Санкт-Петербургский Национальный Исследовательский Университет

Информационных Технологий, Механики и Оптики

МФКТиУ, кафедра ПИиКТ

Лабораторная работа №4 по "Информатике"

Исследование языков разметки документов.

Выполнил: Цыпандин Николай Петрович

Группа: Р3110

Вариант: 7

Санкт-Петербург

2020 г.

**Цель.**

Овладеть знаниями о различных современных языках разметки документов и форматах данных, навыками обработки данных с помощью языка Python 3.x.

**Задание.**

* Исходя из структуры расписания конкретного дня сформировать файл с расписанием в формате, указанном в задании в качестве исходного.
* Написать программу на языке Python, которая бы осуществляла парсинг и конвертацию исходного файла в новый

|  |  |  |  |
| --- | --- | --- | --- |
| ***№ варианта*** | ***Исходный формат*** | ***Результирующий формат*** | ***День недели*** |
| ***7*** | ***yml*** | ***json*** | ***Вторник*** |
|  |  |  |  |

**Вторник.yml**

Schedule:

Tuesday:

Class1:

Time: "10:00-11:30"

Week: "нечетная неделя"

Room: "550 АУД."

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

Lesson: "Физика(ЛЕК)"

Teacher: "Коробков Максим Петрович"

Classformat: "Очно - дистанционный"

Class2:

Time: "11:40-13:10"

Week: "четная неделя"

Room: "545 АУД."

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

Lesson: "Физика(ЛАБ)"

Teacher: "Коробков Максим Петрович"

Classformat: "Очно - дистанционный"

Class3:

Time: "11:40-13:10"

Week: "нечетная неделя"

Room: "545 АУД."

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

Lesson: "Физика(ЛАБ)"

Teacher: "Коробков Максим Петрович"

Classformat: "Очно - дистанционный"

**Результат**

**Вторник.json**

{

"Schedule":

{

"Tuesday":

{

"Class1":

{

"Time":"10:00-11:30",

"Week":"нечетная неделя",

"Room":"550 АУД.",

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

"Lesson":"Физика(ЛЕК)",

"Teacher":"Коробков Максим Петрович",

"Classformat":"Очно - дистанционный"

},

"Class2":

{

"Time":"11:40-13:10",

"Week":"четная неделя",

"Room":"545 АУД.",

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

"Lesson":"Физика(ЛАБ)",

"Teacher":"Коробков Максим Петрович",

"Classformat":"Очно - дистанционный"

},

"Class3":

{

"Time":"11:40-13:10",

"Week":"нечетная неделя",

"Room":"545 АУД.",

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

"Lesson":"Физика(ЛАБ)",

"Teacher":"Коробков Максим Петрович",

"Classformat":"Очно - дистанционный"

}

}

}

}

**Парсер**

**Парсер.py**

1 # -\*- coding: utf-8 -\*-

2 inputf = open('Вторник.yml', 'r', encoding = "utf-8")

3 outputf = open('Вторник.json','w', encoding = "utf-8")

4 newline = inputf.readline()

5 data = list()

6 dopstring = list()

7 linecnt = 0

8 while newline:

9 data.append(newline)

10 linecnt +=1

11 newline = inputf.readline()

12 inputf.close()

13 start\_k = len(data[0]) - len(data[0].lstrip())

14 outputf.write("{\n")

15 gap=0

16 count=0

17 for i in range(0, linecnt-1):

18 if(i!=(linecnt-1)):

19 dopstring = data[i].lstrip().split(':', maxsplit = 1)

20 outputf.write((gap)\*" "+' "' + dopstring[0].lstrip() + '":')

21 if len(dopstring[1][:-1])!=0 :

22 count+=1

23 if(count==7) :

24 outputf.write(dopstring[1][:-1].lstrip()+"\n")

25 count=0

26 else :

27 outputf.write(dopstring[1][:-1].lstrip()+','+"\n")

28 else : outputf.write(dopstring[1][:-1].lstrip()+"\n")

29 end\_k = len(data[i+1]) - len(data[i+1].lstrip())

30 if end\_k < start\_k:

31 outputf.write((gap-1)\*" "+" },"'\n')

32 gap=gap-1

33 if end\_k > start\_k:

34 outputf.write('\n')

35 outputf.write((gap)\*" "+" {"'\n')

36 gap=gap+1

37 start\_k = end\_k

38 dopstring = data[linecnt-1].lstrip().split(':', maxsplit = 1)

39 outputf.write(" "+'"'+dopstring[0].lstrip() + '":'+dopstring[1][:-1].lstrip()+'"'+"\n")

40 for i in range(0,gap+1):

41 outputf.write((+gap-i)\*" "+'}'+'\n')

42 outputf.close()

**Задание 8**

**sravn.py**

#pip pyyaml

import json

import yaml

with open('sravn.yml') as f:

dataMap = yaml.safe\_load(f)

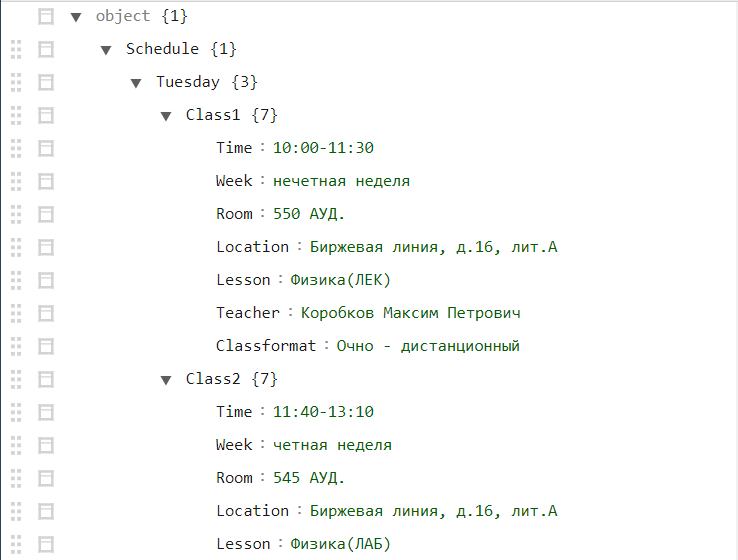
print (json.dumps(dataMap))

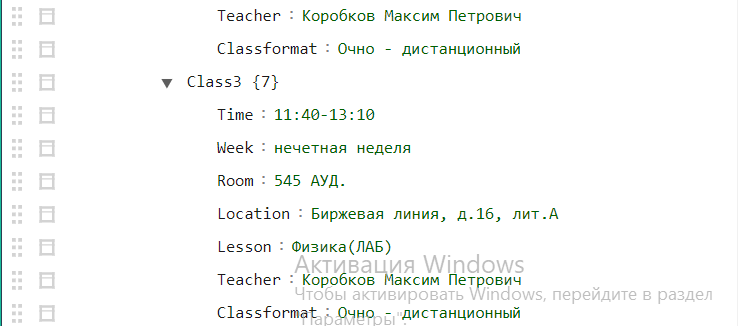
**Результат**

{"Schedule": {"Tuesday": {"Class1": {"Time": "10:00-11:30", "Week": "\u0420\u0405\u0420\u00b5\u0421\u2021\u0420\u00b5\u0421\u201a\u0420\u0405\u0420\u00b0\u0421\u040f \u0420\u0405\u0420\u00b5\u0420\u0491\u0420\u00b5\u0420\u00bb\u0421\u040f", "Room": "550 \u0420\u0452\u0420\u0408\u0420\u201d.", "Location": "\u0420\u2018\u0420\u0451\u0421\u0402\u0420\u00b6\u0420\u00b5\u0420\u0406\u0420\u00b0\u0421\u040f \u0420\u00bb\u0420\u0451\u0420\u0405\u0420\u0451\u0421\u040f, \u0420\u0491.16, \u0420\u00bb\u0420\u0451\u0421\u201a.\u0420\u0452", "Lesson": "\u0420\u00a4\u0420\u0451\u0420\u00b7\u0420\u0451\u0420\u0454\u0420\u00b0(\u0420\u203a\u0420\u2022\u0420\u0459)", "Teacher": "\u0420\u0459\u0420\u0455\u0421\u0402\u0420\u0455\u0420\u00b1\u0420\u0454\u0420\u0455\u0420\u0406 \u0420\u045a\u0420\u00b0\u0420\u0454\u0421\u0403\u0420\u0451\u0420\u0458 \u0420\u045f\u0420\u00b5\u0421\u201a\u0421\u0402\u0420\u0455\u0420\u0406\u0420\u0451\u0421\u2021", "Classformat": "\u0420\u045b\u0421\u2021\u0420\u0405\u0420\u0455 - \u0420\u0491\u0420\u0451\u0421\u0403\u0421\u201a\u0420\u00b0\u0420\u0405\u0421\u2020\u0420\u0451\u0420\u0455\u0420\u0405\u0420\u0405\u0421\u2039\u0420\u2116"}, "Class2": {"Time": "11:40-13:10", "Week": "\u0421\u2021\u0420\u00b5\u0421\u201a\u0420\u0405\u0420\u00b0\u0421\u040f \u0420\u0405\u0420\u00b5\u0420\u0491\u0420\u00b5\u0420\u00bb\u0421\u040f", "Room": "545 \u0420\u0452\u0420\u0408\u0420\u201d.", "Location": "\u0420\u2018\u0420\u0451\u0421\u0402\u0420\u00b6\u0420\u00b5\u0420\u0406\u0420\u00b0\u0421\u040f \u0420\u00bb\u0420\u0451\u0420\u0405\u0420\u0451\u0421\u040f, \u0420\u0491.16, \u0420\u00bb\u0420\u0451\u0421\u201a.\u0420\u0452", "Lesson": "\u0420\u00a4\u0420\u0451\u0420\u00b7\u0420\u0451\u0420\u0454\u0420\u00b0(\u0420\u203a\u0420\u0452\u0420\u2018)", "Teacher": "\u0420\u0459\u0420\u0455\u0421\u0402\u0420\u0455\u0420\u00b1\u0420\u0454\u0420\u0455\u0420\u0406 \u0420\u045a\u0420\u00b0\u0420\u0454\u0421\u0403\u0420\u0451\u0420\u0458 \u0420\u045f\u0420\u00b5\u0421\u201a\u0421\u0402\u0420\u0455\u0420\u0406\u0420\u0451\u0421\u2021", "Classformat": "\u0420\u045b\u0421\u2021\u0420\u0405\u0420\u0455 - \u0420\u0491\u0420\u0451\u0421\u0403\u0421\u201a\u0420\u00b0\u0420\u0405\u0421\u2020\u0420\u0451\u0420\u0455\u0420\u0405\u0420\u0405\u0421\u2039\u0420\u2116"}, "Class3": {"Time": "11:40-13:10", "Week": "\u0420\u0405\u0420\u00b5\u0421\u2021\u0420\u00b5\u0421\u201a\u0420\u0405\u0420\u00b0\u0421\u040f \u0420\u0405\u0420\u00b5\u0420\u0491\u0420\u00b5\u0420\u00bb\u0421\u040f", "Room": "545 \u0420\u0452\u0420\u0408\u0420\u201d.", "Location": "\u0420\u2018\u0420\u0451\u0421\u0402\u0420\u00b6\u0420\u00b5\u0420\u0406\u0420\u00b0\u0421\u040f \u0420\u00bb\u0420\u0451\u0420\u0405\u0420\u0451\u0421\u040f, \u0420\u0491.16, \u0420\u00bb\u0420\u0451\u0421\u201a.\u0420\u0452", "Lesson": "\u0420\u00a4\u0420\u0451\u0420\u00b7\u0420\u0451\u0420\u0454\u0420\u00b0(\u0420\u203a\u0420\u0452\u0420\u2018)", "Teacher": "\u0420\u0459\u0420\u0455\u0421\u0402\u0420\u0455\u0420\u00b1\u0420\u0454\u0420\u0455\u0420\u0406 \u0420\u045a\u0420\u00b0\u0420\u0454\u0421\u0403\u0420\u0451\u0420\u0458 \u0420\u045f\u0420\u00b5\u0421\u201a\u0421\u0402\u0420\u0455\u0420\u0406\u0420\u0451\u0421\u2021", "Classformat": "\u0420\u045b\u0421\u2021\u0420\u0405\u0420\u0455 - \u0420\u0491\u0420\u0451\u0421\u0403\u0421\u201a\u0420\u00b0\u0420\u0405\u0421\u2020\u0420\u0451\u0420\u0455\u0420\u0405\u0420\u0405\u0421\u2039\u0420\u2116"}}}}

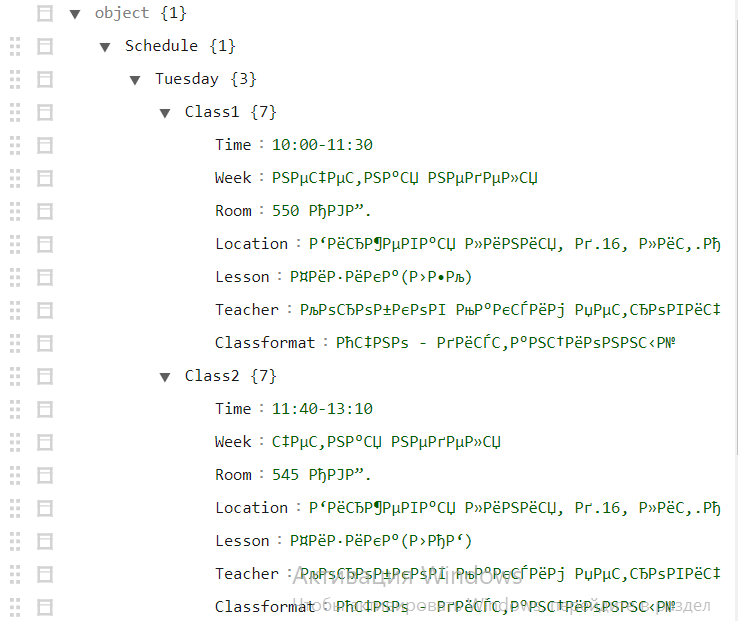
Не поддерживает русские символы, в остальном одно и то же.

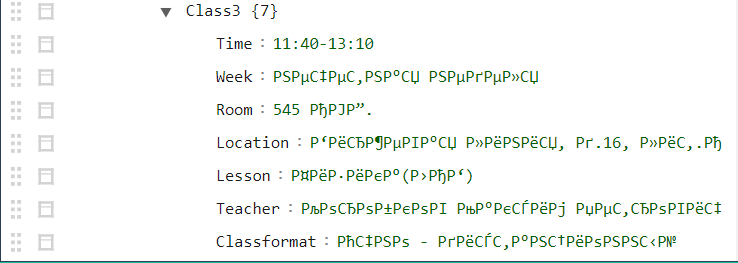
**Вторник.json** на сайте **https://jsonformatter.org/json-editor**





**Результат программы** на сайте **https://jsonformatter.org/json-editor**





**Задание 9**

Результат работы моей программы

T1 = 0.0019232

T2 = 0.0012071

T3 = 0.0011825

T4 = 0.0012828

T5 = 0.0017189

T6 = 0.0046338

T7 = 0.0042771

T8 = 0.0041687

T9 = 0.0043515

T10 = 0.0052615 Tсрсвоя = 0.0030007

Результат работы программы с использованием готовых библиотек

T1 = 0.0077233

T2 = 0.0068973

T3 = 0.0070388

T4 = 0.0062162

T5 = 0.0068486

T6 = 0.0073462

T7 = 0.0063040

T8 = 0.0064788

T9 = 0.0072083

T10 = 0.0116313

Tсрбибл = 0.0073693

В итоге Tсрсвоя < Tсрбибл, это объясняется тем, что через библиотеку исходный код YML проверяется на валидность, также там создаются свои структуры, из-за чего время выполнения увеличивается.

**Выводы.**

Я овладел знаниями о различных современных языках разметки документов, в особенности yaml и json и форматах данных, навыками обработки данных с помощью языка Python 3.x. Также научился измерять время работы программы.