Университет ИТМО

Факультет программной инженерии и компьютерной техники

Лабораторная работа №4

по информатике

«Исследование протоколов,

форматов обмена информацией и языков разметки

документов»

Вариант №20

Выполнил:

Степанов Артур

Группа: P3112

Приняла:

Рудникова Тамара Владимировна

???

?

г. Санкт-Петербург

2021 г.

Оглавление

[Основное задание 2](#_Toc90362357)

[Текст задания 2](#_Toc90362358)

[Исходный код программы 3](#_Toc90362359)

[Вывод программы 4](#_Toc90362360)

[Доп. Задание №1 4](#_Toc90362361)

[Текст задания 4](#_Toc90362362)

[Исходный код программы 5](#_Toc90362363)

[Вывод программы 5](#_Toc90362364)

[Доп. Задание №2 6](#_Toc90362365)

[Текст задания 6](#_Toc90362366)

[Исходный код программы 6](#_Toc90362367)

[Вывод программы 7](#_Toc90362368)

[Доп. Задание №2 8](#_Toc90362369)

[Вывод 8](#_Toc90362370)

# Основное задание

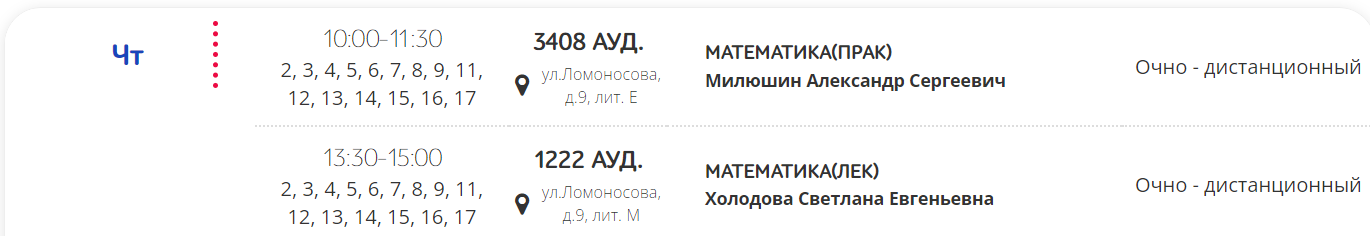
## Текст задания

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.

Расписание занятий на четверг:



Расписание на четверг в формате XML:

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16 | <schedule>  <day name="Четверг">  <lesson name="Математика(ПРАК)">  <time time="10:00-11:30"/>  <place audience="3408" building="ул.Ломоносова, д.9,лит. Е"/>  <teacher name="Милюшин Александр Сергеевич"/>  <format format="Очно - дистанционный"/>  </lesson>  <lesson name="Математика(ЛЕК)">  <time time="13:30-15:00"/>  <place building="ул.Ломоносова, д.9, лит. М"/>  <teacher name="Холодова Светлана Евгеньевна"/>  <format format="Очно - дистанционный"/>  </lesson>  </day>  </schedule> |

## Исходный код программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57 | file = open('xml.txt', 'r', encoding="utf-8")  json = open('json.txt', 'w')  xml = file.read()  xml2=[]  k=len(xml)-**1**  **for** i **in** range(k):  j=i+**1**  w=j+**1**  **if** xml[i]=='<' **and** xml[j]=='/' **and** xml[w]=='l':  xml = xml.replace('</lesson', '} ', **1**)  print('}', end ="", file=json)    **if** xml[i]=='<' **and** xml[j]=='/' **and** xml[w]=='d':  xml = xml.replace('</day', '} ', **1**)      **if** xml[i]=='<' **and** xml[j]=='/' **and** xml[w]=='s':  xml = xml.replace('</schedule', '} ', **1**)    **if** xml[i]=='e' **and** xml[j]=='>':  xml = xml.replace('> ', '**\"**: {', **1**)    **if** xml[i]=='<':  xml = xml.replace('<', '**\"**', **1**)    **if** xml[i]=='>':  xml = xml.replace('>', ' ', **1**)  **if** xml[i]=='=':  xml = xml.replace('=', ' ', **1**)  print('**\"**:', end ="", file=json)    **if** xml[i]==' ' **and** xml[j]=='n':  print('**\"**:{', file=json)  xml = xml.replace(' n', '**\"**n', **1**)    **if** xml[i]==' ' **and** xml[j]=='t':  print('**\"**:{', file=json)  xml = xml.replace(' t', '**\"**t', **1**)    **if** xml[i]==' ' **and** xml[j]=='a':  print('**\"**:{', file=json)  xml = xml.replace(' a', '**\"**a', **1**)    **if** xml[i]==' ' **and** xml[j]=='f':  print('**\"**:{', file=json)  xml = xml.replace(' f', '**\"**f', **1**)  **if** xml[i]==' ' **and** xml[j]=='b':  print('**\"**:{', file=json)  xml = xml.replace(' b', '**\"**b', **1**)    **if** xml[i]=='/':  xml = xml.replace('/', '}', **1**)  print(xml[i], end ="", file=json)  json.close() |

## Вывод программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28 | "schedule": {  "day":{  "name": "Четверг"  "lesson":{  "name": "Математика(ПРАК)"  "time":{  "time": "10:00-11:30"}  "place":{  "audience": "3408"":{  "building": "ул.Ломоносова, д.9,лит. Е"}  "teacher":{  "name": "Милюшин Александр Сергеевич"}  "format":{  "format": "Очно - дистанционный"}  }}  "lesson":{  "name": "Математика(ЛЕК)"  "time":{  "time": "13:30-15:00"}  "place":{  "building": "ул.Ломоносова, д.9, лит. М"}  "teacher":{  "name": "Холодова Светлана Евгеньевна"}  "format":{  "format": "Очно - дистанционный"}  }}  }  } |

# Доп. Задание №1

## Текст задания

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.

## Исходный код программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12 | **import** **json**  **import** **xmltodict**  **with** open("xml.txt", encoding="utf-8") **as** xml\_file:  data\_dict = xmltodict.parse(xml\_file.read())  xml\_file.close()    json\_data = json.dumps(data\_dict)    **with** open("json1.txt", "w") **as** json\_file:  json\_file.write(json\_data)  json\_file.close() |

## Вывод программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20 | {"schedule": {"day":  {"@name": "\u0427\u0435\u0442\u0432\u0435\u0440\u0433",  "lesson": [{  "@name": "\u041c\u0430\u0442\u0435\u043c\u0430\u0442\u0438\u043a\u0430(\u041f\u0420\u0410\u041a)",  "time":  {"@time": "10:00-11:30"},  "place":  {"@audience": "3408", "@building": "\u0443\u043b.\u041b\u043e\u043c\u043e\u043d\u043e\u0441\u043e\u0432\u0430, \u0434.9,\u043b\u0438\u0442. \u0415"},  "teacher":  {"@name": "\u041c\u0438\u043b\u044e\u0448\u0438\u043d \u0410\u043b\u0435\u043a\u0441\u0430\u043d\u0434\u0440 \u0421\u0435\u0440\u0433\u0435\u0435\u0432\u0438\u0447"},  "format":  {"@format": "\u041e\u0447\u043d\u043e - \u0434\u0438\u0441\u0442\u0430\u043d\u0446\u0438\u043e\u043d\u043d\u044b\u0439"}},  {"@name": "\u041c\u0430\u0442\u0435\u043c\u0430\u0442\u0438\u043a\u0430(\u041b\u0415\u041a)",  "time": {"@time": "13:30-15:00"},  "place":  {"@building": "\u0443\u043b.\u041b\u043e\u043c\u043e\u043d\u043e\u0441\u043e\u0432\u0430, \u0434.9, \u043b\u0438\u0442. \u041c"},  "teacher":  {"@name": "\u0425\u043e\u043b\u043e\u0434\u043e\u0432\u0430 \u0421\u0432\u0435\u0442\u043b\u0430\u043d\u0430 \u0415\u0432\u0433\u0435\u043d\u044c\u0435\u0432\u043d\u0430"},  "format":  {"@format": "\u041e\u0447\u043d\u043e - \u0434\u0438\u0441\u0442\u0430\u043d\u0446\u0438\u043e\u043d\u043d\u044b\u0439"}}]}}} |

# Доп. Задание №2

## Текст задания

a) Переписать исходный код, добавив в него использование

регулярных выражений.

b) Сравнить полученные результаты и объяснить их

сходство/различие.

## Исходный код программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63 | **import** **re**  file = open('xml.txt', 'r', encoding="utf-8")  json = open('json.txt', 'w')  xml = file.read()  xml2=[]  k=len(xml)-**1**  result = re.search(r"<", xml)  result1 = re.search(r">", xml)  a=result.group(**0**)  a1=result1.group(**0**)  **for** i **in** range(k):  j=i+**1**  w=j+**1**  **if** xml[i]==a **and** xml[j]=='/' **and** xml[w]=='l':  xml = xml.replace('</lesson', '} ', **1**)  print('}', end ="", file=json)    **if** xml[i]==a1 **and** xml[j]=='/' **and** xml[w]=='d':  xml = xml.replace('</day', '} ', **1**)      **if** xml[i]=='<' **and** xml[j]=='/' **and** xml[w]=='s':  xml = xml.replace('</schedule', '} ', **1**)    **if** xml[i]=='e' **and** xml[j]=='>':  xml = xml.replace('> ', '**\"**: {', **1**)    **if** xml[i]=='<':  xml = xml.replace('<', '**\"**', **1**)    **if** xml[i]=='>':  xml = xml.replace('>', ' ', **1**)  **if** xml[i]=='=':  xml = xml.replace('=', ' ', **1**)  print('**\"**:', end ="", file=json)    **if** xml[i]==' ' **and** xml[j]=='n':  print('**\"**:{', file=json)  xml = xml.replace(' n', '**\"**n', **1**)    **if** xml[i]==' ' **and** xml[j]=='t':  print('**\"**:{', file=json)  xml = xml.replace(' t', '**\"**t', **1**)    **if** xml[i]==' ' **and** xml[j]=='a':  print('**\"**:{', file=json)  xml = xml.replace(' a', '**\"**a', **1**)    **if** xml[i]==' ' **and** xml[j]=='f':  print('**\"**:{', file=json)  xml = xml.replace(' f', '**\"**f', **1**)  **if** xml[i]==' ' **and** xml[j]=='b':  print('**\"**:{', file=json)  xml = xml.replace(' b', '**\"**b', **1**)    **if** xml[i]=='/':  xml = xml.replace('/', '}', **1**)  print(xml[i], end ="", file=json)  json.close() |

## Вывод программы

|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28 | "schedule": {  "day":{  "name": "Четверг"  "lesson":{  "name": "Математика(ПРАК)"  "time":{  "time": "10:00-11:30"}  "place":{  "audience": "3408"":{  "building": "ул.Ломоносова, д.9,лит. Е"}  "teacher":{  "name": "Милюшин Александр Сергеевич"}  "format":{  "format": "Очно - дистанционный"}  }}  "lesson":{  "name": "Математика(ЛЕК)"  "time":{  "time": "13:30-15:00"}  "place":{  "building": "ул.Ломоносова, д.9, лит. М"}  "teacher":{  "name": "Холодова Светлана Евгеньевна"}  "format":{  "format": "Очно - дистанционный"}  }}  "}day  } |

# Доп. Задание №2

0.020003318786621094 seconds – время работы программы без библиотек и регулярных выражений.

0.2800726890563965 seconds – время работы программы с готовыми библиотеками.

0.0020008087158203125 seconds – время работы программы с регулярными выражениями.

# Вывод

Я понял, как использовать простой парсер и узнал как пользоваться готовыми библиотеками для парсинга.