

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Федеральное государственное автономное
образовательное учреждение высшего образования
«Санкт-Петербургский национальный исследовательский университет ИТМО»

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНОЙ ТЕХНИКИ

ЛАБОРАТОРНАЯ РАБОТА №4

по дисциплине

‘ИНФОРМАТИКА’

«Исследование протоколов, форматов обмена информацией и языков
разметки документов»

Вариант №13

Выполнил:

Студент группы Р3118

Павлов Александр Сергеевич

Преподаватель:

Балакшин Павел Валерьевич



УНИВЕРСИТЕТ ИТМО

Санкт-Петербург, 2021

Оглавление

Задание:.....	3
Основные этапы выполнения:	3
Вывод:.....	19
Список литературы:	19

Задание:

Обязательное задание: Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного JSON-файла в новый формата XML без использования регулярных выражений и готовых библиотек.

Дополнительное задание №1: Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов, переписать исходный код, используя данные библиотеки без использования регулярных выражений. Сравнить полученные результаты и объяснить их сходство/различия.

Дополнительное задание №2: Переписать исходный код, добавив в него использование регулярных выражений, сравнить полученные результаты и объяснить их сходство/различия.

Дополнительное задание №3: Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и дополнительного задания №2, сравнить десятикратное время выполнения парсинга + конвертации в цикле.

Основные этапы выполнения:

Исходный JSON-файл:

```
{
  "day": "Среда",
  "date": 10,
  "lessons": [
    {
      "name": "Программирование(Лаб)",
      "time": "08:20-09:50",
      "weeks": [3,5,7,9,11,13,15,17],
      "address": "Кронверский пр., д.49, лит.А",
      "classroom": "372",
      "teacher": "Сорокин Роман Борисович",
```

```

    "form": "Очно-дистанционная"
  },
  {
    "name": "Программирование(Лаб)",
    "time": "10:00-11:30",
    "weeks": [3,5,7,9,11,13,15,17],
    "address": "Кронверский пр., д.49, лит.А",
    "classroom": "372",
    "teacher": "Сорокин Роман Борисович",
    "form": "Очно-дистанционная"
  },
  {
    "name": "Основы профессиональной деятельности(Лаб)",
    "time": "11:40-13:10",
    "weeks": [3,5,7,9,11,13,15,17],
    "address": "Кронверский пр., д.49, лит.А",
    "classroom": "303/2(Усл)",
    "teacher": "Перминов Илья Валентинович",
    "form": "Очно-дистанционная"
  },
  {
    "name": "Основы профессиональной деятельности(Лаб)",
    "time": "13:30-15:00",
    "weeks": [3,5,7,9,11,13,15,17],
    "address": "Кронверский пр., д.49, лит.А",
    "classroom": "303/2(Усл)",
    "teacher": "Перминов Илья Валентинович",
    "form": "Очно-дистанционная"
  }
}

```

```
]
}
```

Обязательное задание:

```
from codecs import open
f = open('raspisanie.json','r','utf_8_sig')
xml = '<?xml version="1.0" encoding="UTF-8" ?>\n<root>\n\t'
while 1:
    line = f.readline()
    if not line:
        xml += '</root>'
        break
    counter = 0
    teg = ""
    massiv = line.split(':',1)
    teg = massiv[0].replace("\t","")
    if teg[0] == "'":
        soderjimoe = massiv[1]
        if soderjimoe[-4] + soderjimoe[-3] != '],':
            xml += '<' + teg[1:len(teg)-1] + '>'
        if soderjimoe[1] == "" and soderjimoe[-4] + soderjimoe[-3] != '],':
            xml += soderjimoe[2:len(soderjimoe)-4] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
        if ord(soderjimoe[1]) >= 47 and ord(soderjimoe[1]) <= 58:
            xml += soderjimoe[1:len(soderjimoe)-3] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
        if soderjimoe[1] == '[':
            if line[-3] == '[':
                xml += "\n" + "\t"
                const_teg = teg[1:len(teg)-1]
            else:
                weeks = soderjimoe[2:len(soderjimoe)-4].split(',')

```

```

for c in weeks:

    xml += '<' + teg[1:len(teg)-1] + '>' + c + '</' + teg[1:len(teg)-1] + '>' + '\n' + '\t'

if '}',' in line:

    xml += '</' + const_teg + '>' + '\n' + '\t'

    xml += '<' + const_teg + '>' + '\n' + '\t'

if '}' in line:

    if line [0] != '}' and line[-3] == '}':

        xml += '</' + const_teg + '>' + '\n'

file = open("raspisanie.xml", 'w', 'utf_8_sig')

file.write(xml)

file.close()

```

Результат выполнения(отдельный XML-файл):

```

<?xml version="1.0" encoding="UTF-8" ?>

<root>

    <day>Среда</day>

    <date>10</date>

    <lessons>

        <name>Программирование(Лаб)</name>

        <time>08:20-09:50</time>

        <weeks>3</weeks>

        <weeks>5</weeks>

        <weeks>7</weeks>

        <weeks>9</weeks>

        <weeks>11</weeks>

        <weeks>13</weeks>

        <weeks>15</weeks>

        <weeks>17</weeks>

        <address>Кронверский пр., д.49, лит.А</address>

        <classroom>372</classroom>

```

<teacher>Сорокин Роман Борисович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Программирование(Лаб)</name>
<time>10:00-11:30</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>
<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>372</classroom>
<teacher>Сорокин Роман Борисович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Основы профессиональной деятельности(Лаб)</name>
<time>11:40-13:10</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>

```

<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>303/2(Усл)</classroom>
<teacher>Перминов Илья Валентинович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Основы профессиональной деятельности(Лаб)</name>
<time>13:30-15:00</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>
<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>303/2(Усл)</classroom>
<teacher>Перминов Илья Валентинович</teacher>
<form>Очно-дистанционна</form>
</lessons>
</root>

```

Дополнительное задание №1:

```

from json2xml import json2xml
from json2xml.utils import readfromstring
from codecs import open
f = open('raspisanie.json','r','utf_8_sig').read()
data = readfromstring(f)

```



```

xml = json2xml.Json2xml(data).to_xml()
file = open('raspisanie_libr.xml','w','utf_8_sig')
file.write(xml)
file.close()

```

Результат выполнения(отдельный XML-файл):

```

<?xml version="1.0" ?>
<all>
  <day type="str">Среда</day>
  <date type="int">10</date>
  <lessons type="list">
    <item type="dict">
      <name type="str">Программирование(Лаб)</name>
      <time type="str">08:20-09:50</time>
      <weeks type="list">
        <item type="int">3</item>
        <item type="int">5</item>
        <item type="int">7</item>
        <item type="int">9</item>
        <item type="int">11</item>
        <item type="int">13</item>
        <item type="int">15</item>
        <item type="int">17</item>
      </weeks>
      <address type="str">Кронверский пр., д.49, лит.А</address>
      <classroom type="str">372</classroom>
      <teacher type="str">Сорокин Роман Борисович</teacher>
      <form type="str">Очно-дистанционная</form>
    </item>
    <item type="dict">

```

```

<name type="str">Программирование(Лаб)</name>
<time type="str">10:00-11:30</time>
<weeks type="list">
    <item type="int">3</item>
    <item type="int">5</item>
    <item type="int">7</item>
    <item type="int">9</item>
    <item type="int">11</item>
    <item type="int">13</item>
    <item type="int">15</item>
    <item type="int">17</item>
</weeks>
<address type="str">Кронверский пр., д.49, лит.А</address>
<classroom type="str">372</classroom>
<teacher type="str">Сорокин Роман Борисович</teacher>
<form type="str">Очно-дистанционная</form>
</item>
<item type="dict">
    <name type="str">Основы профессиональной деятельности(Лаб)</name>
    <time type="str">11:40-13:10</time>
    <weeks type="list">
        <item type="int">3</item>
        <item type="int">5</item>
        <item type="int">7</item>
        <item type="int">9</item>
        <item type="int">11</item>
        <item type="int">13</item>
        <item type="int">15</item>
        <item type="int">17</item>
    </weeks>
    <address type="str">Кронверский пр., д.49, лит.А</address>
    <classroom type="str">372</classroom>
    <teacher type="str">Сорокин Роман Борисович</teacher>
    <form type="str">Очно-дистанционная</form>
    </item>
    </dict>
</item>

```

```

</weeks>
<address type="str">Кронверский пр., д.49, лит.А</address>
<classroom type="str">303/2(Усл)</classroom>
<teacher type="str">Перминов Илья Валентинович</teacher>
<form type="str">Очно-дистанционная</form>
</item>
<item type="dict">
  <name type="str">Основы профессиональной деятельности(Лаб)</name>
  <time type="str">13:30-15:00</time>
  <weeks type="list">
    <item type="int">3</item>
    <item type="int">5</item>
    <item type="int">7</item>
    <item type="int">9</item>
    <item type="int">11</item>
    <item type="int">13</item>
    <item type="int">15</item>
    <item type="int">17</item>
  </weeks>
  <address type="str">Кронверский пр., д.49, лит.А</address>
  <classroom type="str">303/2(Усл)</classroom>
  <teacher type="str">Перминов Илья Валентинович</teacher>
  <form type="str">Очно-дистанционная</form>
</item>
</lessons>
</all>

```

Дополнительное задание №2:

```
from codecs import open
import re
f = open('raspisanie.json','r','utf_8_sig')
xml = '<?xml version="1.0" encoding="UTF-8" ?>\n<root>\n\t'
while 1:
    line = f.readline()
    if not line:
        xml += '</root>'
        break
    counter = 0
    teg = ""
    massiv = line.split(':',1)
    teg = re.sub('\t',"",massiv[0])
    if teg[0] == "'":
        soderjimoe = massiv[1]
        if soderjimoe[-4] + soderjimoe[-3] != '],':
            xml += '<' + teg[1:len(teg)-1] + '>'
        if soderjimoe[1] == "" and soderjimoe[-4] + soderjimoe[-3] != '],':
            xml += soderjimoe[2:len(soderjimoe)-4] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
        if ord(soderjimoe[1]) >= 47 and ord(soderjimoe[1]) <= 58:
            xml += soderjimoe[1:len(soderjimoe)-3] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
        if soderjimoe[1] == '[':
            if line[-3] == '[':
                xml += "\n" + "\t"
                const_teg = teg[1:len(teg)-1]
            else:
                weeks = soderjimoe[2:len(soderjimoe)-4].split(',')
                for c in weeks:
```

```

xml += '<' + teg[1:len(teg)-1] + '>' + c + '<' + teg[1:len(teg)-1] + '>' + '\n' + '\t'
p = re.findall('}',line)
p1 = re.findall(r'}[^,]',line)
if len(p) != 0:
    xml += '<' + const_teg + '>' + '\n' + '\t'
    xml += '<' + const_teg + '>' + '\n' + '\t'
if len(p1) != 0:
    xml += '<' + const_teg + '>' + '\n'
file = open("raspisanie_regex.xml",'w','utf_8_sig')
file.write(xml)
file.close()

```

Результат выполнения(отдельный XML-файл):

```

<?xml version="1.0" encoding="UTF-8" ?>
<root>
    <day>Среда</day>
    <date>10</date>
    <lessons>
        <name>Программирование(Лаб)</name>
        <time>08:20-09:50</time>
        <weeks>3</weeks>
        <weeks>5</weeks>
        <weeks>7</weeks>
        <weeks>9</weeks>
        <weeks>11</weeks>
        <weeks>13</weeks>
        <weeks>15</weeks>
        <weeks>17</weeks>
        <address>Кронверский пр., д.49, лит.А</address>
        <classroom>372</classroom>

```

<teacher>Сорокин Роман Борисович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Программирование(Лаб)</name>
<time>10:00-11:30</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>
<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>372</classroom>
<teacher>Сорокин Роман Борисович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Основы профессиональной деятельности(Лаб)</name>
<time>11:40-13:10</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>

```

<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>303/2(Усл)</classroom>
<teacher>Перминов Илья Валентинович</teacher>
<form>Очно-дистанционна</form>
</lessons>
<lessons>
<name>Основы профессиональной деятельности(Лаб)</name>
<time>13:30-15:00</time>
<weeks>3</weeks>
<weeks>5</weeks>
<weeks>7</weeks>
<weeks>9</weeks>
<weeks>11</weeks>
<weeks>13</weeks>
<weeks>15</weeks>
<weeks>17</weeks>
<address>Кронверский пр., д.49, лит.А</address>
<classroom>303/2(Усл)</classroom>
<teacher>Перминов Илья Валентинович</teacher>
<form>Очно-дистанционна</form>
</lessons>
</root>

```

Дополнительное задание №3:

```

from codecs import open
import re
from json2xml import json2xml
from json2xml.utils import readfromstring
from time import time

```

```

start_time_1 = time()
for i in range(10):
    f = open('raspisanie.json','r','utf_8_sig')
    xml = '<?xml version="1.0" encoding="UTF-8" ?>\n<root>\n\t'
    while 1:
        line = f.readline()
        if not line:
            xml += '</root>'
            break
        counter = 0
        teg = ""
        massiv = line.split(':',1)
        teg = massiv[0].replace('\t','')
        if teg[0] == "'":
            soderjimoe = massiv[1]
            if soderjimoe[-4] + soderjimoe[-3] != '],':
                xml += '<' + teg[1:len(teg)-1] + '>'
            if soderjimoe[1] == "" and soderjimoe[-4] + soderjimoe[-3] != '],':
                xml += soderjimoe[2:len(soderjimoe)-4] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
            if ord(soderjimoe[1]) >= 47 and ord(soderjimoe[1]) <= 58:
                xml += soderjimoe[1:len(soderjimoe)-3] + '</' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
            if soderjimoe[1] == '[':
                if line[-3] == '[':
                    xml += "\n" + "\t"
                    const_teg = teg[1:len(teg)-1]
                else:
                    weeks = soderjimoe[2:len(soderjimoe)-4].split(',')
                    for c in weeks:
                        xml += '<' + teg[1:len(teg)-1] + '>' + c + '</' + teg[1:len(teg)-1] + '>' + '\n' + '\t'

```



```

if '}',' in line:
    xml += '<' + const_teg + '>' + '\n' + '\t'
    xml += '<' + const_teg + '>' + '\n' + '\t'
if '}' in line:
    if line [0] != '}' and line[-3] == '}'':
        xml += '<' + const_teg + '>' + '\n'
file = open("raspisanie.xml",'w','utf_8_sig')
file.write(xml)
file.close()
print(time()-start_time_1)
#libraries
start_time_2 = time()
for i in range(10):
    f = open('raspisanie.json','r','utf_8_sig').read()
    data = readfromstring(f)
    xml = json2xml.Json2xml(data).to_xml()
    file = open('raspisanie_libr.xml','w','utf_8_sig')
    file.write(xml)
    file.close()
print(time()-start_time_2)
#regex
start_time_3 = time()
for i in range(10):
    f = open('raspisanie.json','r','utf_8_sig')
    xml = '<?xml version="1.0" encoding="UTF-8" ?>\n<root>\n\t'
    while 1:
        line = f.readline()
        if not line:
            xml += '</root>'

```

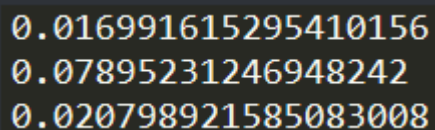
```

break
counter = 0
teg = ""
massiv = line.split(':',1)
teg = re.sub('\t',"",massiv[0])
if teg[0] == "":
    soderjimoe = massiv[1]
    if soderjimoe[-4] + soderjimoe[-3] != '],':
        xml += '<' + teg[1:len(teg)-1] + '>'
    if soderjimoe[1] == "" and soderjimoe[-4] + soderjimoe[-3] != '],':
        xml += soderjimoe[2:len(soderjimoe)-4] + '<' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
    if ord(soderjimoe[1]) >= 47 and ord(soderjimoe[1]) <= 58:
        xml += soderjimoe[1:len(soderjimoe)-3] + '<' + teg[1:len(teg)-1] + '>' + "\n" + "\t"
    if soderjimoe[1] == '[':
        if line[-3] == '[':
            xml += "\n" + "\t"
            const_teg = teg[1:len(teg)-1]
        else:
            weeks = soderjimoe[2:len(soderjimoe)-4].split(',')
            for c in weeks:
                xml += '<' + teg[1:len(teg)-1] + '>' + c + '<' + teg[1:len(teg)-1] + '>' + '\n' + '\t'
p = re.findall('}',',',line)
p1 = re.findall(r'}[^,]',line)
if len(p) != 0:
    xml += '<' + const_teg + '>' + '\n' + '\t'
    xml += '<' + const_teg + '>' + '\n' + '\t'
if len(p1) != 0:
    xml += '<' + const_teg + '>' + '\n'
file = open("raspisanie_regex.xml",'w','utf_8_sig')

```

```
file.write(xml)
file.close()
print(time()-start_time_3)
```

Результаты выполнения:



```
0.016991615295410156
0.07895231246948242
0.020798921585083008
```

Рисунок 1 – результат выполнения дополнительного задания №3, где первое число – время выполнения обязательного задания, второе число – время выполнения первого дополнительного задания(с использованием библиотек), третье число – время выполнения второго дополнительного задания(с использованием регулярных выражений)

Вывод:

Отличие итоговых XML-файлов, полученных в результате парсинга JSON-файла разными способами объясняется в неоднозначности моделирования объектов в XML – один объект можно представить по-разному.

В ходе выполнения данной лабораторной работы я познакомился с такими форматами хранения информации, как JSON, YAML и XML. Научился переводить JSON-формат в XML-формат.

Список литературы:

1. Презентация «Современные языки разметки» Балакшин П.В – 2021 год [Электронный ресурс]. – URL:
https://isu.ifmo.ru/pls/apex/f?p=2002:0:106072181065638:DWNLD_F:NO::FILE:0F46B554A5A631053633DCD0435A10E6
2. Статья «YAML vs JSON vs XML | What is the Difference Between Them?» Aniruddha Chaudhari – 2019 год [Электронный ресурс]. – URL:
<https://www.csestack.org/yaml-vs-json-vs-xml-difference/>