Университет ИТМО

Факультет ФПИ и КТ

Лабораторная работа №3

"Исследование языков разметки документов"

Вариант 28

Выполнил

Стасик Илья Сергеевич

Группа P3111

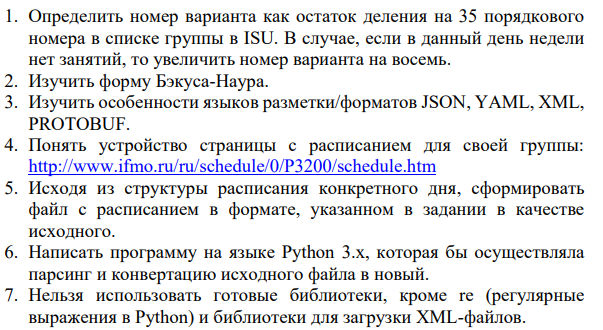
Преподаватель

П.В. Балакшин

г. Санкт-Петербург 2020

Цель: изучить особенности языков разметки различных форматов. Научиться писать парсер, осуществляющий конвертацию из формата JSON в XML

Задание



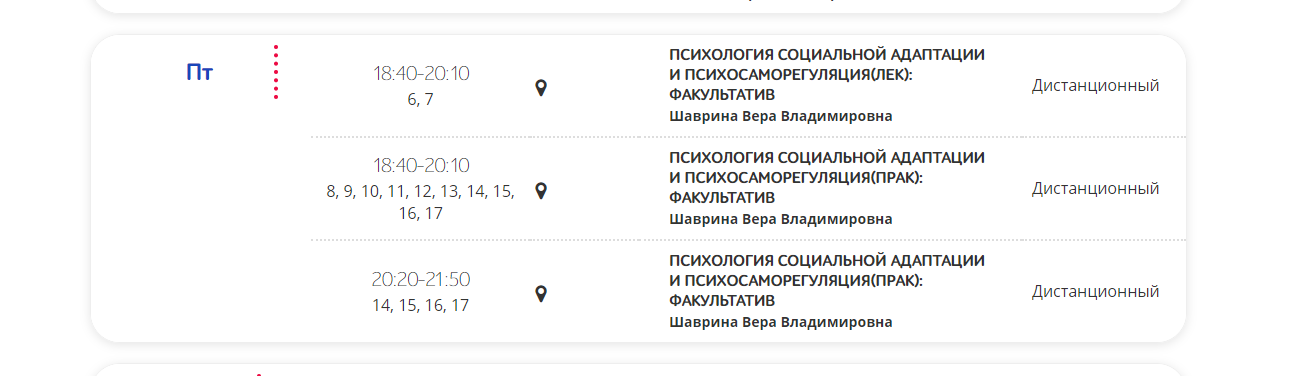
Отчёт онлайн:

<https://drive.google.com/file/d/1IJm7ifFrV8MpFGEoOTMniq9tgFz4Xqh5/view?usp=sharing>

GitHub:

<https://github.com/BeatenPixel/ITMO_programming/tree/master/year0/part0/inf/lab3>

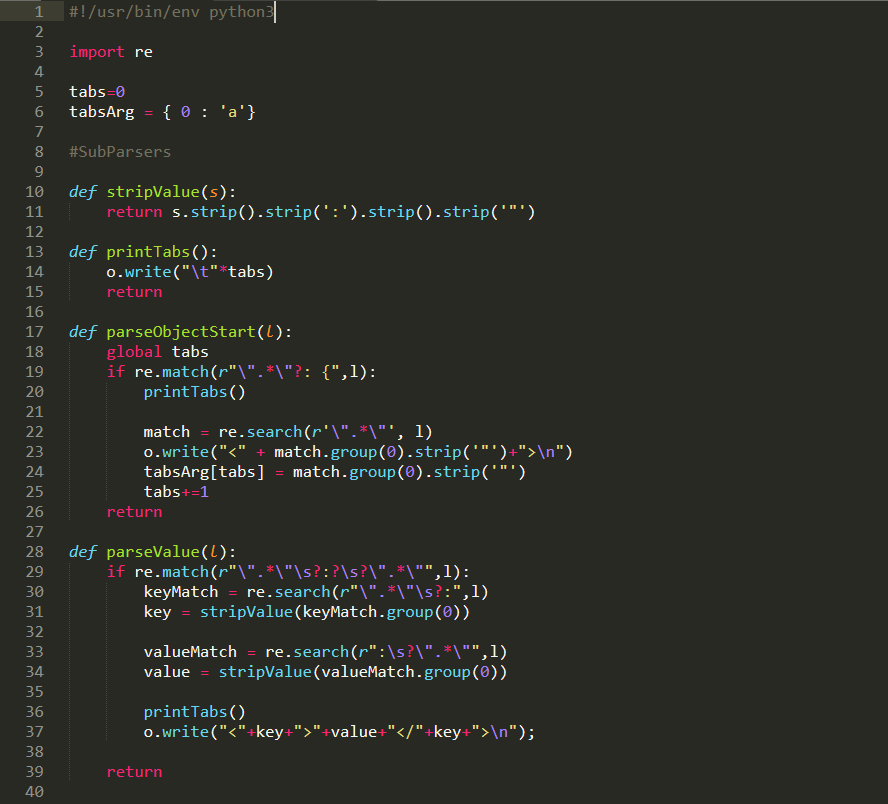
Выполнение

На основе расписания составил JSON файл (schedule.json):



Написал парсер на языке Python (parser.py)

При выполнении использовалась библиотека “re”



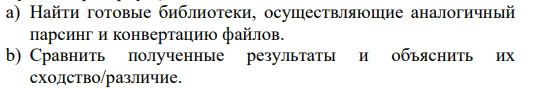




Получившийся файл schedule\_generated.xml соответствует формату XML:



Дополнительное задание 1:



Нашёл библиотеку json2xml

Для использования библиотеки её нужно установить:

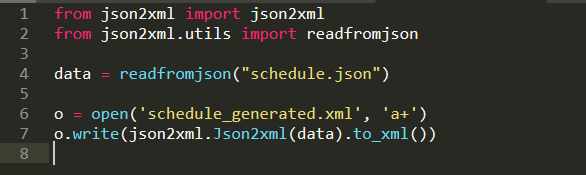
*pip install json2xml*

И импортировать в скрипт:

*from json2xml import json2xml*

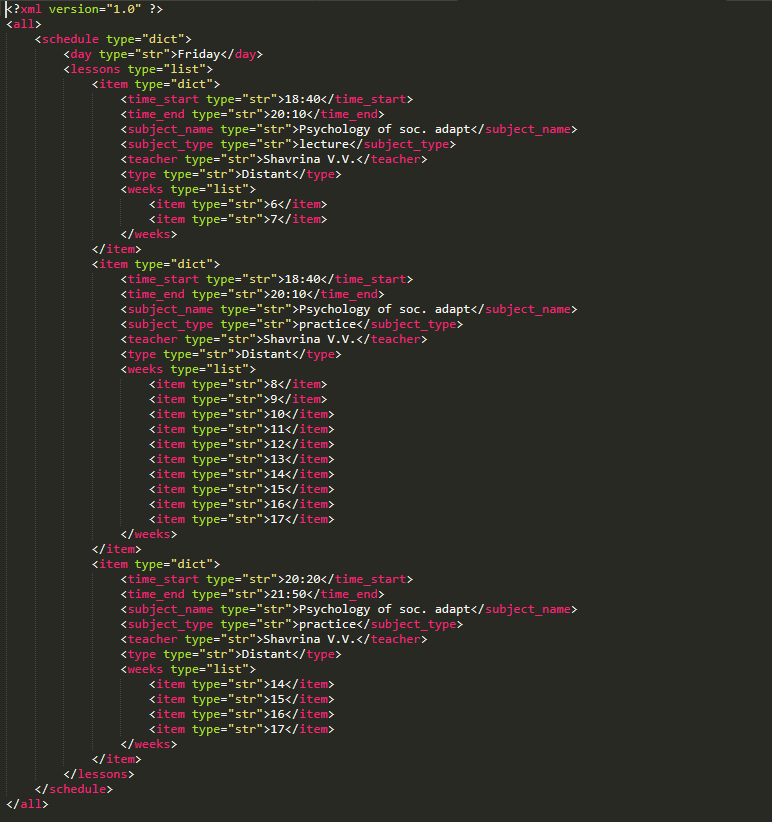
*from json2xml.utils import readfromjson*

Парсинг с помощью библиотеки json2xml:

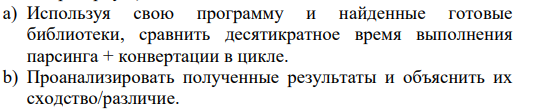


Использование готовой библиотеки упрощает программу в разы

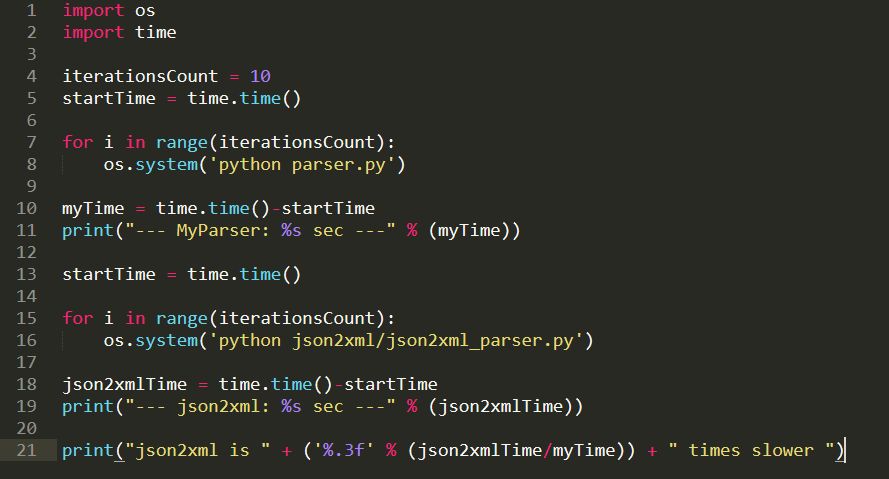
Файл сгенерированный библиотекой сложнее по структуре, в нём присутствуют дополнительные атрибуты (например, атрибут type обозначающий тип)



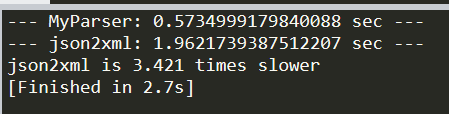
Дополнительное задание 2:



Написал дополнительный скрипт, который сравнивает время выполнения (10 итераций) моего парсера и парсера, использующего json2xml:



Вывод программы:



Можем сделать вывод, что библиотека json2xml приблизительно в 3 раза медленнее собственно написанного парсера. Это связано с тем, что библиотека рассчитана на более универсальный, усложнённый парсинг, нужно учитывать все возможные варианты, выводить более детальную информацию, что и замедляет работу программы.

Вывод:

Изучил особенности языков разметки различных форматов. Написал парсер, осуществляющий конвертацию из формата JSON в XML, нашёл стороннюю библиотеку, осуществляющую схожую конвертацию. Сравнил скорость работы своего парсера с готовой библиотекой json2xml