Оглавление

[Реферат 5](#_Toc515282508)

[Введение 6](#_Toc515282509)

[1. Обзор предметной области поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282510)

[1.1 Описание предметной области поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282511)

[2. Обследование объекта сервиса поиск абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282512)

[2.1 Формулировка требований к сервису поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282513)

[2.2 Обзор аналогов сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282514)

[2.3 Проектирование сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282515)

[2.4 Проектирование логической модели сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282516)

[2.5 Идентификация сценариев сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282517)

[2.6 Проектирование классов сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282518)

[2.7 Разработка физической модели сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282519)

[3. Основания выбора средств разработки сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282520)

[3.1 Выбор СУБД для сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 6](#_Toc515282521)

[3.2 Выбор языка программирования сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 8](#_Toc515282522)

[4. Разработка сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 8](#_Toc515282523)

[5. Тестирование сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 8](#_Toc515282524)

[Заключение 8](#_Toc515282525)

[Список использованных источников 8](#_Toc515282526)

[Приложения 8](#_Toc515282527)

[Приложение А Техническое задание на создание сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 8](#_Toc515282528)

[Приложение Б Диаграмма вариантов использования 8](#_Toc515282529)

[Приложение В Диаграмма классов 8](#_Toc515282530)

[Приложение Г Диаграмма развёртывания 8](#_Toc515282531)

[Приложение Д Диаграмма базы данных сервиса поиска абитуриентов на базе технологий NLP и машинного обучения 8](#_Toc515282532)

# Реферат

Выпускная квалификационная работа, ##с., # рис., # табл., # источников, #прилож.

**Ключевые слова:** Python, NLP, база данных, #.

**Объектом** сервис для поиска абитуриентов.

**Цель работы** – разработка сервиса для поиска абитуриентов на базе технологий NLP и машинного обучения.

В процессе работы производилось определение требований к сервису и его реализация.

**Достигнутые технико-эксплуатационные показатели:** #

**Степень внедрения:** #

**Результат:** В результате разработки был получен полноценный работающий сервис поиска абитуриентов #.

# Введение

В 2008 году ученые из университета Хэриот-Уотт (Эдинбург) во главе с профессором Адрианом Нортом, главой кафедры прикладной психологии, решили проверить, связаны ли музыкальные предпочтения с интеллектом и характером слушателей. Их исследования выявили очень интересную особенность: самые высокие результаты IQ-тестов - у любителей тяжелой музыки и рока. На основе данных о том, что музыка взаимосвязана с уровнем интеллекта, вывели гипотезу, что она так же взаимосвязана с предрасположенностью к какому-либо направлению обучения. Для подтверждения данной гипотезы было разработано программное обеспечение, которое на основе выборки студентов по разным направлениям, находит зависимости между ними.

# Обзор предметной области поиска абитуриентов на базе технологий NLP и машинного обучения

## Описание предметной области поиска абитуриентов на базе технологий NLP и машинного обучения

Определения:

* **Natural language processing (NLP)** - общее направление искусственного интеллекта и математической лингвистики. Оно изучает проблемы компьютерного анализа и синтеза естественных языков. Применительно к искусственному интеллекту анализ означает понимание языка, а синтез — генерацию грамотного текста. [#]
* Таргетинг (англ. target — цель) — рекламный механизм, позволяющий выделить из всей имеющейся аудитории только ту часть, которая удовлетворяет заданным критериям (целевую аудиторию), и показать рекламу именно ей.

Между университетами происходит конкуренция за лучше умы поступающих. На сегодняшний день финансирование вузов зависит от результатов приемной компании. Поэтому для вуза важен каждый студент, но если этот студент учится на направлении, к которому у него нет предрасположенности, то вероятность его отчисления из ВУЗа велика. Поэтому для создания списка первоочередных абитуриентов нужен их анализ, который мы будем брать из социальной сети Вконтакте, на основе его профиля и музыки, которую он слушает.

Стратегия проста, выявить школьников с предрасположенностью к определенному направлению обучаемом в вузе и пригласить его на обучение.

Прямых аналогов не выявлено. Косвенные аналоги – программы для таргетинга.

Задачи:

* Выявить зависимость между музыкой исследуемого и его предрасположенностью к определенному виду деятельности.
* Создать список первоочередных абитуриентов по направлениям.

Актуальность: Тема актуальна, так как каждый ВУЗ (государственные и коммерческие) ведет борьбу за студентов.

«Худший формат дня открытых дверей, который можно придумать, — посадить в фойе тетушек пенсионного возраста с буклетами вуза. Это будет первый и последний день, когда абитуриент переступит порог этого института, — рассказывает психолог, карьерный консультант, программный директор Школы Осознанного Развития Елена Рагозина. — Современные дети не понимают ту информацию, которую им доносят на днях открытых дверей, проходящих по классической схеме. Поэтому побеждают ВУЗы, распространяющие информацию о себе заблаговременно и не в традиционных направлениях.

# Обследование объекта сервиса поиск абитуриентов на базе технологий NLP и машинного обучения

## Формулировка требований к сервису поиска абитуриентов на базе технологий NLP и машинного обучения

Функционал сервиса:

* Составлять прогноз абитуриентов
* Самообучение системы
* Обновление БД
* Обработка называний музыкальных групп
* Работа с файлом

## Обзор аналогов сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

Определения:

* **Парсер** (от англ. *parser*) или **граббер** (от англ. *grabber*) — программное обеспечение, предназначенное для анализа и разбора исходных данных, с целью их обработки и дальнейшей использования в требуемом виде.
* **Парсинг –** автоматический анализ и сбор данных по заданному признаку.
* **Целевая группа, целевая аудитория** — термин, используемый в маркетинге или рекламе для обозначения группы людей, объединенных общими признаками, или объединенной ради какой-либо цели или задачи.

Задачи:

* Анализ студентов ХМАО и разбиение их на целевые группы по направлениям обучения
* Научить машину распознавать абитуриента по профилю социальной сети ВКонтакте

Критерии для сравнения аналогов:

* #
* Цена
* Работа с АПИ ВК
* Тип ПО
* Скорость обработки
* Машинное обучение (умение анализировать)

В ходе анализа не было выявлено ни одного прямого аналога. В результате чего, был произведен поиск косвенных аналогов, занимающихся схожим функционалом.

Аналоги:

Аналоги:

* #
* Сервис поиска целевой аудитории ВКонтакте vk.barkov.net
* Vk Audio Parser v2.2
* Яндекс.Музыка
* Pepper.ninja

В ходе сравнения аналогов, была составлена сравнительная таблица:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Критерии/ПО | Цена руб/мес. | Работа с музыкой | Тип ПО | Скорость обработки | Анализ данных |
| vk.barkov.net | 390 | ✔ (Только со страницы профиля) | Web-сервис | 30000 | ➖ |
| Vk Audio Parser v2.2 | 1200 | ✔ | Настольное ПО | 150 | ➖ |
| Pepper.ninja | 690 | ➖ | Web-сервис | 2000 | ➖ |
| Яндекс.Музыка | 100 | ✔ | Web-сервис | Персонально | ✔ |

# расширить таблицу, добавить другие критерии взять из диалога аналог

Как видно из таблицы сравнения, одним из главных критериев был анализ данных, ни один аналог, кроме Яндекс.Музыки не удовлетворяет данным условиям.

Разработка данного сервиса актуальна для университетов, прямых аналогов не имеет.

## Проектирование сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

## Проектирование логической модели сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

## Идентификация сценариев сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

## Проектирование классов сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

## Разработка физической модели сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

# Основания выбора средств разработки сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

## Выбор СУБД для сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

База данных - совокупность взаимосвязанных данных, организованных в соответствии со схемой базы данных таким образом, чтобы с ними мог работать пользователь. [# ГОСТ 34.321-96]

Система управления базами данных(СУБД) - Совокупность программ и языковых средств, предназначенных для управления данными в базе данных, ведения базы данных и обеспечения взаимодействия ее с прикладными программами. [# ГОСТ 20886-85]

Реляционная база данных представляет собой множество взаимосвязанных таблиц, каждая из которых содержит информацию об объектах определенного вида. Каждая строка таблицы содержит данные об одном объекте, а столбцы таблицы содержат различные характеристики этих объектов – атрибуты [# Академия Microsoft: Технология Microsoft ADO .NET Технология Microsoft ADO .NET Ч.А. Кариев]

Выбор систему управления баз данных (СУБД) представляет сложную задачу. Один из важнейших этапов разработки баз данных. Выбранный программный продукт должен удовлетворять многим критериям: текущим потребностям, будущим потребностям, финансовым затратам на приобретение оборудования, программного обеспечения и его использования.

При разработке тестировались схожие СУБД: MySQL, Oracle, PostgreSQL, SQLite.

# +- каждой

Сервис поиска абитуриентов будет работать СУБД PostgerSQL. Поэтому в рамках выпускной квалификационной работы будет использоваться она. Графические оболочки для работы с СУБД PostreSQL: HeidiSQL и pgAdmin 4. Данный выбор заключается в том, что они хорошо дополняют друг друга. PgAdmin 4 хорош для создания самой базы данных и ее настройки. HeidiSQL удобен для резервирования БД, просмотра и редактирования данных. Так же был рассмотрена графическая оболочка dbForge Studio Express for PostgreSQL. Данный продукт находится на стадии разработки и функционал продукта очень скуден, на момент написания ВКР был возможет лишь просмотр данных. Проектирование БД проводилось в сервисе dbdesigner.net

## Выбор языка программирования сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

# Разработка сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

# Тестирование сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

# Заключение

# Список использованных источников

# Приложения

# Приложение А Техническое задание на создание сервиса поиска абитуриентов на базе технологий NLP и машинного обучения

# Приложение Б Диаграмма вариантов использования

# Приложение В Диаграмма классов

# Приложение Г Диаграмма развёртывания

# Приложение Д Диаграмма базы данных сервиса поиска абитуриентов на базе технологий NLP и машинного обучения