|  |  |  |
| --- | --- | --- |
|  |  |  |
| МИНОБРНАУКИ РОССИИ | | |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** | | |
| Институт искусственного интеллекта | | |
| Кафедра программного обеспечения систем радиоэлектронной аппаратуры | | |

|  |  |
| --- | --- |
| **ОТЧЕТ** | |
| **ПО НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ** | |
|  | |
| **Тема: «Парсер расписания в JSON/БД»** | |
| Студент группы           КМБО-05-22 | *И.В Пахалков* |
| Руководитель практики | *В.И. Смирнов* |

|  |  |  |
| --- | --- | --- |
| «Отчет представлен к рассмотрению» | «\_\_\_» \_\_\_\_\_\_\_\_\_\_2024 г. | *(подпись студента)* |
|  |  |  |
| «Отчет утвержден.  Допущен к защите» | «\_\_\_» \_\_\_\_\_\_\_\_\_\_2024 г. | *(подпись руководителя)* |

Москва 2024

**ОГЛАВЛЕНИЕ**

[Введение 3](#_Toc168864222)

[1 Постановка задачи 3](#_Toc168864223)

[1.1 Задание 3](#_Toc168864224)

[1.2 Планируемые результаты 3](#_Toc168864225)

[1.3 Исходные данные 4](#_Toc168864226)

[2 Описание работ и результатов 4](#_Toc168864227)

[2.1 Используемые средства разработки 4](#_Toc168864228)

[2.2 Содержание и результат выполненных работ 4](#_Toc168864229)

[2.3 Выполненные задачи 7](#_Toc168864230)

[2.4 Доступ к результатам 10](#_Toc168864231)

[3 Заключение 11](#_Toc168864232)

[Список литературы 12](#_Toc168864233)

# Введение

В современном мире информационных технологий автоматизация обработки данных является ключевым аспектом стабильной работы различных систем и приложений. В данной научно-исследовательской работе рассматривается разработка парсера расписания и импортирование его в файлы формата JSON и базу данных (БД) PostgreSQL, что позволяет эффективно обрабатывать и хранить данные расписания учебных групп.

Целью данной работы является изучение и реализация методов парсинга расписания учебных занятий для групп кафедры 536 (КМБО-..., KMMO-...), а также импорт этих данных в БД для последующего использования.

# 1 Постановка задачи

1.1 Задание

Необходимо разработать и реализовать скрипт на языке python, который парсит расписание из EXCEL файла в формат JSON для последующего импортирования в базу данных. Также должен быть реализован скрипт на языке python, который парсит полученные значения из JSON файла и импортирует их в таблицы sc\_rasp, sc\_prep, sc\_group и sc\_desc.

1.2 Планируемые результаты

* загрузка в базу данных сведений о расписании занятий групп кафедры ВЕГА;
* отчет с описанием возможностей улучшения.

1.3 Исходные данные

* EXCEL файлы, хранящие расписание всех групп института искусственного интеллекта (рисунок 1);

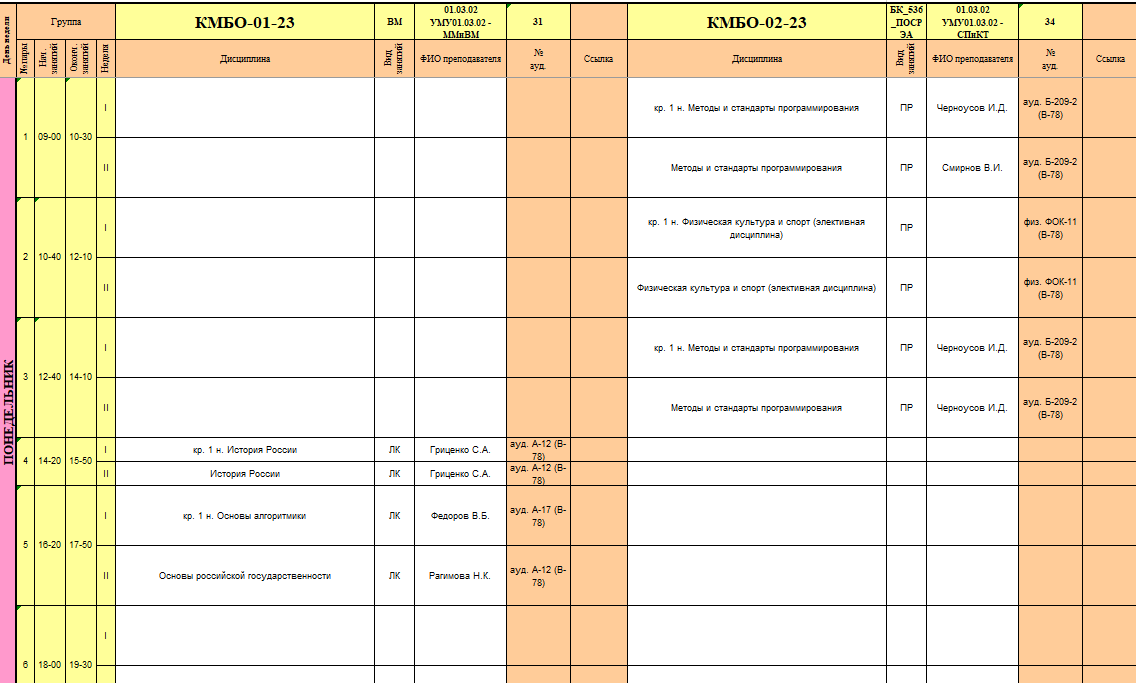


Рисунок 1 – структура EXCEL файла

- заготовленная база данных с определённым форматом.

# 2 Описание работ и результатов

2.1 Используемые средства разработки

Для написания исходного кода были использованы библиотеки openpyxl, json, и psycopg2.

* 1. Содержание и результат выполненных работ

В результате импорта данных из JSON в БД получаем таблицы:

* таблица sc\_rasp хранит в себе всю нужную информацию про расписание (рисунок 2);
* таблица sc\_prep хранит информацию о преподавателях (рисунок 3);
* таблица sc\_desc хранит информацию о дисциплинах занятий (рисунок 4);
* таблица sc\_group хранит информацию о группах (рисунок 5).

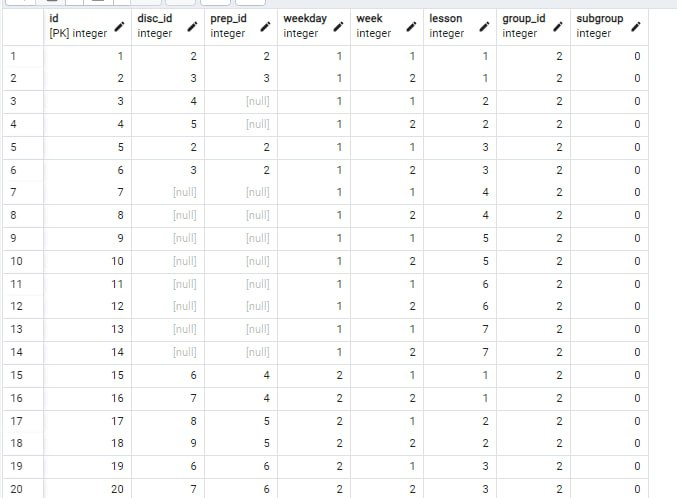


Рисунок 2 – таблица sc\_rasp



Рисунок 3 – таблица sc\_prep

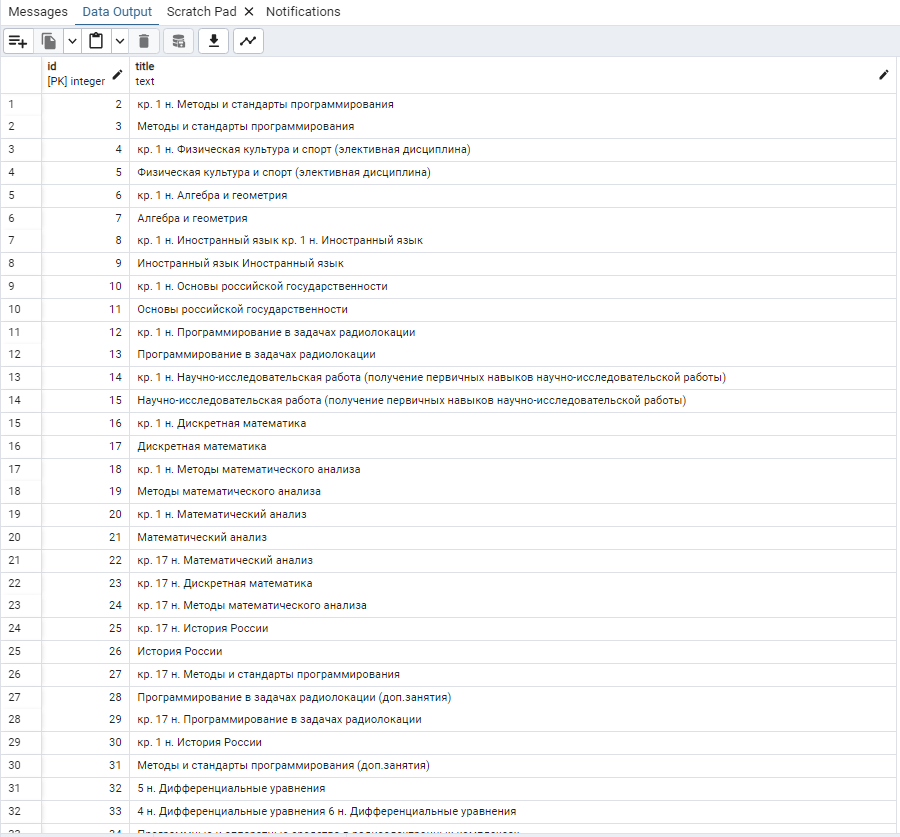


Рисунок 4 – таблица sc\_disc

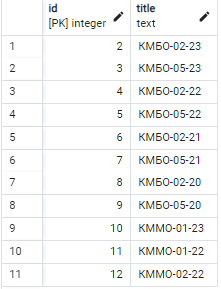


Рисунок 5 – таблица sc\_group

2.3 Выполненные задачи

В соответствии с планом выполнения индивидуального задания, были выполнены следующие задачи:

* изучение библиотек для работы с файлами EXCEL на языке Python. Анализ формата хранения данных в таблице расписания и выделение основных полей, необходимых для парсинга;
* разработана функция get\_main для получения основной информации расписания, а именно: день недели, номер пары, нач. занятия, оконч. занятия, неделя занятия;
* разработана функция get\_group, которая получает группы только 536 кафедры;
* разработана функция для парсинга данных с EXCEL. Концепция реализации данной функции представлена на рисунке 6;

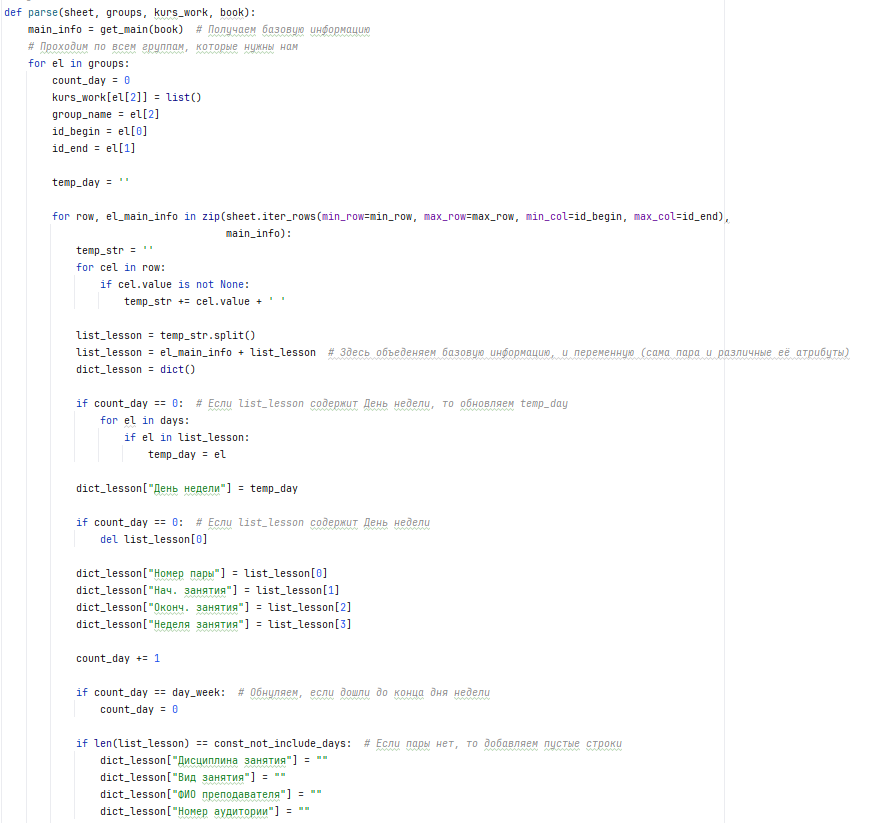


Рисунок 6 – функция parse

* разработана функция, которая преобразует полученные данные в формат JSON, используя функции описанные выше (рисунок 7). На рисунке 8 представлен их вид;

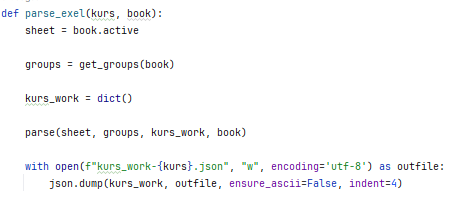


Рисунок 7 – функция parse\_exel

  
Рисунок 8 – данные в формате JSON

* разработана функция get\_weekday\_number, которая переводит дни недели в нужный формат для БД;
* разработана функция get\_week\_number, которая переводит четный формат в нужный формат для БД;
* импорт распаршенных данных в итоговую базу данных. На рисунке 9 показана концепция этой программы.

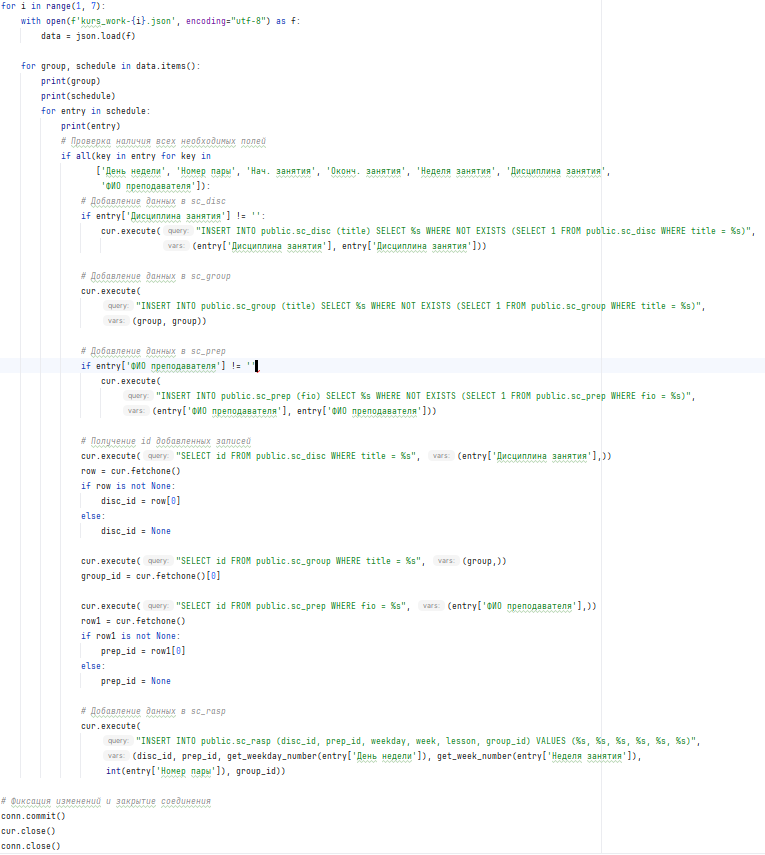


Рисунок 9 – импорт данных в БД

2.4 Доступ к результатам

Результаты практики расположены в репозитории Gitlab и отправлены руководителю в виде [ссылки](https://vgit.mirea.ru/kub78/314-practica).

# 3 Заключение

В ходе выполнения данной работы была создана система для парсинга и обработки расписания учебных групп кафедры 536 (КМБО-..., KMMO-...). Проведено детальное изучение библиотек для работы с файлами EXCEL на языке Python, что позволило эффективно извлекать и обрабатывать необходимые данные. Были выделены основные поля таблицы расписания, осуществлен парсинг данных по указанным группам и запись в формат JSON.

Дополнительно была изучена работа с базой данных PostgreSQL на языке Python, что позволило импортировать распашенные поля в базу данных. Все этапы работы были выполнены в соответствии с поставленными задачами, с учетом замечаний руководителя проекта.

Цель работы достигнута.

# Список литературы

* [Онлайн-справочник по языку Python](https://docs.python.org/3/reference/index.html)
* [Документация JSON](https://docs.python.org/3/library/json.html)
* [Документация openpyxl](https://docs-python.ru/packages/modul-openpyxl/)