

Санкт-Петербургский Национальный Исследовательский Университет
Информационных Технологий, Механики и Оптики Факультет Программной Инженерии и
Компьютерной Техники



Вариант № 6
Лабораторная работа № 3
по дисциплине
'Информатика'

Выполнил:
Веласкес Хуан
291005;
Р3113;
Преподаватель:
Балакшин Павел Валерьевич

Санкт-Петербург 2020 г.

Текст задания:

Лабораторная работа No3. «Исследование языков разметки документов»

Порядок выполнения работы

- 1) Изучить форму Бэкуса-Наура.
- 2) Изучить особенности языков разметки/форматов JSON, YAML, XML, PROTOBUF.
- 3) Понять устройство страницы с расписанием для своей группы:
<http://www.ifmo.ru/ru/schedule/0/P3200/schedule.htm>
- 4) Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного.
- 5) Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.
- 6) Нельзя использовать готовые библиотеки, кроме re (регулярные выражения в Python) и библиотеки для загрузки XML-файлов.
- 7) 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
- 8) 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.
- 9) Проверить, что все пункты задания выполнены и выполнены верно.
- 10) Написать отчёт о проделанной работе.
- 11) Подготовиться к устным вопросам на защите.

Программа без библиотеки:

```
import time
from timeit import timeit

def split(content):

    fileR_array= []
    for line in content:
        stripped_line = line.strip()
        fileR_array.append(stripped_line)
    return logic(fileR_array)

def logic(array):
    aux=[]
    level=-1
    for line in array:
        if(line=="{"):
            level+=1
        elif(line=="},"):
            level-=1
        elif(line=="}"):
            break
    else:
```

```

        aux.append([level, line.replace(",","")])
    return aux

def addSpace(stringElement,level):
    string_length=len(stringElement)+level    # will be adding level extra spaces
    string_revised=stringElement.rjust(string_length)
    return string_revised

def construct(arrayWithLevels):
    outArrAux=[]
    for elem in arrayWithLevels:
        numLevel = elem[0]
        strElem = elem[1]
        outArrAux.append(addSpace(strElem,numLevel))
    return outArrAux

def parsing():
    with open('timetable.json', 'r') as rf:
        with open('timetableY.yaml', 'w') as wf:
            rf_content=rf.readlines()
            string_without_line_breaks = ""
            arrayBlock=[]
            outputArray=[]

            arrayBlock=split(rf_content)
            outputArray=construct(arrayBlock)
            for element in outputArray:
                element=element+'\n'
                wf.write(element)

start_time = time.time()
parsing()
print("%.6f" % (time.time() - start_time))

```

Программа с библиотекой:

```

import sys
import json
import yaml
import time

def parsingLibrary():

    outputFile = open("timetableLibrery.yaml", "w", encoding="utf-8")

    with open('timetable.json', encoding='utf-8') as fh:
        data = json.load(fh)
        outputFile.write(yaml.dump(data, allow_unicode=True))

start_time = time.time()
parsingLibrary()
print("Library %s seconds ---" % (time.time() - start_time))

```

input.JSON

```

{
  "Расписание":
  {
    "Вторник":
    {
      "Пара1":
      {
        "Время": "11:40-13:10",
        "Неделя": "всегда",
        "Ауд": "427 АУД.",

```

```

      "Корпус": "Кронверкский пр., д.49, лит.А",
      "Предмет": "МАТЕМАТИКА (БАЗОВЫЙ УРОВЕНЬ) (ПРАК) ",
      "Преподаватель": ""
    },
    "Пара2":
    {
      "Время": "15:20-16:50",
      "Неделя": "всегда",
      "Ауд": "550 АУД.",
      "Корпус": "Биржевая линия, д.16, лит.А",
      "Предмет": "ФИЗИКА (ЛЕК) ",
      "Преподаватель": "Смирнов Александр Витальевич"
    },
    "Пара3 (чёт) ":
    {
      "Время": "17:00-18:30",
      "Неделя": "четная неделя",
      "Ауд": "535 АУД.",
      "Корпус": "Биржевая линия, д.16, лит.А",
      "Предмет": "ФИЗИКА (ЛАБ) ",
      "Преподаватель": "Смирнов Александр Витальевич"
    },
    "Пара4 (нечёт) ":
    {
      "Время": "17:00-18:30",
      "Неделя": "нечетная неделя",
      "Ауд": "535 АУД.",
      "Корпус": "Биржевая линия, д.16, лит.А",
      "Предмет": "ФИЗИКА (ЛАБ) ",
      "Преподаватель": "Смирнов Александр Витальевич"
    }
  }
}

```

Результат работы программы без библиотеки файл YAML :

"Расписание":

"Вторник":

"Пара1":

"Время": "11:40-13:10"

"Неделя": "всегда"

"Ауд": "427 АУД."

"Корпус": "Кронверкский пр. д.49 лит.А"

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

"Преподаватель": ""

"Пара2":

"Время": "15:20-16:50"

"Неделя": "всегда"

"Ауд": "550 АУД."

"Корпус": "Биржевая линия д.16 лит.А"

"Предмет": "ФИЗИКА(ЛЕК)"

"Преподаватель": "Смирнов Александр Витальевич"

"Пара3(чёт)":

"Время": "17:00-18:30"

"Неделя": "четная неделя"

"Ауд": "535 АУД."

"Корпус": "Биржевая линия д.16 лит.А"

"Предмет": "ФИЗИКА(ЛАБ)"

"Преподаватель": "Смирнов Александр Витальевич"

"Пара4(нечёт)":

"Время": "17:00-18:30"

"Неделя": "нечетная неделя"

"Ауд": "535 АУД."

"Корпус": "Биржевая линия д.16 лит.А"

"Предмет": "ФИЗИКА(ЛАБ)"

"Преподаватель": "Смирнов Александр Витальевич"

Результат работы программы с библиотекой, файл YAML :

Расписание:

Вторник:

Пара1:

Ауд: 427 АУД.

Время: 11:40-13:10

Корпус: Кронверкский пр., д.49, лит.А

Неделя: всегда

Предмет: МАТЕМАТИКА(БАЗОВЫЙ УРОВЕНЬ)(ПРАК)

Преподаватель: "

Пара2:

Ауд: 550 АУД.

Время: 15:20-16:50

Корпус: Биржевая линия, д.16, лит.А

Неделя: всегда

Предмет: ФИЗИКА(ЛЕК)

Преподаватель: Смирнов Александр Витальевич

Пара3(чёт):

Ауд: 535 АУД.

Время: 17:00-18:30

Корпус: Биржевая линия, д.16, лит.А

Неделя: четная неделя

Предмет: ФИЗИКА(ЛАБ)

Преподаватель: Смирнов Александр Витальевич

Пара4(нечёт):

Ауд: 535 АУД.

Время: 17:00-18:30

Корпус: Биржевая линия, д.16, лит.А

Неделя: нечетная неделя

Предмет: ФИЗИКА(ЛАБ)

Преподаватель: Смирнов Александр Витальевич

Весь код: <https://github.com/Odzen/Lab3-Files-python/tree/master>

Results of testing

MyProgram [s]	Library [s]
0.001921	0.006844
0.001216	0.007237
0.001027	0.004395
0.001359	0.004439
0.001049	0.004450
0.001440	0.004032
0.001024	0.005081

0.000962	0.004212
0.001676	0.004760
0.001549	0.0049583

Выводы:

Благодаря этой работе я смог изучить некоторые особенности форматов JSON и YAML, а также научиться обрабатывать файлы и работать с их данными в python.

Моя программа работает примерно в 5 раз быстрее, чем библиотека, но потому что я проверяю гораздо меньше вещей для разбора. Моя программа является самой простой, и она может закончиться с ошибками или не анализировать весь возможный синтаксис, присутствующий во входных данных.