

Федеральное государственное автономное образовательное
учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной
техники

Направление подготовки 09.03.04 «Программная
инженерия»

Системное и прикладное программное обеспечение

Отчёт

По лабораторной работе №4

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

Вариант: 5

Работу выполнил:

Поленов Кирилл Александрович

Группа Р3113

Работу принял:

Рыбаков Степан Дмитриевич

г. Санкт-Петербург 2023

Оглавление

Основное задание	3
Доп. Задание 1	5
Доп. Задание 2	7
Заключение.....	8
Список литературы.....	11

Основное задание

Задание (обязательное)

Написать парсер из XML в YAML без использования сторонних библиотек и регулярных выражений

Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый путём простой замены метасимволов исходного формата на метасимволы результирующего формата.

Рисунок 1

Исходный файл (используется в основном задании и первом и втором дополнительных заданиях)

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <schedule>
3    <day name="Понедельник">
4      <classes>
5        <class>
6          <title>Мат. анализ</title>
7          <type>Лекция</type>
8          <weeks>Чётная, Нечётная</weeks>
9          <begin_time>08:20</begin_time>
10         <end_time>09:50</end_time>
11         <teacher>Рванова Алла Сергеевна</teacher>
12         <building>Кронверкский пр., д. 49, лит. А</building>
13         <classroom>1404</classroom>
14       </class>
15
16       <class>
17         <title>Мат. анализ</title>
18         <type>Практика</type>
19         <weeks>Чётная, Нечётная</weeks>
20         <begin_time>10:00</begin_time>
21         <end_time>11:30</end_time>
22         <teacher>Исаева Татьяна Тимофеевна</teacher>
23         <building>Кронверкский пр., д. 49, лит. А</building>
24         <classroom>2412</classroom>
25       </class>
26     </classes>
27   </day>
28 </schedule>
```

Рисунок 3

Программа

```
1  """
2  Вариант 5
3  XML ---> YAML
4  день недели: Понедельник
5  """
6
7  scheduleXml = open('schedule.xml', mode='r', encoding='utf-8')
8  scheduleYaml = open('task1_schedule.yaml', mode='a', encoding='utf-8')
9  schedule = [line.rstrip() for line in scheduleXml]
10
11  for line in range(1, len(schedule)):
12      s = schedule[line]
13      for symbol in range(0, len(schedule[line])-1):
14          if s[symbol] == '<' and s[symbol + 1] == '/':
15              s = s[:symbol]
16              break
17          if s[symbol] == '<':
18              s = s.replace(s[symbol], '*', 1)
19          if s[symbol] != '/' and s[symbol+1] == '>':
20              s = s.replace(s[symbol+1], '&', 1)
21          if s[symbol] == s[symbol+1] == ' ':
22              s = s.replace(s[symbol], '?', 1)
23              s = s.replace(s[symbol+1], '^', 1)
24
25      s = s.replace('*', '').replace('&', ': ').replace('?^', ' ')
26      if s == '      class: ':
27          s = '      - class: '
28      if s == '  day name="Понедельник": ':
29          s = "- day: \n      '@name': Понедельник"
30
31      scheduleYaml.write(s + '\n')
32  scheduleXml.close()
33  scheduleYaml.close()
```

Рисунок 4

Тест основного задания

```
1  schedule:
2  - day:
3      '@name': Понедельник
4      classes:
5      - class:
6          title: Мат. анализ
7          type: Лекция
8          weeks: Чётная, Нечётная
9          begin_time: 08:20
10         end_time: 09:50
11         teacher: Рванова Алла Сергеевна
12         building: Кронверкский пр., д. 49, лит. А
13         classroom: 1404
14
15
16     - class:
17         title: Мат. анализ
18         type: Практика
19         weeks: Чётная, Нечётная
20         begin_time: 10:00
21         end_time: 11:30
22         teacher: Исаева Татьяна Тимофеевна
23         building: Кронверкский пр., д. 49, лит. А
24         classroom: 2412
25
26
27
```

Рисунок 5

Доп. Задание 1

Доп. 1

Переписать код с библиой для парсинга

- Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
- Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
- Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Рисунок 6

Программа

```
1  """
2  использую библиотеку xmlplain 1.6.0
3  """
4  import xmlplain
5
6  with open('schedule.xml', encoding='utf-8') as inf:
7      xmlSchedule = xmlplain.xml_to_obj(inf, strip_space=True, fold_dict=True)
8
9  with open('task2_schedule.yaml', 'a', encoding='utf-8') as outf:
10     xmlplain.obj_to_yaml(xmlSchedule, outf)
```

Рисунок 7

Тест доп. задания 1

```
1  schedule:
2    day:
3      '@name': Понедельник
4    classes:
5      - class:
6          title: Мат. анализ
7          type: Лекция
8          weeks: Чётная, Нечётная
9          begin_time: 08:20
10         end_time: 09:50
11         teacher: Рванова Алла Сергеевна
12         building: Кронверкский пр., д. 49, лит. А
13         classroom: '1404'
14      - class:
15          title: Мат. анализ
16          type: Практика
17          weeks: Чётная, Нечётная
18          begin_time: '10:00'
19          end_time: '11:30'
20          teacher: Исаева Татьяна Тимофеевна
21          building: Кронверкский пр., д. 49, лит. А
22          classroom: '2412'
```

Рисунок 5

Вывод: библиотеки значительно ускоряют написание кода и делают его компактнее, не сильно ухудшая его качество, но уменьшая возможность разработчика в обслуживании кода.

Доп. Задание 2

Доп. 2

Переписать код с использованием регулярок

- а) Переписать исходный код, добавив в него использование регулярных выражений.
- б) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Рисунок 6

Программа

```
1  import re
2
3  openTag = r'<([^\s]\w+)>'
4  closeTag = r'</\w+>'
5  lstClass = r'<class>'
6  lstDay = r'<day name="(\w+)">'
7  tab = r'<tab>'
8
9  scheduleXml = open('schedule.xml', mode='r', encoding='utf-8')
10 scheduleYaml = open('task3_schedule.yaml', mode='a', encoding='utf-8')
11
12 s = scheduleXml.read()[39:]
13
14 s = re.sub(lstClass, ' - class:', s)
15 s = re.sub(lstDay, "- day:\n '@name': '\\1", s)
16 s = re.sub(openTag, '\\1: ', s)
17 s = re.sub(closeTag, '', s)
18 s = re.sub(tab, ' ', s)
19
20 scheduleYaml.write(s)
21 scheduleXml.close()
22 scheduleYaml.close()
```

Рисунок 7

Тест доп. задания 2

```
1  schedule:
2    - day:
3      '@name': Понедельник
4      classes:
5        - class:
6          title: Мат. анализ
7          type: Лекция
8          weeks: Чётная, Нечётная
9          begin_time: 08:20
10         end_time: 09:50
11         teacher: Рванова Алла Сергеевна
12         building: Кронверкский пр., д. 49, лит. А
13         classroom: 1404
14
15
16     - class:
17       title: Мат. анализ
18       type: Практика
19       weeks: Чётная, Нечётная
20       begin_time: 10:00
21       end_time: 11:30
22       teacher: Исаева Татьяна Тимофеевна
23       building: Кронверкский пр., д. 49, лит. А
24       classroom: 2412
25
26
27
```

Рисунок 8

Вывод: код стал значительно компактнее, на его написание я затратил меньше времени, соответственно, регулярные выражения значительно упрощают реализацию программного продукта.

Доп. Задание 3

Доп. 3

Сделать код применимым к общему случаю парсинга (в моём случае любой XML должен быть правильно конвертирован в YAML)

- а) Переписать исходный код таким образом, чтобы для решения задачи использовались формальные грамматики. То есть ваш код должен уметь осуществлять парсинг и конвертацию любых данных, представленных в исходном формате, в данные, представленные в результирующем формате: как с готовыми библиотеками из дополнительного задания №1.
- б) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
- в) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Рисунок 6

Исходный файл

https://github.com/bilyardvmetro/ITMO-System-Application-Software/blob/main/Informatics/Labs/Lab4/lab4/task4_schedule.xml

1

Программа

<https://github.com/bilyardvmetro/ITMO-System-Application-Software/blob/main/Informatics/Labs/Lab4/lab4/task4.py>

Тест доп. задания 3

https://github.com/bilyardvmetro/ITMO-System-Application-Software/blob/main/Informatics/Labs/Lab4/lab4/task4_schedule.yaml

Вывод: код стал применим к общему случаю парсинга и выходной файл выглядит значительно привлекательнее, однако сложность кода разительно отличается от предыдущих программ.

Заключение

В ходе выполнения данной лабораторной работы я познакомился с синтаксисом языков разметки XML и YAML. Изучил их назначение и особенности, написал собственный парсер из XML в YAML, используя формальные грамматики.

Список литературы

Формальные языки и грамматики . – URL:

<https://habr.com/ru/articles/177109/> (Дата обращения: 15.11.2023)

Регулярные выражения в Python от простого к сложному.

Подробности, примеры, картинки, упражнения. – URL:

<https://habr.com/ru/articles/349860/> (Дата обращения: 18.11.2023)

**П.В. Балакшин, В.В. Соснин, И.В. Калинин, Т.А. Малышева,
С.В. Раков, Н.Г. Рущенко, А.М. Дергачев Информатика:**

лабораторные работы и тесты [Электронный ресурс] –

https://t.me/balakshin_students (Дата обращения: 13.11.2023)