курсов.

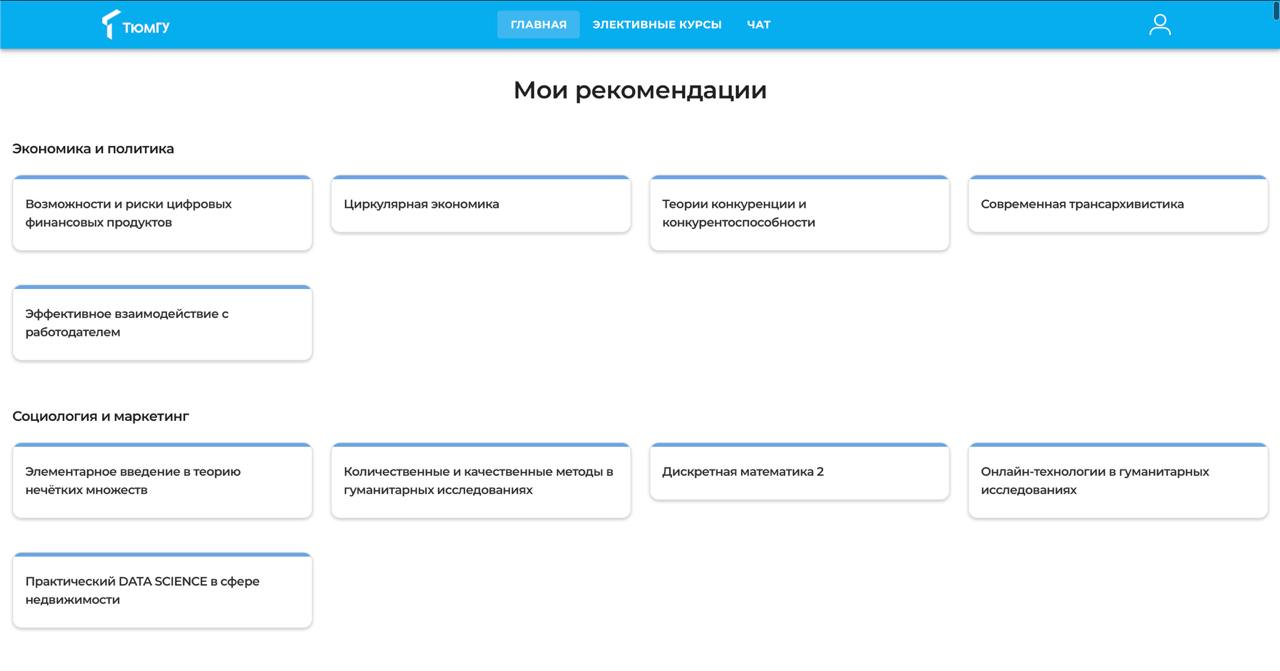
### 

### 9.3 ПРИМЕРЫ ИНТЕРФЕЙСА

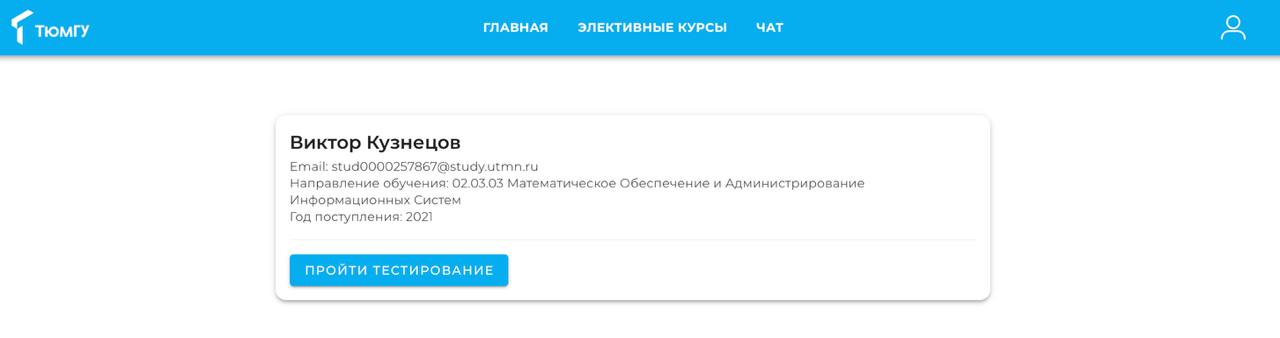
Главная страница:

  
*Рисунок 3. Главная страница.*

Страница рекомендаций:

*Рисунок 4. Страница рекомендаций.*

Личный кабинет пользователя:

  
*Рисунок 5. Личный кабинет пользователя.*

## 

## ГЛАВА 10. ОРГАНИЗАЦИЯ ТЕСТИРОВАНИЯ ПРОЕКТА

### 10.1 МЕТОДЫ ТЕСТИРОВАНИЯ

1. Этап тестирования рекомендаций:

На этом этапе цель — оценить качество и релевантность рекомендаций, предоставляемых системой.

1. Небольшая группа студентов, представляющая различные курсы и специализации.
2. Предоставить тестовой группе доступ к функционалу рекомендательной системы. Студенты вводят свои предпочтения или информацию о ранее пройденных курсах. Система генерирует рекомендации на основе алгоритма (TF-IDF + K-Means).
3. Студенты заполняют анкету с обратной связью. Примеры вопросов: Насколько рекомендации соответствовали вашим интересам? Что было неудобно в интерфейсе? Какие улучшения вы бы предложили?
4. Этап тестирования веб-сервиса в реальных условиях:

На этом этапе цель — проверить, насколько эффективно веб-сервис работает в сравнении с привычным инструментом (Яндекс Форма).

1. Выборка студентов (не менее 20-30 человек) из общей аудитории, которым ранее предлагали использовать Яндекс Форму для подачи запросов на перезапись.
2. Разделить студентов на две группы:  
   Контрольная группа: Использует Яндекс Форму, как это делалось ранее.  
   Экспериментальная группа: Использует веб-сервис для подачи заявок, просмотра информации о курсах и получения рекомендаций.
3. Пользователи заполняют анкеты обратной связи:   
   Насколько удобным был интерфейс веб-сервиса? Легко ли было найти нужные курсы? Насколько полезны были рекомендации?

## ЗАКЛЮЧЕНИЕ

Разработанная система частично решает задачу автоматизации процесса перезаписи на элективные курсы, предоставляя пользователям персонализированные рекомендации. Для этого использовались методы машинного обучения, такие как кластеризация и тематическое моделирование. В ходе проверки прототипа было подтверждено, что система корректно анализирует данные курсов и предпочтения пользователей, предлагая релевантные варианты.

На данный момент реализована только часть системы, отвечающая за категоризацию курсов и формирование рекомендаций. Функционал автоматизированной перезаписи на курсы, включая проверку ограничений доступности, находится в стадии разработки. Для подтверждения эффективности и удобства всей системы планируется проведение тестирования с участием конечных пользователей, что позволит оценить точность рекомендаций и надежность механизма перезаписи.

## 

## 

## СПИСОК ЛИТЕРАТУРЫ

1. Вьюгин В. В. Математические основы машинного обучения и прогнозирования. — М.: МЦНМО, 2013. — 256 с.
2. Домингос П. Верховный алгоритм. Как машинное обучение изменит наш мир. — М.: Манн, Иванов и Фербер, 2016. — 352 с.
3. Траск Э. Грокаем глубокое обучение. — СПб.: Питер, 2019. — 448 с.
4. Лейн Х., Хапке Х., Коул К. Обработка естественного языка в действии. — СПб.: Питер, 2020. — 400 с.
5. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по данным. — СПб.: Питер, 2017. — 392 с.