

С. А. Коровин, В. А. Бородина, К. В. Кочеткова,
К. А. Слободяник, М. В. Аврискин

Тюменский государственный университет, г. Тюмень

УДК 004.855

ВЫБОР ЭЛЕКТИВНЫХ ДИСЦИПЛИН В ТЮМГУ НА ОСНОВЕ ЦИФРОВЫХ ПОРТРЕТОВ СТУДЕНТОВ, СФОРМИРОВАННЫХ НА БАЗЕ АНАЛИЗА ПОДПИСОК ПОЛЬЗОВАТЕЛЯ В VK

***Аннотация.** В статье представлен процесс разработки системы рекомендаций, основанной на анализе взаимосвязи цифровых портретов студентов, сформированных на базе подписок на сообщества в социальной сети «VKонтакте», и выбираемых ими элективных дисциплин, для помощи обучающимся вузов в построении индивидуальной образовательной траектории.*

***Ключевые слова:** анализ данных, рекомендательная система, машинное обучение, классификация, кластеризация, коллаборативная фильтрация, индивидуальная образовательная траектория, элективная дисциплина, цифровой портрет.*

Введение. В связи с последними актуальными тенденциями персонализации высшего образования, наблюдаемыми в образовательном пространстве России [1] и закрепленными в качестве одного из стратегических направлений развития отечественного образования в нормативно-правовых документах РФ¹, многие вузы нашей страны внедряют индивидуальные образовательные траектории (ИОТ).

¹ Закон об образовании вводит такое понятие, как индивидуальный учебный план, который представляет собой «учебный план, обеспечивающий освоение образовательной программы на основе индивидуализации ее

Тюменский государственный университет был одним из первых вузов России, начавшим в 2017 г. выстраивать новую образовательную модель [2], среди важных нововведений которой стало непосредственное участие студента в формировании своей учебной программы, реализуемое, в том числе, за счет выбираемой части дисциплин — элективных курсов (Electives). К 2022 г. в ТюмГУ представлено более 600 таких курсов в различных областях знаний.

В таком многообразии студенты порой испытывают трудности в выборе интересных для себя элективных дисциплин, несмотря на большое количество подходов для их выбора [3].

Методы решения. Вследствие постоянного роста информационных потоков образовательного процесса и возможности получения детализированной информации о деятельности обучающихся возникает необходимость использования методов и технологий «больших данных» (Big Data) для оперативного управления образовательным процессом [4]. Сотрудники ТюмГУ уже давно занимаются разработками и исследованиями моделей прогнозирования для таких задач, как выбор образовательной программы (ОП) для абитуриентов, формирования команды для групповой проектной работы и выбора элективных курсов [5].

В качестве подходов к подбору дисциплин на основе анализа цифрового (информационного) следа студентов использовались поиск на основе текстового запроса за счет «предварительной кластеризации текстов программ, разметки текстов и построения классификатора для быстрого поиска подходящей дисциплины в соответствующем классе» [6] и формирование списка рекомендованных курсов на основе выбранных областей знаний [7].

В рамках данной работы предлагается решение задачи выбора элективных дисциплин, основанное на подборе таких курсов исходя из интересов студентов, вычисленных на базе анализа подписок на сообщества в социальной сети «ВКонтакте», и методе коллаборативной

содержания с учетом особенностей и образовательных потребностей конкретного обучающегося» (Федеральный закон от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации»).

фильтрации, отличающееся от решения подобной задачи, описанного в работе И. Г. Захаровой, М. С. Воробьевой, Ю. В. Боганюк [5].

Постановка задачи. Разработать систему рекомендаций, основанную на анализе взаимосвязи цифровых портретов студентов, сформированных на базе информационного следа в виде подписок на сообщества в социальной сети «ВКонтакте», и выбираемых ими элективных дисциплин, и реализовать интерфейс для взаимодействия с конечным пользователем в виде чат-бота в одноименной соцсети для помощи в построении индивидуальной образовательной траектории.

Предварительный этап: сбор данных

Для обучения системы рекомендаций необходимо собрать данные о студентах и уже пройденных ими элективах из платформы управления индивидуальными образовательными траекториями в университетах и ДПО ИС “Modeus”, а также сопоставить выгруженных студентов с их профилями в социальной сети «ВКонтакте» для дальнейшего анализа подписок и формирования цифровых портретов.

Выгрузка всех необходимых данных из ИС “Modeus” была реализована с помощью скрипта на языке программирования Python с использованием библиотеки Selenium — инструмента для автоматизации действий в веб-браузере, таких как заполнение полей форм, нажатие кнопок, поиск элементов на странице.

В результате были получены следующие наборы данных:

1) «элективы» — данные обо всех элективах: наименование, ссылка на модуль учебной программы (МУП), категория, список посетивших курс студентов;

2) «студенты» — данные о студентах: ФИ студентов — с удалением совпадающих пар «фамилия — имя», для того чтобы избежать ошибок при поиске профилей «ВКонтакте».

Для того чтобы сформировать цифровые портреты студентов, необходимо найти их профили в соцсети «ВКонтакте», с этой целью был написан скрипт, работающий с инструментарием для разработчиков VK API [8] на языке программирования Python.

Для уменьшения времени работы скрипта и числа запросов на сервер поиск профилей студентов реализован двумя методами:

1) поиск среди подписчиков сообществ ТюмГУ в VK: с помощью id каждого сообщества выполняется запрос для получения всех подписанных на него пользователей, после этого из собранного массива данных удаляются пары «фамилия, имя: id профиля» с повторяющимися фамилией и именем. Далее выполняется поиск всех полученных из ИС “Modeus” студентов в собранном массиве данных;

2) глобальный поиск по фамилии, имени и коду университета (в соцсети «ВКонтакте») для определения университета используется текстовое значение и идентификатор, который был получен с помощью метода database.getUniversities() [8]) выполняется для всех студентов, id профилей которых не были найдены предыдущим методом.

С целью хранения полученных данных, необходимых для обучения и корректной работы системы рекомендаций, была создана БД с использованием СУБД PostgreSQL, состоящая из 4 главных и 2 вспомогательных таблиц, диаграмма представлена на рис. 1.

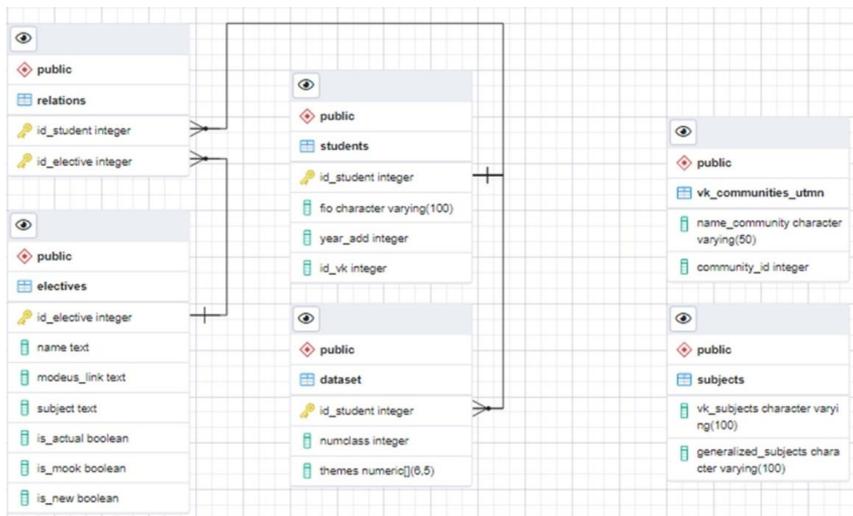


Рис. 1. Структура БД

Разработка системы рекомендаций

Система рекомендаций работает в три этапа:

1) распределение студентов из обучающей выборки по кластерам, исходя из долей подписок по обобщенным тематикам сообществ VK;

2) определение пользователя в одну из полученных на предыдущем шаге групп в соответствии с анализом его интересов;

3) создание перечня рекомендованных пользователю элективов.

На первом этапе близкие тематики сообществ «ВКонтакте» вручную объединяются в обобщенные тематики, при этом категория «Юмор» и тематики, связанные с университетом, были исключены, так как они в равной степени присутствуют у большинства студентов. Далее средствами библиотеки для работы с VK API высчитываются доли подписок на сообщества по обобщенным тематикам от общего числа подписок каждого студента из выборки.

Для распределения студентов на группы использовался метод иерархического кластерного анализа [9; 436-445]. В качестве признаков были использованы доли подписок пользователя, вычисленные на предыдущем шаге.

В результате было получено разбиение на 61 кластер, при этом для представителей каждого были рассмотрены средние значения по всем признакам с целью проверки, что для того или иного кластера выражен некоторый интерес. Например, для кластера № 57 это автомобили, а для № 58 — спорт (рис. 2).

	Транспорт	Автомобили	Развлечение	Кино	Спорт	Дизайн	История	Медицина	Финансы	Еда	...
cluster											
1	0.000209	0.000846	0.017954	0.025217	0.008559	0.038928	0.006678	0.002249	0.000833	0.006829	...
2	0.000060	0.002939	0.018220	0.025839	0.015986	0.043290	0.011946	0.004753	0.003033	0.017078	...
3	0.000000	0.001656	0.016424	0.032455	0.016681	0.021650	0.011945	0.003931	0.003020	0.012955	...
4	0.000000	0.000000	0.065558	0.000000	0.012987	0.000000	0.061688	0.000000	0.000000	0.000000	...
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
...
57	0.000125	0.302098	0.019365	0.016161	0.035052	0.008423	0.013300	0.001351	0.004224	0.007897	...
58	0.000240	0.033985	0.008580	0.009097	0.517738	0.001934	0.001457	0.000000	0.007604	0.015628	...
59	0.000877	0.023972	0.018866	0.052304	0.274634	0.005565	0.009035	0.002410	0.005865	0.012920	...
60	0.000135	0.018957	0.021069	0.032885	0.179846	0.014197	0.014908	0.002492	0.005481	0.011337	...
61	0.000250	0.026739	0.021578	0.038653	0.134259	0.016524	0.008571	0.002232	0.005346	0.015017	...

Рис. 2. Скриншот фрагмента таблицы средних значений по всем признакам для каждого кластера

На втором этапе аналогичным образом анализируются подписки пользователя, которому необходимо выдать список рекомендованных элективов. Для определения группы, к которой следует отнести этого пользователя, используется классификация методом *k*-ближайших соседей [10; 67-71], реализованным в библиотеке *sklearn*.

Класс определяется с точностью в 75%, а в остальных 25% случаев пользователь попадает в группу, которая менее точно отражает его интересы, но поскольку представили этой группы оказались рядом с ним, у них все же есть общие интересы, и система рекомендаций подберет элективы в соответствии с ними.

На последнем этапе рассматриваются только те элективные дисциплины, которые выбирал хотя бы один представитель группы, в которую был определен пользователь. В результате работы система рекомендаций формирует список элективов, отсортированный в порядке невозрастания по частоте выбора студентами этой группы.

Для разработки системы рекомендаций был выбран язык программирования Python актуальной версии 3.10, для работы использованы библиотеки *numpy*, *pandas*, *sklean*, *scipy*, *vk_api* и *psycorg2*.

Пользовательский интерфейс

В качестве интерфейса для взаимодействия пользователя с созданной системой рекомендаций был выбран чат-бот, реализованный в социальной сети «ВКонтакте». Для его работы был написан скрипт на языке программирования Python с использованием библиотеки *vk_api*.

Так как целевой аудиторией программного продукта являются студенты, обучающиеся на программах бакалавриата и специалитета ТюмГУ, и абитуриенты, необходимо было проработать две отдельные ветки: для студентов и для абитуриентов (рис. 3).

Проанализировав подписки пользователя, при условии, что его профиль в соцсети «ВКонтакте» открыт, система рекомендаций выдаст 10 самых релевантных элективных курсов на основе анализа частоты выбора тех или иных дисциплин среди студентов группы, к которой отнесен текущий пользователь.

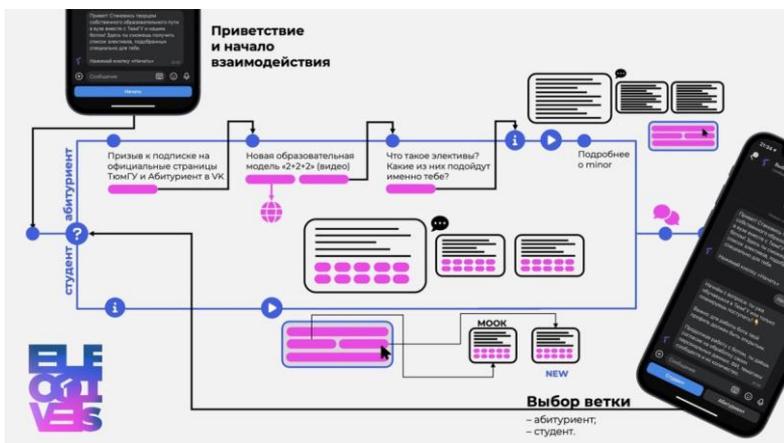


Рис. 3. Схема работы чат-бота с ветками

У пользователя имеется возможность дополнительно запросить 20 дисциплин из числа следующих по частоте выбора. Также система рекомендаций предложит пользователю интересные ему дисциплины, реализуемые в формате MOOK и 10 новых элективов, которые были созданы и представлены к выбору студентам в текущем году (рис. 4).



Рис. 4. Выдача результатов работы системы рекомендаций

Заключение. В результате работы из ИС “Modeus” были выгружены данные о 562 элективных курсах и выбиравших их 5178 студентах 2-3 курса за последние 4 семестра обучения, для 3209 из которых были найдены профили в социальной сети «ВКонтакте» и сформированы цифровые портреты на базе анализа подписок на сообщества в этой соцсети.

При тестировании системы рекомендаций на малом круге студентов были успешно предложены элективные курсы, исходя из интересов, вычисленных на базе анализа подписок на сообщества VK. Есть предположение, что при увеличении объема или изменении наборов данных, можно добиться более точных рекомендаций.

Данная разработка может быть внедрена и в других вузах страны, в том числе она будет актуальна вследствие перехода на новую образовательную модель «2+2+2» в рамках государственной программы поддержки «Приоритет 2030», в которой Тюменский государственный университет также принимает участие [11].

Решение позволяет студенту получить вектор, иначе говоря направление, для дальнейшего рассмотрения интересных и полезных для него элективных дисциплин, что в свою очередь позволит упростить изучение большого по объему списка таких курсов. Возможно, данное решение сможет увеличить шанс просмотра и последующей записи на интересные студенту элективы, на которые он мог бы не обратить свое внимание.

Авторы работы считают, что необходимо продолжать исследования и разработки в данном направлении, так как проблема построения индивидуальной образовательной траектории остается такой же актуальной, а предложенное решение при должном подходе может дать положительный эффект.

СПИСОК ЛИТЕРАТУРЫ

1. Константинова Л. В. Результаты мониторинга информации о тенденциях развития высшего образования в мире и в России. Вып. 2. Образовательные инновации в вузе: индивидуализация и смешанное обучение / Л. В. Константинова, Н. Н. Гагиев, Е. А. Смирнова [и др.]. — Текст : электронный // Научно-исследовательский институт развития образования Российского экономического университета им. Г. В. Плеханова. — Москва, 2021. — URL: <https://www.rea.ru/>

- ru/org/managements/Nauchno-issledovatel'skijj-institut-razvitiya-obrazovaniya/ Pages/monitoringi_informacii_o_sostoyanii_vischego_obrazovaniya.aspx (дата обращения: 24.05.2022).
2. Федорова Н. К. Индивидуализация образования: модель Тюменского государственного университета / Н. К. Федорова. — Текст : непосредственный // EdCrunch Томск : материалы Международной конференции по новым образовательным технологиям, 29-31 мая 2019 г. — Томск : Издательский дом Томского государственного университета, 2019. — С. 301-305.
 3. Индивидуальная образовательная траектория vs Традиционная образовательная модель в вузе. — Текст : электронный // Forbes Education Russian : [сайт]. — 2020. — 31 дек. — URL: <https://education.forbes.ru/special-projects/iot-main/iot-unmn-vs> (дата обращения: 20.05.2022).
 4. Захарова И. Г. Big Data и управление образовательным процессом / И. Г. Захарова — Текст : непосредственный // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. — 2017. — Т. 3, № 1. — С. 210-219.
 5. Захарова И. Г. Сопровождение индивидуальных образовательных траекторий на основе концепции объяснимого искусственного интеллекта / И. Г. Захарова, М. С. Воробьева, Ю. В. Боганюк. — Текст : непосредственный // Образование и наука. — 2022. — Т. 24, № 1. — С. 163-190.
 6. Захарова И. Г. Цифровой след студента: от данных к прогнозам и рекомендациям / И. Г. Захарова, М. В. Аврискин. — Текст : непосредственный // Информатизация образования и методика электронного обучения: цифровые технологии в образовании : материалы V Международной научной конференции : в 2 ч. (Красноярск, 21-24 сентября 2021 г.) / под общ. ред. М. В. Носкова. — Красноярск : Сибирский федеральный университет, 2021. — С. 120-124.
 7. Воробьева М. С. Диагностика предпочтений студентов при проектировании индивидуальных образовательных траекторий / М. С. Воробьева, М. Н. Первалова. — Текст : непосредственный // Информатизация образования и методика электронного обучения: цифровые технологии в образовании : материалы V Международной научной конференции : в 2 ч. (Красноярск, 21-24 сентября 2021 г.) / под общ. ред. М. В. Носкова. — Красноярск : Сибирский федеральный университет, 2021. — С. 60-63.

8. Документация по работе с VK API. — Текст : электронный // VK для разработчиков : [сайт]. — 2022. — URL: <https://dev.vk.com/reference> (дата обращения: 25.05.2022).
9. Рашка С. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow2 / С. Рашка, В. Мирджалили ; пер. с англ. — 3-е изд. — Санкт-Петербург : ООО «Диалектика», 2020. — 848 с. — Текст : непосредственный.
10. Lantz В. Machine Learning with R / В. Lantz. — Birmingham : Packt Publishing Ltd., 2013. — 396 p. — Text : direct.
11. Мультипрофильное образование по модели «2+2+2» в ТюмГУ. — Текст : электронный // Тюменский государственный университет : [сайт]. — 2022. — URL: <https://www.utmn.ru/obrazovanie/priority-education/> (дата обращения: 27.05.2022).

**Д. Д. КРИВОРОГОВ, А. А. ФАЗЛЫЕВ, А. Н. ХОДЫРЕВ,
Д. В. ШУШАРИН, А. А. СТУПНИКОВ**

Тюменский государственный университет, г. Тюмень

УДК 004.92

СИСТЕМА ВИЗУАЛИЗАЦИИ ХРОНОЛОГИЧЕСКИХ ЛЕНТ

***Аннотация.** В работе представлено описание реализации программного продукта, решающего проблему визуализации данных, расположенных в хронологическом порядке.*

***Ключевые слова:** хронологические ленты, система визуализации, события, анализ событий.*

Введение. Структурирование информации — это один из важных способов ее представления и анализа. Нередким является накопление информации в объемах, не поддающихся визуальному представлению стандартными средствами. В таких случаях исследователи прибегают к использованию специальных инструментов, способствующих оптимальному распределению информации на экране. Одним из таких инструментов может стать комплекс хронологических лент, содержащих последовательные события на временном интервале. Основной особенностью такого инструмента является содержание в компактном формате текста, изображения, видео и прочие файлы.