Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

*Факультет Программной инженерии и компьютерной техники*

**Лабораторная работа №4**

Исследование протоколов, форматов обмена информацией и языков разметки документов

Вариант №10

Группа: P3112

Выполнил: Кобелев Р.П.

Проверил:

к.т.н. преподаватель Белозубов А.В.



г.Санкт—Петербург

2022г

Оглавление

[Введение 3](#_Toc120102937)

[Задания 4](#_Toc120102938)

[1 задание 4](#_Toc120102939)

[2 задание 4](#_Toc120102940)

[Input.yaml 4](#_Toc120102941)

[3 задание 5](#_Toc120102942)

[YAML\_XML.py 5](#_Toc120102943)

[output.xml 6](#_Toc120102944)

[4 задание 7](#_Toc120102945)

[YAML\_XML\_Lib.py 7](#_Toc120102946)

[Output1.xml 7](#_Toc120102947)

[5 задание 9](#_Toc120102948)

[YAML\_XML\_REGEX.py 9](#_Toc120102949)

[Output2.xml 9](#_Toc120102950)

[6 задание 11](#_Toc120102951)

[YAML\_XML\_TEST.py 11](#_Toc120102952)

[Вывод программы 13](#_Toc120102953)

[7 задание 14](#_Toc120102954)

[YAML\_CSV.py 14](#_Toc120102955)

[output.csv 14](#_Toc120102956)

[Заключение 16](#_Toc120102957)

[Список литературы 17](#_Toc120102958)

# Введение

В данной лабораторной работе я буду писать алгоритм парсинга и конвертации YAML файла в XML и CSV формат с помощью встроенных библиотек, с регулярными выражениями и без них.

# Задания

## 1 задание

Определить номер варианта как остаток деления на 36 порядкового номера в списке группы в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь.

10%36 = 10 вариант

|  |  |  |  |
| --- | --- | --- | --- |
| **№ Варианта** | **Исходный формат** | **Результирующий формат** | **День недели** |
| 10 | YAML | XML | Вторник |

## 2 задание

Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы в выбранном дне было не менее двух занятий (можно использовать своё персональное). В случае, если в данный день недели нет таких занятий, то увеличить номер варианта ещё на восемь.

### Input.yaml

"ИНФОРМАТИКА(ЛЕК): ZOOM":  
 время: 08:20-09:50  
 неделя: 5  
 Ауд./Корпус: None  
 Преподаватель: Балакшин Павел Валерьевич  
 формат занятий: Дистанционный  
  
"ИНФОРМАТИКА(ЛЕК): АКТОВЫЙ ЗАЛ":  
 время: 08:20-09:50  
 неделя: 3, 7, 9, 11, 13, 15, 17  
 Ауд./Корпус: "ул.Ломоносова, д.9, лит. М"  
 Преподаватель: Балакшин Павел Валерьевич  
 формат занятий: Очно - дистанционный  
  
"ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛЕК): ZOOM":  
 время: 10:00-11:30  
 неделя: 5  
 Ауд./Корпус: None  
 Преподаватель: Клименков Сергей Викторович  
 формат занятий: Дистанционный  
  
"ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛЕК): АКТОВЫЙ ЗАЛ":  
 время: 10:00-11:30  
 неделя: 3, 7, 9, 11, 13, 15, 17  
 Ауд./Корпус: "ул.Ломоносова, д.9, лит. М"  
 Преподаватель: Клименков Сергей Викторович  
 формат занятий: Очно - дистанционный  
  
"ПРОГРАММИРОВАНИЕ(ЛЕК): ZOOM":  
 время: 11:40-13:10  
 неделя: 5  
 Ауд./Корпус: None  
 Преподаватель: Письмак Алексей Евгеньевич  
 формат занятий: Дистанционный  
  
"ПРОГРАММИРОВАНИЕ(ЛЕК)":  
 время: 11:40-13:10  
 неделя: 3, 7, 9, 11, 13, 15, 17  
 Ауд./Корпус: "1223 АУД. ул.Ломоносова, д.9, лит. М"  
 Преподаватель: Письмак Алексей Евгеньевич  
 формат занятий: Очно - дистанционный

## 3 задание

Обязательное задание (позволяет набрать до 65 процентов от максимального числа баллов БаРС за данную лабораторную): написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

### YAML\_XML.py

input = open("input.yaml", "r", encoding="UTF-8")  
output = open("output.xml", "w", encoding="UTF-8")  
condition = True  
lastline = ''  
numdots = 0  
linebool = True  
tag1 = " "  
  
output.write('<Расписание>\n')  
for i in input.readlines():  
 line = i.rstrip()  
 if line == '':  
 condition = True  
 output.write(" </"+lastline+">\n\n")  
 else:  
 if condition == True:  
 condition = False  
 line = line[1:len(line)-2]  
 line = line.replace('(', '.', -1).replace(')','.', -1).replace(' ', '', -1).replace(':', '', -1)  
 output.write(" <"+line+">\n")  
 lastline = line  
 else:  
 line = line.replace('/', '', -1)  
 s = line.split(" ")  
 if s[-1][-1] !=":":  
 for b in range(len(s)):  
 if s[0] !='' and s[0] !='-':  
 break  
 else:  
 s.pop(0)  
  
 output.write(' <'+tag1+'>'+' '.join(s)+"</"+tag1+">\n")  
 else:  
 tag1 = line[1:len(line)-1].replace(' ', '', -1)  
  
output.write(" </"+lastline+">\n")  
output.write('</Расписание>')  
  
input.close()  
output.close()

После парсинга и конвертации данные записываются в output.xml

### output.xml

<Расписание>  
 <ИНФОРМАТИКА.ЛЕК.ZOOM>  
 <время>08:20-09:50</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.ZOOM>  
  
 <ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>08:20-09:50</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
 <время>10:00-11:30</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>10:00-11:30</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
 <время>11:40-13:10</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.>  
 <время>11:40-13:10</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>1223 АУД. ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.>  
</Расписание>

## 4 задание

Дополнительное задание №1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).

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

b) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.

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

### YAML\_XML\_Lib.py

import yaml  
output = open("output1.xml", "w", encoding="UTF-8")  
output.write('<Расписание>\n')  
with open('input.yaml', "r", encoding="UTF-8") as fh:  
 read\_data = yaml.load(fh, Loader=yaml.SafeLoader)  
 for i in read\_data.keys():  
 output.write(' <'+i.replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '', -1)+'>\n')  
 for j in read\_data.get(i):  
 for k in read\_data.get(i).get(j):  
 output.write(' <'+j.replace('/', '', -1).replace(' ', '', -1)  
 +'>'+str(k)+'</'  
 +j.replace('/', '', -1).replace(' ', '', -1)+'>\n')  
 output.write(' </' + i.replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '', -1) + '>\n')  
  
output.write('</Расписание>\n')  
output.close()

После парсинга и конвертации данные записываются в output1.xml

### Output1.xml

<Расписание>  
 <ИНФОРМАТИКА.ЛЕК.ZOOM>  
 <время>08:20-09:50</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.ZOOM>  
 <ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>08:20-09:50</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
 <время>10:00-11:30</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>10:00-11:30</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
 <время>11:40-13:10</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.>  
 <время>11:40-13:10</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>1223 АУД. ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.>  
</Расписание>

Результаты без использования библиотеки и наоборот одинаковы. Это произошло из-за того, что программа отличается только способом считывания данных.

## 5 задание

Дополнительное задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).

a) Переписать исходный код, добавив в него использование

регулярных выражений.

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

### YAML\_XML\_REGEX.py

import re  
  
input = open("input.yaml", "r", encoding="UTF-8")  
output = open("output2.xml", "w", encoding="UTF-8")  
  
pat1 = r'"[а-яА-я():\s\w]+":$'  
pat2 = r'(?<=^\s\s)[а-яА-Я\.\/\s]+(?<!:)'  
pat3 = r'^[ -]+(?<!\w)'  
lastline = ''  
output.write('<Расписание>\n')  
  
condition = True  
  
for i in input.readlines():  
 line = i.rstrip()  
 if line == '':  
 condition = True  
 output.write(" </"+lastline+">\n\n")  
 else:  
 if condition == True:  
 condition = False  
 line = re.search(pat1, line)  
 line = line.group().replace('(', '.', -1).replace(')','.', -1).replace(' ', '', -1).replace(':', '', -1).replace('"', '', -1)  
 output.write(" <"+line+">\n")  
 lastline = line  
 else:  
 if line[-1] != ':':  
 k = re.sub(pat3, '', line)  
 output.write(' <'+tag1+'>'+k+"</"+tag1+">\n")  
 else:  
 tag1 = re.search(pat2, line)  
 tag1 = tag1.group().replace('/', '', -1).replace(' ', '', -1)  
output.write(" </"+lastline+">\n")  
output.write('</Расписание>')  
  
input.close()  
output.close()

После парсинга и конвертации данные записываются в output2.xml

### Output2.xml

<Расписание>  
 <ИНФОРМАТИКА.ЛЕК.ZOOM>  
 <время>08:20-09:50</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.ZOOM>  
  
 <ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>08:20-09:50</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Балакшин Павел Валерьевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ИНФОРМАТИКА.ЛЕК.АКТОВЫЙЗАЛ>  
  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
 <время>10:00-11:30</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.ZOOM>  
  
 <ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
 <время>10:00-11:30</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Клименков Сергей Викторович</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ОСНОВЫПРОФЕССИОНАЛЬНОЙДЕЯТЕЛЬНОСТИ.ЛЕК.АКТОВЫЙЗАЛ>  
  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
 <время>11:40-13:10</время>  
 <неделя>5</неделя>  
 <Ауд.Корпус>None</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.ZOOM>  
  
 <ПРОГРАММИРОВАНИЕ.ЛЕК.>  
 <время>11:40-13:10</время>  
 <неделя>3</неделя>  
 <неделя>7</неделя>  
 <неделя>9</неделя>  
 <неделя>11</неделя>  
 <неделя>13</неделя>  
 <неделя>15</неделя>  
 <неделя>17</неделя>  
 <Ауд.Корпус>1223 АУД. ул.Ломоносова, д.9, лит. М</Ауд.Корпус>  
 <Преподаватель>Письмак Алексей Евгеньевич</Преподаватель>  
 <форматзанятий>Очно - дистанционный</форматзанятий>  
 </ПРОГРАММИРОВАНИЕ.ЛЕК.>  
</Расписание>

Результаты без использования регулярных выражений и наоборот одинаковы. Это произошло из-за того, что программа отличается только способом считывания данных.

## 6 задание

Дополнительное задание №3 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).

a) Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и программу из дополнительного задания №2, сравнить стократное время выполнения парсинга + конвертации в цикле.

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

### YAML\_XML\_TEST.py

import timeit  
  
print("-----------------------------------Мой алгоритм-----------------------------------")  
start1 = timeit.default\_timer()  
  
for i in range(100):  
 input = open("input.yaml", "r", encoding="UTF-8")  
 output = open("output.xml", "w", encoding="UTF-8")  
 condition = True  
 lastline = ''  
 numdots = 0  
 linebool = True  
 tag1 = " "  
  
 output.write('<Расписание>\n')  
 for i in input.readlines():  
 line = i.rstrip()  
 if line == '':  
 condition = True  
 output.write(" </" + lastline + ">\n\n")  
 else:  
 if condition == True:  
 condition = False  
 line = line[1:len(line) - 2]  
 line = line.replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '', -1)  
 output.write(" <" + line + ">\n")  
 lastline = line  
 else:  
 line = line.replace('/', '', -1)  
 s = line.split(" ")  
 if s[-1][-1] != ":":  
 for b in range(len(s)):  
 if s[0] != '' and s[0] != '-':  
 break  
 else:  
 s.pop(0)  
  
 output.write(' <' + tag1 + '>' + ' '.join(s) + "</" + tag1 + ">\n")  
 else:  
 tag1 = line[1:len(line) - 1].replace(' ', '', -1)  
  
 output.write(" </" + lastline + ">\n")  
 output.write('</Расписание>')  
  
 input.close()  
 output.close()  
  
stop1 = timeit.default\_timer()  
print("Test succesfully done\nTime: "+str(stop1-start1))  
  
print("----------------------------Программа с помощью библиотеки------------------------")  
start2 = timeit.default\_timer()  
  
for i in range(100):  
 import yaml  
  
 output = open("output1.xml", "w", encoding="UTF-8")  
 output.write('<Расписание>\n')  
 with open('input.yaml', "r", encoding="UTF-8") as fh:  
 read\_data = yaml.load(fh, Loader=yaml.SafeLoader)  
 for i in read\_data.keys():  
 output.write(  
 ' <' + i.replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '', -1) + '>\n')  
 for j in read\_data.get(i):  
 for k in read\_data.get(i).get(j):  
 output.write(' <' + j.replace('/', '', -1).replace(' ', '', -1)  
 + '>' + str(k) + '</'  
 + j.replace('/', '', -1).replace(' ', '', -1) + '>\n')  
 output.write(' </' + i.replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '',  
 -1) + '>\n')  
  
 output.write('</Расписание>\n')  
 output.close()  
  
stop2 = timeit.default\_timer()  
print("Test succesfully done\nTime: "+str(stop2-start2))  
print("--------------------Программа с помощью регулярных выражений----------------------")  
start3 = timeit.default\_timer()  
  
for i in range(100):  
 import re  
  
 input = open("input.yaml", "r", encoding="UTF-8")  
 output = open("output2.xml", "w", encoding="UTF-8")  
  
 pat1 = r'"[а-яА-я():\s\w]+":$'  
 pat2 = r'(?<=^\s\s)[а-яА-Я\.\/\s]+(?<!:)'  
 pat3 = r'^[ -]+(?<!\w)'  
 lastline = ''  
 output.write('<Расписание>\n')  
  
 condition = True  
  
 for i in input.readlines():  
 line = i.rstrip()  
 if line == '':  
 condition = True  
 output.write(" </" + lastline + ">\n\n")  
 else:  
 if condition == True:  
 condition = False  
 line = re.search(pat1, line)  
 line = line.group().replace('(', '.', -1).replace(')', '.', -1).replace(' ', '', -1).replace(':', '',  
 -1).replace(  
 '"', '', -1)  
 output.write(" <" + line + ">\n")  
 lastline = line  
 else:  
 if line[-1] != ':':  
 k = re.sub(pat3, '', line)  
 output.write(' <' + tag1 + '>' + k + "</" + tag1 + ">\n")  
 else:  
 tag1 = re.search(pat2, line)  
 tag1 = tag1.group().replace('/', '', -1).replace(' ', '', -1)  
 output.write(" </" + lastline + ">\n")  
 output.write('</Расписание>')  
  
 input.close()  
 output.close()  
  
stop3 = timeit.default\_timer()  
print("Test succesfully done\nTime: "+str(stop3-start3))

### Вывод программы

-----------------------------------Мой алгоритм-----------------------------------

Test succesfully done

Time: 0.06726959999650717

----------------------------Программа с помощью библиотеки------------------------

Test succesfully done

Time: 0.6463474999764003

--------------------Программа с помощью регулярных выражений----------------------

Test succesfully done

Time: 0.0701382999541238

Программа с регулярными выражениями работает дольше, чем оригинальная программа, потому что импорт библиотеки для регулярных выражений требует время. Программа с использованием библиотеки работает дольше, чем оригинальная программа, потому что в библиотеке прописано множество функций, которые обеспечивают нам правильную выборку данных.

## 7 задание

Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).

a) Переписать исходную программу, чтобы она осуществляла парсинг и конвертацию исходного файла в любой другой формат (кроме JSON, YAML, XML, HTML): PROTOBUF, TSV, CSV, WML и т.п.

b) Проанализировать полученные результаты, объяснить особенности использования формата.

### YAML\_CSV.py

import csv, yaml  
  
def refractor(g):  
 s = str(g[0])  
 if len(g)>=1:  
 for i in range(1, len(g)):  
 s+=(', '+str(g[i]))  
 return s  
  
with open('input.yaml', "r", encoding="UTF-8") as inp:  
 with open("output.csv", "w", newline='') as out:  
 suc = []  
 fieldnames = ['Предмет']  
 var = False  
 read\_data = yaml.load(inp, Loader=yaml.SafeLoader)  
 for k, v in read\_data.items():  
 row = [k]  
 for a, b in v.items():  
 row.append(refractor(b))  
 if not var:  
 fieldnames+=[a]  
 if not var:  
 writer = csv.writer(out)  
 suc.append(fieldnames)  
 var = True  
 suc.append(row)  
 writer.writerows(suc)

Результат парсинга и конвертации записывается в output.csv

### output.csv

Предмет,время,неделя,Ауд./Корпус,Преподаватель,формат занятий  
ИНФОРМАТИКА(ЛЕК): ZOOM,08:20-09:50,5,None,Балакшин Павел Валерьевич,Дистанционный  
ИНФОРМАТИКА(ЛЕК): АКТОВЫЙ ЗАЛ,08:20-09:50,"3, 7, 9, 11, 13, 15, 17","ул.Ломоносова, д.9, лит. М",Балакшин Павел Валерьевич,Очно - дистанционный  
ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛЕК): ZOOM,10:00-11:30,5,None,Клименков Сергей Викторович,Дистанционный  
ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛЕК): АКТОВЫЙ ЗАЛ,10:00-11:30,"3, 7, 9, 11, 13, 15, 17","ул.Ломоносова, д.9, лит. М",Клименков Сергей Викторович,Очно - дистанционный  
ПРОГРАММИРОВАНИЕ(ЛЕК): ZOOM,11:40-13:10,5,None,Письмак Алексей Евгеньевич,Дистанционный  
ПРОГРАММИРОВАНИЕ(ЛЕК),11:40-13:10,"3, 7, 9, 11, 13, 15, 17","1223 АУД. ул.Ломоносова, д.9, лит. М",Письмак Алексей Евгеньевич,Очно – дистанционный

Формат csv – это текстовый файл с разделителями, который использует запятую для разделения значений. Данный текстовый формат создан для представления табличных данных

# Заключение

В результате выполненной работы я познакомился с языками разметки документов XML, CSV, YAML. Написал алгоритм парсинга и конвертации, ознакомился с библиотеками yaml для чтения «.yaml» файлов. Узнал, почему код с библиотеками и регулярными выражениями работает дольше, чем без них.

# Список литературы

**CSV - Википедия** [В Интернете] // https://ru.wikipedia.org/wiki/CSV. - Wikipedia.

**Пишем изящный парсер на питоне** / Хабр [В Интернете] / авт. saluev // https://habr.com/ru/post/309242/. - Хабр, 5 сентября 2016 г..

**Форма Бэкуса — Наура** — Википедия [В Интернете] // https://ru.wikipedia.org/wiki/Форма\_Бэкуса\_—\_Наура. - Wikipedia.