Министерство наук и высшего образования Российской Федерации Государственное образовательное учреждение высшего образования

ГОУ ВПО «Пермский государственный научно-исследовательский университет»

**Отчет по проекту**

**GPT агрегатор терминов**

**беспилотных авиационных систем**

По дисциплине «Групповая проектная работа»

Выполнили:

Студенты 1-го курса ИКНТ

группа «Аноним8»:

Зырянова Ника Артуровна ПМИ-7,8

Сакулин Ярослав Александрович ПМИ-7,8

Резепин Антон Львович ПМИ-7,8

Черемных Архип Олегович ПМИ-7,8

Зернин Максим Юрьевич ИТХ-3,4

Пугач Дмитрий Николаевич ИТХ-3,4

Проверил

Кандидат филологических наук  
Заведующий кафедрой английского языка профессиональной

Коммуникации ПГНИУ

**Исаева Екатерина Владимировна**

Подпис

Оглавление

[**1.** **Постановка задачи** 3](#_Toc167449130)

[**2.** **Цели и задачи проекта** 3](#_Toc167449131)

[**2.1** **цель проекта** 3](#_Toc167449132)

[**2.2** **Задачи проекта** 3](#_Toc167449133)

[**3.** **Диаграмма Ганта** 3](#_Toc167449134)

[**4.** **Этапы работы** 3](#_Toc167449135)

[**5.** **Риски** 4](#_Toc167449136)

[**6.** **Описание продукта** 5](#_Toc167449137)

[**6.1** **Функции, выполняемые программой** 5](#_Toc167449138)

[**6.2** **Технологии реализации** 5](#_Toc167449139)

[**7.** **Тестирование программы** 5](#_Toc167449140)

[**8.** **Результаты проекта** 6](#_Toc167449141)

[**9.** **Заключение** 6](#_Toc167449142)

# **Постановка задачи**

Задача проекта заключается в создании модуля, позволяющего сформировать актуальную базу данных о беспилотных авиационных системах (БАС), для автоматизации работы лингвистов по разработке терминов БАС с использованием нейронных сетей.

# **Цели и задачи проекта**

## **цель проекта**

Целью проекта «GPT-агрегатор терминов БАС» является создание модуля, позволяющего сформировать актуальную базу данных терминов БАС с использованием нейронных сетей, для определения терминов и присвоению им категорий, соответствующих в базе данных.

## **Задачи проекта**

1. Изучение теории работы с нейросетемя
2. Создание парсера для сайтов БАС
3. Изучение работы с базами данных
4. Доработка готовой базы данных
5. Создание модуля
6. Написание отчетной документации

# **Диаграмма Ганта**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1.03 | 15.03 | 31.03 | 15.04 | 30.04 | 1.05 | 15.05 | 27.05 |
| Зырянова Н. |  |  |  |  |  |  |  |  |
| Сакулин Я. |  |  |  |  |  |  |  |  |
| Черемных А. |  |  |  |  |  |  |  |  |
| Резепин А. |  |  |  |  |  |  |  |  |
| Зернин М. |  |  |  |  |  |  |  |  |
| Пугач Д. |  |  |  |  |  |  |  |  |

# **Этапы работы**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Сроки выполнения** | **Этап разработки** | **Фактические сроки выполнения** | **Комментарий** | **Ответственный** |
| 1.03 – 3.04 | Разработка парсера | 1.03 – 15.04 | Изменение состава команды | Пугач Д.  Черемных А.  Зырянова Н. |
| 1.03 – 3.04 | Разработка базы данных | 1.03 – 5.05 | Отсутствие знаний в этой области | Резепин А. |
| 1.03 – 9.03 | Изучение и выбор модели нейронной сети | 2.03 – 15.03 | Проблемы с изучением материала | Сакулин Я. |
| 9.03 – 30.03 | Разработка архитектуры нейронной сети | 15.03 – 31.03 | Недостаточно опыта работы | Сакулин Я. |
| 30.03 – 17.04 | Программирование нейронной сети | 31.04 – 19.04 | Недостаточно опыта работы, изменение состава команды | Сакулин Я.  Черемных А. |
| 17.04 – 11.05 | Обучение нейронной сети | 19.04 – 10.05 | Проблемы с обучением, изменение состава команды | Сакулин Я.  Черемных А. |
| 17.04 – 15.05 | Сборка программы | 10.05 – 15.05 | Из-за сдвигов в других этапах и изменение состава команды | Зернин М. |
| 15.05 – 17.05 | Тестирование программы | 15.05 – 19.05 | Из-за сдвигов в других этапах | Все участники команды |
| 17.05 – 19.05 | Внесение правок в программу | 19.05 – 21.05 | Из-за сдвигов в других этапах | Зернин М. |
| 19.05 – 22.05 | Написание отчета | 21.05 – 24.05 | Из-за сдвигов в других этапах | Зырянова Н. |
| 22.05 – 25.05 | Подготовка презентации | 24.05 – 26.05 | Из-за сдвигов в других этапах | Зырянова Н. |
| 27.05 | Сдача проекта заказчику | 27.05 | Без изменений | Все участники |

# **Риски**

1. Ожидаемые

|  |  |
| --- | --- |
| Риск | Решение |
| Риск невыполнения работ | Разработать более детальный план работ и следовать этому графику |
| Риск нарушения сроков выполнения | Составить примерный график работ для каждого участника в соответствии с его учебным планом |
| Недостаток знаний в сфере работы с нейросетями | Поиск и изучение необходимой теории необходимой для проекта |
| Недостаток опыта работы в сфере работы с нейросетями | Методом проб и ошибок изучать работу нейросетей |

1. Появившиеся в процессе выполнения

|  |  |
| --- | --- |
| Риск | Решение |
| Недостаток рабочей силы | Объединиться с группой, у которой такой же проект, если это возможно |
| Проблема с обучением нейронной сети | Найти человека, у которого есть опыт работы с нейронными сетями |
| Проблема в реализации базы данных | Найти человека, у которого есть опыт работы с базами данных/обратиться к старшекурсникам, с которыми совместный проект |

# **Описание продукта**

## **Функции, выполняемые программой**

Перечень функций, выполняемых программой:

* Автоматический сбор данных с сайтов по беспилотным авиационным системам: термины, контексты, характеристики
* Добавление данных с сайтов в базу данных
* Выделение терминов из контекстов с помощью нейросети, присвоение им категории, соответствующих в базе данных
* Запись новых терминов с присвоенными им категориями в базу данных
* Ручное редактирование терминов и запись в базу данных; удаление терминов
* Формирование актуального словаря с терминологией БАС

## **Технологии реализации**

* Языки программирования
  + Python

Использовался для создания базы данных, парсера (для автоматического сбора данных), бэкенда

* + BERT

Использовался для выделения терминов из контекста и присвоению им категории, машинное обучение

* + MySQL

Использовался для взаимодействия с базой данных

# **Тестирование программы**

1. Проверка работоспособности базы данных на ручное редактирование данных, добавление данных как ручное, так и автоматическое, удаление данных.
2. Проверка работоспособности нейронной сети с определением терминов из контекстов и присвоению им категорий. С какой точностью она определяет термины
3. Проверка автоматического сбора данных с сайтов про беспилотные авиационные системы. Занесение данных в базу данных без определения категории нейронной сетью
4. Общее тестирование всей программы на работоспособность и функциональность
5. Отладка модуля

# **Результаты проекта**

Результатом проекта является функциональный модуль, который формирует базу данных с терминологией БАС.

# **Заключение**

Программа была успешно реализована и в настоящий момент готова к использованию, но с некоторыми изменениями требований Заказчика. На начальном этапе было запланировано, что, помимо вышеперечисленных функций, которые выполняет программа, будут осуществлены следующие функции: лемматизация терминов (языковых единиц на русском языке, имеющих специфическое значение для предметной области БАС), сохранение гиперссылок на веб-сайт источника термина, но из-за большого плана на реализацию и маленького опыта работы с базами данных, нейронными сетями эти функции не удалось реализовать.

В дальнейшем можно решать следующие задачи:

1. Модификация модуля и добавление недостающих функций
2. Оптимизация программы
3. Выход на Российский рынок

В процессе работы над проектом участники команды научились:

1. Создавать и работать с базами данных
2. Изучили работу с нейросетями
3. Понимание работы Backend
4. Научились писать ТЗ
5. Умение работы в команде
6. Умение парсить данные с сайтов
7. Создание модуля на Python