

# Разработка алгоритма поиска студента учебного заведения в экосистеме социальных сервисов на примере социальной сети “ВКонтакте”

Жариков А.В., Костырко А.В.  
*Алтайский государственный университет*  
*zharikov@math.asu.ru, kostirko3@mail.ru*

## Аннотация

В статье ставится задача разработать алгоритм идентификации студентов университета в социальной сети “ВКонтакте” и провести тестирование предложенного алгоритма с целью выявления степени применимости программных средств для подобного рода задач идентификации.

Системы идентификации находят все большее применение при решении различных технических задач для IT-специалистов в рамках компании, в которых необходимо определить принадлежность человека к какой-либо экосистеме социальных сервисов или других системах и сообществах.

Процесс идентификации может содержать несколько этапов, где в зависимости от сложности этапа, составляется свой алгоритм поиска свойств объекта для его полной идентификации. Причем, результат поиска свойств может быть пустым. Подобный результат возникает из-за неоднородности заполняемых данных пользователями экосистем социальных сервисов.

Таким образом, возникает задача разработать алгоритм идентификации человека или группы людей в социальных сетях на примере студентов учебных заведений и протестировать на реальных данных.

В качестве инструмента для поиска группы студентов в интернет-сообществах необходимо выбрать сервис, работа которого не противоречила действующему законодательству.

Проведен обзор и анализ существующих сервисов идентификации: “Яндекс Люди” [1], “FindFace” [2], “Поиск людей” [3] и популярных социальных сетей.

В результате, анализа функциональности предложенных сервисов были выявлены недостатки и преимущества, представленные в таблице 1.

Таблица 1

Основные недостатки и преимущества функциональности сервисов

Сервис	Поиск по признакам	Бесплатность использования	Поиск одновременно нескольких лиц	Точность определения
Яндекс Люди	да	да	нет	Высокая
FindFace	нет	нет	нет	Высокая
“Поиск людей”	нет	да	нет	Низкая

Приведенные результаты, показывают, что ни один из предложенных сервисов не удовлетворяет условиям поиска группы студентов одновременно.

На основе статистических данных за 2015 год возраст наиболее активной аудитории социальных сетей России составляет 17-24 года, что соответствует возрасту большинства студентов. Самая высокая популярность в данной возрастной категории соответствует

социальной сети “ВКонтакте”, которая официально предоставляет информацию о пользователях посредством API (application programming interface) при соблюдении правил платформы работы с данными пользователей.

Таким образом, был разработан алгоритм, схема которого включает в себя три этапа:

1. Идентификация студентов с заполненной информацией об образовании.
2. Идентификация студентов через друзей уже найденных студентов.
3. Выгрузка публичных данных идентифицированных студентов.

Рассмотрим более подробно действия каждого этапа.

Действия этапа 1:

- 1) выгрузка информации о студенте из локальной базы данных;
- 2) запрос к API социальной сети “ВКонтакте” на поиск профиля по фамилии, имени, городу. В качестве ответа API возвращает список идентификаторов найденных профилей пользователей в социальной сети.
- 3) запрос к API по идентификатору для получения информации о профиле
- 4) проверяется информация об образовании. Если информация соответствует информации из локальной базы данных, то переходим к действию 5), иначе выполняем возвращаемся к действию 3).
- 5) обновляется в локальной базе данных информация о найденном студенте.
- 6) выполняется действие 1) пока не закончится список студентов локальной базы данных.

Действия этапа 2:

- 1) выгрузка информации из локальной базы данных идентифицированного студента
- 2) отправляется запрос к API с выгрузкой списка друзей идентифицированного студента
- 3) проверяется информация по фамилии, имени, полученная от API с той информацией, что находится в локальной базе данных по всем студентам (аналогично пункту 4) этапа 1);
- 4) выполняется действие 1) пока не закончится список студентов, что находятся в локальной базе данных;

Действия этапа 3:

- 1) выгрузка информации из локальной базы данных идентифицированного студента;
- 2) запрос к API для выгрузки данных по идентификатору пользователя;
- 3) полученный url-адрес фотографии записывается в базу данных;
- 4) выполняется действие 1) пока не закончится список студентов

В качестве локальной базы данных выступает СУБД MySQL.

Предложенный алгоритм был реализован в виде web-сервиса, с помощью языка программирования PHP 5 и Open API 5.4 “ВКонтакте”.

Тестирование алгоритма было проведено на примере нескольких групп студентов АГУ и других учебных заведений города Барнаула. Результаты работы алгоритма представлены в таблице 2.

Таблица 2

## Результаты работы разработанного алгоритма

Название учебного заведения (факультет)	Номер группы	Количество студентов в группе	Правильно идентифицированные студенты	Эффективность алгоритма %	Ошибка %
АГУ (ФМиИТ)	432	22	19	86%	14%
АГУ (ФМиИТ)	434	14	14	100%	–
АГУ (ФТФ)	537	23	20	87%	13%
АГУ (ФТФ)	531	8	6	75%	25%
АГУ (ХФ)	631	20	19	95%	5%
АлтГТУ (Физико-технический)	–	22	18	82%	18%
РАНГиГС	–	19	14	74%	26%

В результате проделанных тестов видно, что алгоритм не всегда правильно идентифицирует студента. Данный факт был проанализирован для каждого случая и выявлены основные недостатки предложенного алгоритма:

1. Алгоритм не учитывает имена студентов, написанных в транслитерации, в виде псевдонима и других нестандартных форм записи имени.
2. Направление обучения студента в сети “ВКонтакте” не всегда соответствует названию из ФГОС.
3. Название вуза и факультета должны быть сопоставлены названиям в сети “ВКонтакте”.

В целом, эффективность алгоритма может считаться приемлемой и программный инструмент может быть внедрен для внутренних нужд предприятия или иного учреждения.

## Список литературы

1. Яндекс Люди. — URL: <https://yandex.ru/people>.
2. FindFace. — URL: <http://findface.ru/>.
3. Поиск людей. — URL: <http://poisklyudei.com/poisk-lyudey-vkontakte.html>.