

*Р.М. Якубов<sup>1</sup>, С.М. Бильдин<sup>1</sup>, М.Д. Дроздецкий<sup>1</sup>, М.С. Воробьева<sup>1,2</sup>*

<sup>1</sup> Тюменский государственный университет, г. Тюмень

<sup>2</sup> Научно-технический университет «Сириус», г. Сочи

УДК 004.912

## РАЗРАБОТКА ВЕБ-ПРИЛОЖЕНИЯ ДЛЯ ФОРМИРОВАНИЯ ЦИФРОВОГО ПРОФИЛЯ СТУДЕНТА

**Аннотация.** В статье рассматривается метод формирования цифрового профиля студента на основе полученных данных из социальной сети «ВКонтакте». Исследован словарный подход к классификации сведений о студентах, которые интересуются программированием. В результате спроектировано и разработано веб-приложение для формирования цифрового профиля студента, проведен анализ интересов студентов ИМиКН различных курсов и направлений.

**Ключевые слова:** сбор данных, извлечение данных, анализ данных, классификация, Python, ВКонтакте.

Отсутствие специализированных инструментов для организации и управления студенческими мероприятиями способствует возникновению ряда проблем, в частности определение круга участников события. Организаторы сталкиваются с проблемой поиска целевой аудитории, вследствие чего событие часто обходит стороной компетентных участников.

Поэтому необходимо собрать сведения о потенциальных лицах с целью формирования круга единомышленников, хорошо ориентирующихся в той или иной сфере жизнедеятельности. Для формирования *цифрового профиля* необходимо собрать записи о человеке,

содержащих личную информацию, сведения об интересах, социальную или профессиональную роли.

Студенческие данные можно получить из социальной сети «ВКонтакте» [1], анализируя подписки, записи на стене пользователя и те записи, на которых когда-либо упоминался данный профиль. Очевидно, что студенты, интересующиеся программированием, состоят в IT-группах, читают новости, выставляют свои посты, ставят комментарии к интересующим текстам. Аспектов и направлений в программировании много, А. В. Васеневым в статье [2] была выделена сфера IT-интересов, распределенных по 8 группам:

1. *competitive* – спортивное программирование;
2. *data science* – наука о данных;
3. *desktop* – разработка оконных приложений;
4. *embedded* – программирование встраиваемых систем;
5. *gamedev* – разработка игр;
6. *mobile* – разработка приложений для мобильных устройств;
7. *system administration* – системное администрирование;
8. *web* – разработка веб-приложений.

При выполнении классификации данных из социальной сети «ВКонтакте» был использован словарный подход [3]. Главное требование к классификации – формирование корпуса ключевых слов, которые будут относиться к определенной группе.

Структура словаря:

$$\{ \begin{array}{l} keyword_1: [tag_1, tag_2, tag_3, \dots, tag_p] \\ keyword_2: [tag_1, tag_2, tag_3, \dots, tag_q] \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ keyword_N: [tag_1, tag_2, tag_3, \dots, tag_z] \end{array} \},$$

где  $N$  – размер словаря (количество ключевых слов);

$P, Q, \dots, Z \in \{1, 2, \dots, 8\}$  – количество тегов в списке.

Требуемая информация из социальной сети «ВКонтакте» представляет текстовые данные, поэтому к каждому тексту применяется следующая обработка:

- 1) перевод всех слов в нижний регистр;
- 2) удаление всех цифр и знаков пунктуации;
- 3) удаление стоп-слов (частицы, предлоги, союзы и т. п.);
- 4) разбиение текста на список униграмм.

В результате текстовые данные представляют «мешок слов»  $[w_1, w_2, \dots, w_n]$  – многомерный вектор, каждый элемент которого представляет строку (униграмму) [4].

Рассмотрим алгоритм классификации данных профиля студента из социальной сети «ВКонтакте» (см. рис.1).

Этап 1. Получение методами API VK списка всех подписок студента, текста записей со стены пользователя и записей, на которых был отмечен;

Этап 2. Нормализация полученных на этапе 1 текстовых данных и последующая классификация этих данных, за исключением списка подписок;

Этап 3. Получение текста названия и описания группы методами API VK и нормализация полученных данных;

Этап 4. Повторение этапа 3 для каждой группы из списка подписок.

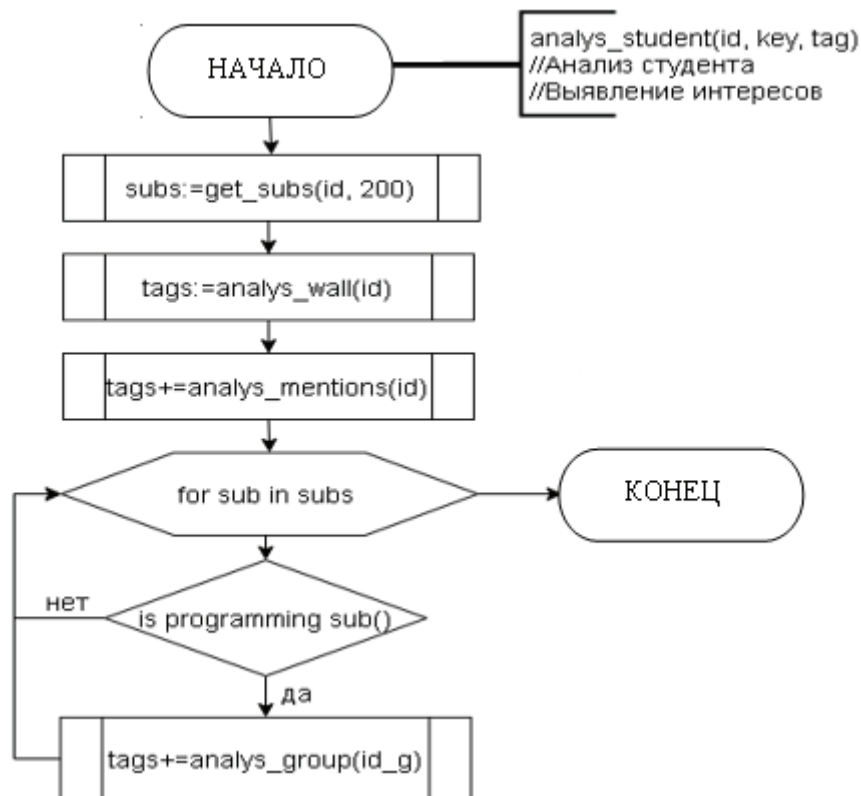


Рис. 1. Блок-схема алгоритма классификации.

Веб-приложение было разработано на языке программирования Python. Стек технологий: Django – фреймворк для написания веб-приложений, HTML, CSS, Bootstrap4 – frontend технологии, SQLite3 - СУБД, интерфейс VK API. Архитектура приложения включает 4 функциональных блока: страница создания аккаунта, страница регистрации, страница профиля студента и классификатор (см. рис.2).

На странице создания аккаунта (Register page) пользователь вводит адрес электронной почты, ссылку на страницу «Вконтакте». На странице создания профиля (Profile create page) заполняет основные данные (ФИО, дата рождения, пол, контактная информация, сведения об образовании). Во время заполнения информации в фоновом режиме запускается алгоритм классификации. После регистрации пользователь перенаправляется на страницу профиля (Profile page), где находится вся введенная информация и выявленные в результате анализа страницы «Вконтакте» навыки.

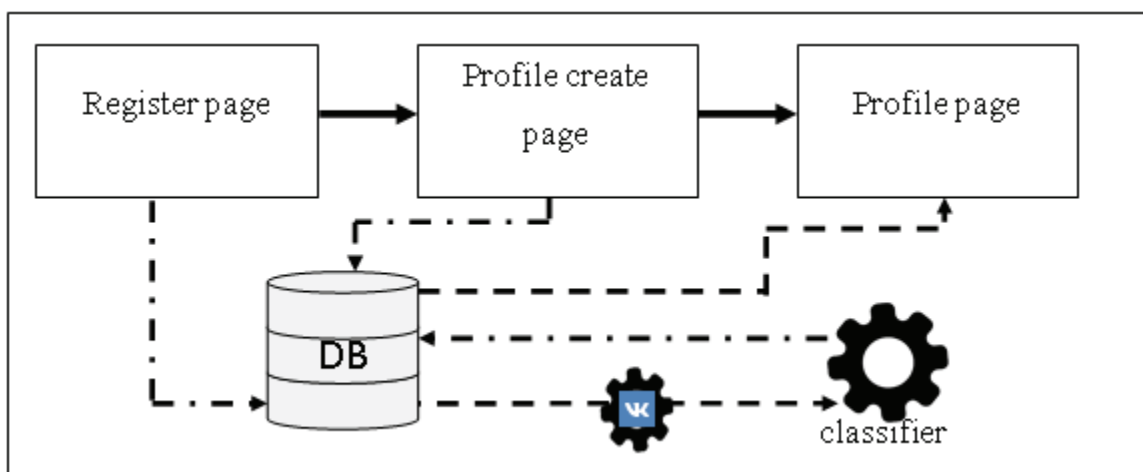


Рис. 2. Архитектура веб-приложения

Во время сбора информации из социальной сети были отобраны 342 студента Института математики и компьютерных наук следующих направлений обучения «Компьютерная безопасность», «Информационная безопасность», «Прикладная информатика», «Математическое обеспечение и администрирование ИС». Были выявлены интересы каждого студента соответствующего направления. Всего получилось 138 записей, каждая из которых содержит долю студентов определенного направления обучения относительно общего числа обучающихся, распределенных по ИТ-группам, согласно выявленным интересам в результате классификации.

После проведения анализа были получены результаты о студентах 2 курса ИМиКН и о студентах всех курсов направления МОиАИС.

Сводные данные о студентах ИМиКН 2 курса направлений ПИ, КБ, ИБ, МОиАИС содержат следующую информацию (см. рис.3):

- 1) основная доля спортивных программистов (*competitive*) приходится на направление МОиАИС, доля которых с каждым годом только возрастает;
- 2) популярная область *Data science* одинаково актуальна для всех направлений;

3) различия в заинтересованности областями *mobile*, *system administration* и *gamedev* могут быть охарактеризованы спецификой учебной программы соответствующих направлений.

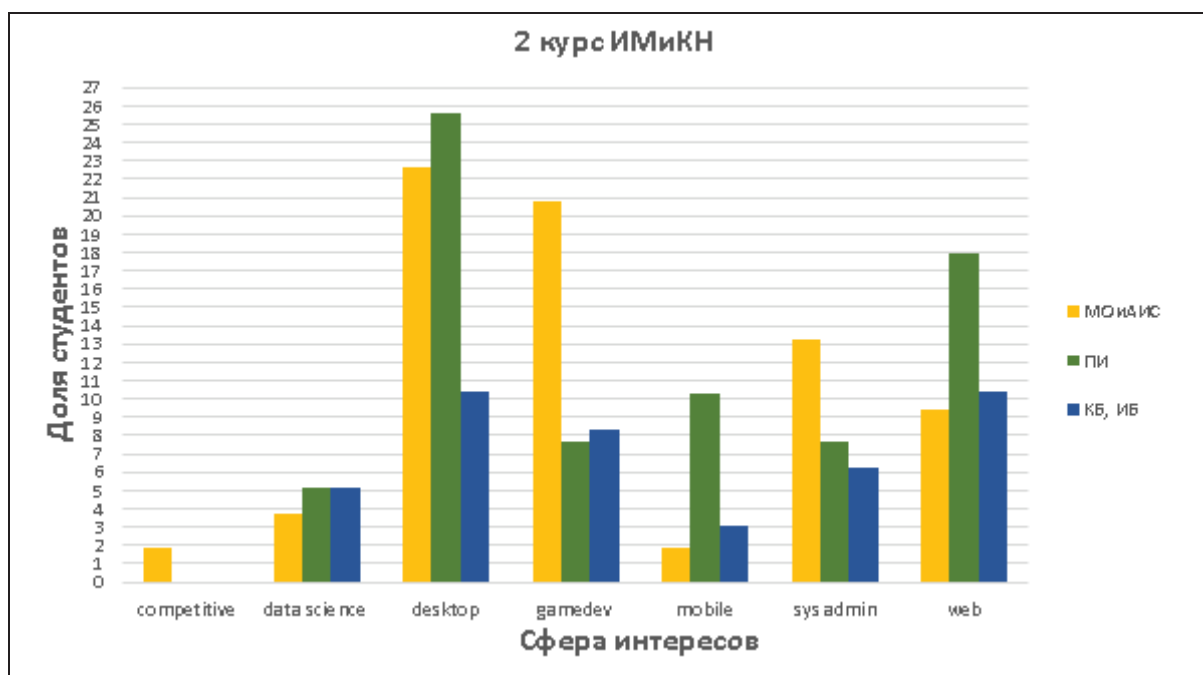


Рис. 3. Доля студентов 2 курса ИМиКН, распределенных по интересам.

Проведен анализ полученной информации по студентам направления «Математическое обеспечение и администрирование информационных систем» всех курсов обучения (см.рис.4):

- 1) интерес к *web*-программированию у студентов МОиАИС с каждым годом только возрастает, следует ожидать прирост *web*-программистов на данном направлении;
- 2) скачок в области *gamedev* у второго курса МОиАИС может быть охарактеризован тематикой курсовых работ за 2 семестр, которые были ориентированы на игровую разработку;
- 3) определение рекомендаций к подробному рассмотрению области мобильной разработки (возможно дополнения к учебной программе), может повлиять на устранение резких скачков в данной области, так как на данный момент это не менее популярное направление.

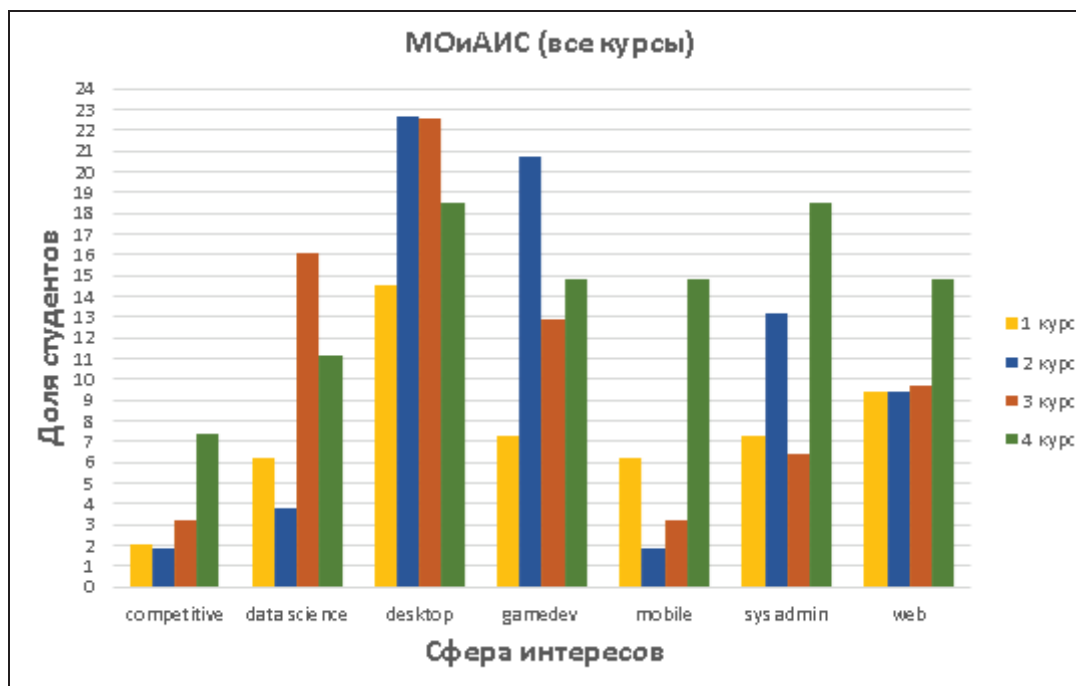


Рис. 4. Доля студентов МОиАИС, распределенных по интересам.

В результате было разработано веб-приложение для формирования цифрового профиля студента на основе данных, полученных из социальной сети, с использованием инструментов сбора информации со страницы «Вконтакте» и последующей классификацией собранных данных по основным направлениям программирования.

### Благодарности

Статья подготовлена в рамках разработки образовательного кейса для НТУ Сириус при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028.

### СПИСОК ЛИТЕРАТУРЫ

1. Сумкин, К. С. Анализ страницы пользователя социальной сети «Вконтакте» [Электронный ресурс]: / К. С. Сумкин, Л. О. Тараненко. – Текст: непосредственный // Молодой ученый. – 2016. – №12 (116). – С. 189-194. – Режим доступа: <https://moluch.ru/archive/116/318842>, (дата обращения: 19.04.2020)

2. Направления в программировании [Электронный ресурс]: / «Мое образование» — Электрон. дан. — 28.01.2016. — Режим доступа: [https://моеобразование.ru/napravleniya\\_v\\_programmirovanii.html](https://моеобразование.ru/napravleniya_v_programmirovanii.html), (дата обращения: 26.04.2020)
3. Воробьева, М. С., Дубаков, А. А. Разработка приложения для определения тематики текста с использованием алгоритмов кластеризации [Электронный ресурс]: / Воробьева М. С., Дубаков А. А. — Электрон. журн. — 2019. — №17: — С. 85-95. — Режим доступа: <https://www.elibrary.ru/item.asp?id=41380880>, (дата обращения: 4.05.2020).
4. Пайвин, Д. Н., Глазкова, А. В. Исследование методов векторного представления естественного языка на примере классификации коротких текстов [Электронный ресурс]: / Пайвин Д. Н., Глазкова А. В. — Электрон. журн. — 2019. — №17: — С. 104-113. — Режим доступа: <https://www.elibrary.ru/item.asp?id=41380883>, (дата обращения: 25.05.2020).