Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Факультет программной инженерии и компьютерной техники

**Лабораторная работа #3**

**"Исследование языков разметки для документов"**

Вариант 8

Выполнил:

Студент 1-го курса, группы P3114

Ермаков Т. С., 311733

Преподаватели:

Балакшин Павел Валерьевич

Ильина Аглая Геннадьевна

Санкт-Петербург, 2020

**Порядок** **выполнения** **работы**

1. Определить номер варианта как остаток деления на 35 порядкового номера в списке группы в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь. **XML -> YAML**

2. Изучить форму Бэкуса-Наура.

3. Изучить особенности языков разметки/форматов JSON, YAML, XML, PROTOBUF.

4. Понять устройство страницы с расписанием для своей группы: <http://www.ifmo.ru/ru/schedule/0/P3200/schedule.htm>

5. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного.

6. Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.

7. Нельзя использовать готовые библиотеки, кроме re (регулярные выражения в Python) и библиотеки для загрузки XML-файлов.

8. Необязательное задание для получения оценки «4» и «5» (позволяет набрать от 75 до 89 процентов от максимального числа баллов БаРС за данную лабораторную).

a) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.

b) Сравнить полученные результаты и объяснить их сходство/различие.

9. Необязательное задание для получения оценки«5» (позволяет набрать от 90 до 100 процентов от максимального числа баллов БаРС за данную лабораторную).

a) Используя свою программу и найденные готовые библиотеки, сравнить десятикратное время выполнения парсинга + конвертации в цикле.

b) Проанализировать полученные результаты и объяснить их сходство/различие.

10.Проверить, что все пункты задания выполнены и выполнены верно. 11.Написать отчёт о проделанной работе.

12.Подготовиться к устным вопросам на защите.

**Выполнение работы.**

**Исходный код XML**

<Timetable>

    <Day>Вторник

        <Lesson>

            <Time>10:00-11:30</Time>

            <Week></Week>

            <Subject>МАТЕМАТИКА (БАЗОВЫЙ УРОВЕНЬ)(ПРАК)</Subject>

            <Teacher>Назаров Равшанжон</Teacher>

            <Build>Кронверкский пр., д.49, лит.А</Build>

            <Room>427</Room>

        </Lesson>

        <Lesson>

            <Time>13:30-15:00</Time>

            <Week>четная неделя</Week>

            <Subject>ФИЗИКА(ЛАБ)</Subject>

            <Teacher>Коробков Максим Петрович</Teacher>

            <Build>Биржевая линия, д.16, лит.А</Build>

            <Room>545</Room>

        </Lesson>

        <Lesson>

            <Time>13:30-15:00</Time>

            <Week>нечетная неделя</Week>

            <Subject>ФИЗИКА(ЛАБ)</Subject>

            <Teacher>Коробков Максим Петрович</Teacher>

            <Build>Биржевая линия, д.16, лит.А</Build>

            <Room>545</Room>

        </Lesson>

        <Lesson>

            <Time>15:20-16:50</Time>

            <Week></Week>

            <Subject>ФИЗИКА(ЛЕК)</Subject>

            <Teacher>Смирнов Александр Витальевич</Teacher>

            <Build>Биржевая линия, д.16, лит.А</Build>

            <Room>550</Room>

        </Lesson>

    </Day>

</Timetable>

**Исходный код parser.py**

import time  
start\_time = time.time()  
input = open("timetable.xml", "r", encoding="UTF-8")  
output = open("timetable.yml", "w", encoding="UTF-8")  
f = input.read().split("\n")  
k = 0  
for line in f:  
 a = line.replace("<", ">").split(">")  
 if line.count("/") == 1:  
 if line.count("<") > 1:  
 output.write(" " \* k + a[1] + ': "' + a[2] + '"\n')  
 else:  
 k -= 1  
 else:  
 output.write(" " \* k + a[1] + ":" + "\n")  
 k += 1  
input.close()  
output.close()  
print((time.time()-start\_time))

**Полученный результат в YAML**

Timetable:

 Day:

  Lesson:

   Time: "10:00-11:30"

   Week: ""

   Subject: "МАТЕМАТИКА (БАЗОВЫЙ УРОВЕНЬ)(ПРАК)"

   Teacher: "Назаров Равшанжон"

   Build: "Кронверкский пр., д.49, лит.А"

   Room: "427"

  Lesson:

   Time: "13:30-15:00"

   Week: "четная неделя"

   Subject: "ФИЗИКА(ЛАБ)"

   Teacher: "Коробков Максим Петрович"

   Build: "Биржевая линия, д.16, лит.А"

   Room: "545"

  Lesson:

   Time: "13:30-15:00"

   Week: "нечетная неделя"

   Subject: "ФИЗИКА(ЛАБ)"

   Teacher: "Коробков Максим Петрович"

   Build: "Биржевая линия, д.16, лит.А"

   Room: "545"

  Lesson:

   Time: "15:20-16:50"

   Week: ""

   Subject: "ФИЗИКА(ЛЕК)"

   Teacher: "Смирнов Александр Витальевич"

   Build: "Биржевая линия, д.16, лит.А"

   Room: "550"

**Задания 8 и 9:**

Перевод из XML в YAML и наоборот можно сделать с помощью библиотеки xmlplain. Результат почти не отличается от моего. При помощи xmlplain пустые строки у room и location, а также длинное название пар были обособленны в кавычки. Время работы xmlplain в 6 раз дольше, чем у моего парсера (0.011 против 0.002) Так получается из-за того, что xmlplain подключает библиотеки xml, yaml и др., а также оно более приспособлено к различным вариантам входящих данных (например, списки).

**Выводы**

В процессе выполнения лабораторной работы я изучил синтаксис Python (навыки работы с файлами, обработки строк с помощью Python, условия, циклы), языки разметки XML, YAML и JSON, работу с внешними библиотеками.