

Федеральное государственное автономное образовательное учреждение высшего образования  
«Санкт-Петербургский национальный исследовательский  
университет информационных технологий, механики и оптики»

**Факультет Программной Инженерии и Компьютерной Техники**

**Дисциплина: Информатика**

**Лабораторная работа №3  
«Исследование языков разметки документов»**

Выполнил: Конаныхина Антонина  
Александровна

Группа: P3115

Вариант: 14

Преподаватель: Балакшин Павел Валерьевич

Санкт-Петербург, 2020г

**Вариант:**

№ Варианта	Исходный формат	Результирующий формат	День недели
14	XML	YAML	Среда

**Задание:** Написать свой парсер на языке Python и выполнить конвертацию файла из формата XML в YAML. Сравнить десятикратное время выполнения программы собственного парсера и готовых библиотек, выяснить, что работает быстрее.

**Основной код программы:**

```
p = open("lab3.xml", "r") #открытие файла исходника
o = open("lab3.yaml", "w") #открытие пустого файла
prevStr = 1
lastTag = 0
for line in p:
    #поиск строк, содержащих открывающий и закрывающий тэги
    одновременно
    if line.count("<") == 2:
        ln = line.split("<") #запись в массив частей строки, разделённых «<»
        if prevStr == 0:
            #Если это первая строка в группе
            o.write(ln[0] + " - " + (ln[1].replace(">", ":")) + "\n")
            #замена на строку с -
        else:
            o.write(ln[0] + " " + (ln[1].replace(">", ":")) + "\n")
            #замена на строку без -
        prevStr = 1
    elif line.count("/") == 0: #поиск строк, не являющихся закрывающим
    тэгом
        prevStr = 0
        line = line.replace("<", "")
        line = line.replace(">", ":")
        if (line != lastTag): #Проверяем, не повторяется ли строка
            o.write(line)
```

```
    lastTag = line #Записываем в переменную найденную строку
p.close()
o.close()
```

**Код со счётом времени работы парсеров - написанного самостоятельно и библиотечного:**

```
import xmlplain
import time

tm = time.time()
loops = 10

for i in range(0, loops):
    p = open("lab3.xml", "r")
    o = open("lab3.yaml", "w")
    prevStr = 1
    lastTag = 0
    for line in p:
        if line.count("<") == 2:
            ln = line.split("<")
            if prevStr == 0:
                o.write(ln[0] + "- " + (ln[1].replace(">", ": ")) + "\n")
            else:
                o.write(ln[0] + " " + (ln[1].replace(">", ": ")) + "\n")
            prevStr = 1
        elif line.count("/") == 0:
            prevStr = 0
            line = line.replace("<", "")
            line = line.replace(">", ":")
            if (line != lastTag):
                o.write(line)
            lastTag = line
    p.close()
    o.close()

print(time.time() - tm)

for i in range(0, loops):

    with open("lab3.xml") as inf:
        root = xmlplain.xml_to_obj(inf, strip_space = True, fold_dict=True)
```

```
with open("lab3_2.yaml", "w") as outf:
    xmlplain.obj_to_yaml(root, outf)
```

```
print(time.time() - tm)
```

### **Результат работы программы:**

```
0.014003992080688477
0.07802057266235352
```

### **Содержание файла в исходном формате (XML):**

```
<day>
  <lesson>
    <time>10:00-11:30</time>
    <week>нечетная неделя</week>
    <subject>ДИСКРЕТНАЯ МАТЕМАТИКА</subject>
    <type>ЛЕК</type>
    <class>2202 АУД.</class>
    <place>ул.Ломоносова, д.9, лит. А</place>
    <teacher>Поляков Владимир Иванович</teacher>
    <format>Очно - дистанционный</format>
  </lesson>
  <lesson>
    <time>11:40-13:10</time>
    <subject>ИНОСТРАННЫЙ ЯЗЫК</subject>
    <type>ПРАК.</type>
    <place>ул.Ломоносова, д.9, лит. А</place>
    <format>Очно - дистанционный</format>
  </lesson>
</day>
```

### **Содержание файла в результирующем формате (YAML):**

```
day:
  lesson:
    - time: 10:00-11:30
      week: нечетная неделя
      subject: ДИСКРЕТНАЯ МАТЕМАТИКА
      type: ЛЕК
      class: 2202 АУД.
      place: ул.Ломоносова, д.9, лит. А
```

teacher: Поляков Владимир Иванович  
format: Очно - дистанционный  
- time: 11:40-13:10  
subject: ИНОСТРАННЫЙ ЯЗЫК  
type: ПРАК.  
place: ул.Ломоносова, д.9, лит. А  
format: Очно – дистанционный

### **Сравнение результатов:**

Парсер, написанный вручную, работает быстрее (примерно за  $10 \cdot O(n)$ ) библиотечного парсера из-за разницы реализованных алгоритмов и их скоростей выполнения. Данный библиотечный парсер медленнее примерно в 7 раз (точность приближительна, так как нагрузка на процессор каждый раз разная), к тому же написанный вручную парсер «заточен» под конкретную задачу, что так же влияет на скорость.

### **Вывод:**

В процессе выполнения лабораторной работы я изучила особенности различных языков разметки, научилась писать простой парсер и вспомнила основной синтаксис языка Python. Хотелось бы как-то пошутить, но все шутки на сегодня кончились.