

ANALYSIS OF THE INTERESTS OF APPLICANTS ON THE VKONTAKTE SOCIAL NETWORK TO SELECT A FACULTY

**Medvedeva T.S. (Russian Federation)
Email: Medvedeva521@scientifictext.ru**

*Medvedeva Tatyana Sergeevna – Student,
DEPARTMENT OF FUNDAMENTAL INFORMATICS, FACULTY OF
MATHEMATICS AND INFORMATION TECHNOLOGIES,
MORDOVIAN STATE UNIVERSITY N.P. OGAREV, SARANSK*

Abstract: *the relevance of the chosen topic is due to the fact that school graduates every year face a difficult choice of faculty and direction of training. But in the modern world, thanks to digital footprints, it is very easy to track what a particular person is interested in. Therefore, it would be interesting to trace the dependence of the interests of applicants in a social network on the interests of existing students of a particular faculty and to identify applicants with the highest level of interest for making a decision on the choice of the direction of study.*

Keywords: *data analysis, digital footprints, social networks, qlik sense, api, vkontakte.*

АНАЛИЗ ИНТЕРЕСОВ АБИТУРИЕНТОВ В СОЦИАЛЬНОЙ СЕТИ ВКОНТАКТЕ ДЛЯ ВЫБОРА ФАКУЛЬТЕТА

Медведева Т.С. (Российская Федерация)

*Медведева Татьяна Сергеевна – студент,
кафедра фундаментальной информатики, факультет математики и
информационных технологий,
Мордовский государственный университет им. Н.П. Огарёва, г. Саранск*

Аннотация: *актуальность выбранной темы обусловлена тем, что выпускники школ каждый год сталкиваются со сложным выбором факультета и направления подготовки. Но в современном мире благодаря цифровым следам очень легко отследить, чем интересуется конкретный человек. Поэтому было бы интересно проследить зависимость интересов абитуриентов в социальной сети от интересов уже действующих студентов конкретного факультета и выявить абитуриентов с наиболее высоким уровнем заинтересованности для принятия решения о выборе направления обучения.*

Ключевые слова: *анализ данных, цифровые следы, социальные сети, qlik sense, api, вконтакте.*

Выпускникам школы бывает очень трудно определиться с выбором факультета для дальнейшего обучения. Каждый год, уже бывшие школьники выбирают с каким направлением связать свою жизнь. Чаще всего факультет выбирают по престижности, совету знакомых, рейтингу и в том числе по собственным интересам. В цифровом мире не сложно узнать, чем интересуется конкретный человек хотя бы на примере страницы Вконтакте. На своей странице люди добавляют понравившуюся музыку, видео, вступают в группы, которые им интересны. А что, если для решения проблемы выбора факультета использовать информационные технологии и анализировать страницу Вконтакте на предмет зависимости интересов пользователя с необходимыми интересами для конкретного факультета? Это так же поможет и университету в подборе новых кадров. Исследование этой проблемы является актуальным, так как выпускники школ сталкиваются с этим каждый год и было бы интересно проследить зависимость их интересов в социальной сети со сферой деятельности конкретного факультета.

Узнавать предрасположенность абитуриентов к факультету будем на основе их интересов, а именно по сообществам в социальной сети Вконтакте, на которые они подписаны. За основу будет выбран факультет математики и информационных технологий МГУ Н.П. Огарева. Получать данные будем с помощью API Вконтакте. Выгрузим друзей автора статьи и сделаем пометки тем пользователям, которые точно являются студентами выбранного факультета. Для выгрузки друзей используется метод *friends.get*, а для получения имени и фамилии метод *users.get*. Всего было выгружено 308 пользователей Вконтакте, из них 94 (30.5%) определены как студенты либо имеющие какое-либо отношение к факультету (рисунок 1).

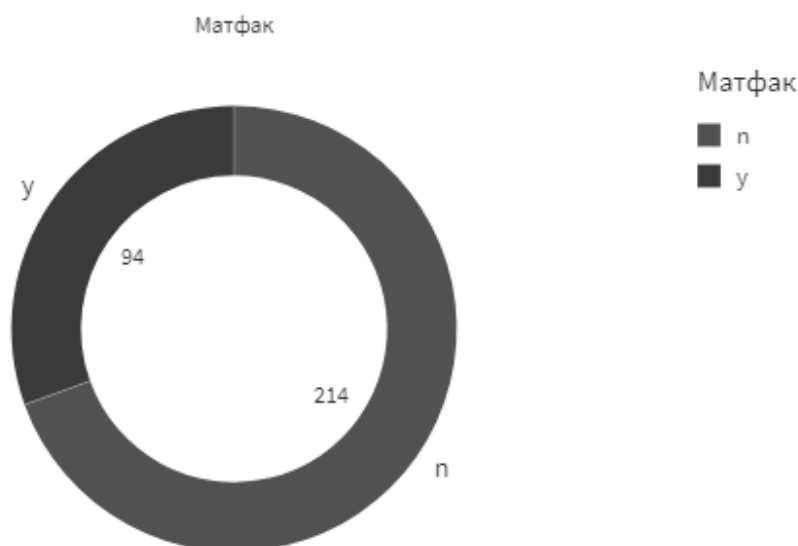


Рис. 1. Распределение количества пользователей по факультету

Выберем только пользователей с пометкой Матфак = 'у' и получим по ним сообщества. Для этого используем метод *groups.get*, в качестве параметра он принимает идентификатор пользователя. Так как некоторые аккаунты являются закрытыми или не содержат ни одной подписки, то такие пользователи были пропущены. В результате для анализа используются группы 81 студента. Проанализируем полученные данные. Найдем самые популярные сообщества (таблица 1). Всего было выгружено 5 801 уникальных групп.

Таблица 1. Топ 15 популярных групп среди студентов

Название группы	Количество пользователей
Привет, сейчас, Саранск.	41
Огарёвские цитаты	38
Новости Балагана	34
Бот-расписание МГУ им.Н.П.Огарева	30
Типичный программист	23
САРАНСК Онлайн	22
SMART-place Кубик Рубикова	22
«Столица С»	21
Библиотека программиста	20
SimbirSoft	20
/dev/null	19
Пикабу	19
MDK	19
Будет в Саранске. Афиша Саранска	18
Киномания	18

При поиске сообществ, по ключевым словам, 77 из 5 801 связаны с математикой, программированием и IT (таблица 2).

Таблица 2. Топ 15 групп среди студентов, связанных с математикой, программированием и IT

Название группы	Количество пользователей
Типичный программист	23
Библиотека программиста	20
Типичный математик	13

ITc сообщество программистов	12
Подслушано у программиста	10
Программирование ITmozg:	10
Программист	10
IT-КОТ	9
Программирование	7
Уютное сообщество программистов	7
FoxDevs - IT-сообщество Саранска	7
TechRocks Программирование и IT новости	6
Дискретная математика	4
For Web — фронтенд, дизайн, программирование	4
ЕГЭ. Математика	3

Теперь выгрузим группы абитуриентов. Абитуриентами будем считать всех школьников 17-ти лет, проживающих в г. Саранск и имеющих заполненное поле со школой в их профиле. Используем метод *users.search*.

Получим 103 человека с учетом условия, что аккаунт является открытым. По аналогии с выгрузкой групп студентов выгрузим группы абитуриентов и запишем в файл. Выведем топ – 15 самых популярных групп (таблица 3). Всего удалось выгрузить 10 144 уникальные группы.

Таблица 3. Топ-15 групп потенциальных абитуриентов

Название группы	Количество пользователей
Привет, сейчас, Саранск.	53
Ищу тебя Саранск	34
«Столица С»	31
Леонардо Дайвинчик	29
Овсянка, сэр!	29
МЕМЫ	26
НОС: Новости о Саранске... и Мордовии	25
Рифмы и Панчи	23
MDK	23
Саранск Доска Позора	22
Смейся до слёз :D	21

ВКонтакте	20
РДШ Мордовия	20
СигнаL	20
СТЫД	20

Объединим результаты и проанализируем, кто из абитуриентов наиболее совпадает по интересам со студентами математического факультета. Загрузим полученные результаты в программу Qlik Sense. Сначала соотнесем все группы, которые были выгружены у студентов математического факультета. Получим следующий результат (таблица 4):

Таблица 4. Топ-10 абитуриентов с совпавшими группами

Школа	Абитуриент	Количество совпадающих групп
Лицей №7	Анастасия Фомина	167
Лицей №7	Дмитрий Хлутчин	160
Школа №8	Юля Макаева	147
Школа №36	Вика Корчинская	98
Школа №25	Дмитрий Зорькин	97
Гимназия №23	Анастасия Барина	96
Школа №9	Артём Асташкин	90
Школа №8	Олеся Пискунова	87
Школа №27	Маша Евсеева	76
Школа №11	Виктор Белов	74

Если проанализировать это большое число совпадающих групп, то становится ясно что все эти сообщества не несут никакой пользы и направлены только на развлечения, поэтому опираться на данный результат не совсем верно.

Теперь соотнесем только профильные группы. Получим следующий результат (таблица 5):

Таблица 5. Все абитуриенты с совпавшими профильными группами

Школа	Абитуриент	Количество совпадающих групп	Название совпавшей группы
Лицей №7	Дмитрий Хлутчин	1	Программист

Исходя из текущей выборки данных, только 0.97% потенциальных абитуриентов имеют заинтересованность 1.3% (Отношение количества

совпавших профильных групп к количеству всех профильных групп) в сфере IT.

Список литературы / References

1. *Маккинни У.* Python и анализ данных / У. Маккинни Изд. 2-е. М.: Эксмо, 2013. 484 с.
2. VK API. [Электронный ресурс], 2019. // Режим доступа: https://vk.com/dev/first_guide/ Заголовок с экрана/ (дата обращения: 17.05.2021).
3. *Бейли Л.* Изучаем SQL / Л. Бейли. СПб. : Питер, 2012. 592 с.
4. *Новиков Б.А.* Основы технологий баз данных / Б.А. Новиков, Е.А. Горшкова. М.: ДМК Пресс, 2019. 240 с.