Федеральное государственное автономное образовательное учереждение высшего образования “Национальный исследовательский университет ИТМО”

Мегафакультет компьютерных технологий и управления

Факультет программной инженерии и компьютерной техники



Лабораторная работа №4  
по информатике

Вариант: 11

Группа: P3114

Студент: Лагус

Максим Сергеевич

Преподаватель: Балакшин Павел Валерьевич

г. Санкт-Петербург

Сентябрь, 2021

Оглавление

[Текст задания: 3](#_Toc1725229023)

[Выполнение заданий: 4](#_Toc1298965261)

[Вывод: 9](#_Toc1396804251)

**Текст задания**

1. 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.

2. Изучить форму Бэкуса-Наура.

3. Изучить особенности протоколов и форматов обмена информацией

между системами: JSON, YAML, XML.

4. Понять устройство страницы с расписанием для своей группы:

https://itmo.ru/ru/schedule/0/P3110/schedule.htm

5. Исходя из структуры расписания конкретного дня, сформировать файл с

расписанием в формате, указанном в задании в качестве исходного.

6. 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.

7. Нельзя использовать готовые библиотеки, в том числе регулярные

выражения в Python и библиотеки для загрузки XML-файлов.

8. Дополнительное задание 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.

9. Дополнительное задание 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.

10. Дополнительное задание 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

десятикратное время выполнения парсинга + конвертации в

цикле.

b) Проанализировать полученные результаты и объяснить их

сходство/различие.

11. Дополнительное задание задание №4 (позволяет набрать +5 процентов

от максимального числа баллов БаРС за данную лабораторную.

a) Переписать исходную, 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

осоебнности использованного формата.

12. Проверить, что все пункты задания выполнены и выполнены верно.

13. Написать отчёт о проделанной работе.

14. Подготовиться к устным вопросам на защите.

**Выполнение работы**

from timeit import Timer  
  
import yaml  
import xmltodict  
import xmlplain  
import re  
  
file = "schedule.xml"  
output = "out"  
  
  
def task\_1():  
 with open(file, 'r') as f:  
 k = 0  
 with open(output, 'w') as o:  
 for a in f.readlines():  
 s = a.replace('\t', '')  
 if s.count('<') == 1:  
 if s.find("</") < 0:  
 o.write(' ' \* k + s.replace('<', '').replace('>', ": ") + '\n')  
 k += 1  
 else:  
 k -= 1  
 elif s.count('<') == 2:  
 o.write(' ' \* k + s[:s.find("</")].replace('<', '').replace('>', ": ") + '\n')  
 with open(output, 'r') as ya:  
 return yaml.safe\_load(ya.read())  
  
  
def task\_2():  
 with open(file, 'r') as f:  
 s = f.read().replace('\t', '').replace('\n', '')  
 with open(output, 'w') as ya:  
 y = xmlplain.xml\_to\_obj(s, strip\_space=True, fold\_dict=True)  
 xmlplain.obj\_to\_yaml(y, ya)  
 with open(output, 'r') as ya:  
 return yaml.safe\_load(ya.read())  
  
  
def task\_3():  
 with open(file, 'r') as f:  
 k = 0  
 match1 = r'/\w+'  
 match2 = r'\w+'  
 with open(output, 'w') as o:  
 for a in f.readlines():  
 s = a.replace('\t', '').replace('\n', '')  
 n = re.findall(match1, s)  
 if len(n) == 0:  
 o.write(' ' \* k + s.replace('<', '').replace('>', ": ") + '\n')  
 k += 1  
 else:  
 n = re.findall(match2, s)  
 if n == 1:  
 k -= 1  
 else:  
 o.write(' ' \* k + s[:s.find("</")].replace('<', '').replace('>', ": ") + '\n')  
 with open(output, 'r') as ya:  
 return yaml.safe\_load(ya.read())  
  
  
def task\_4():  
 print(Timer('from \_\_main\_\_ import task\_1; task\_1()').timeit(100))  
 print(Timer('from \_\_main\_\_ import task\_2; task\_2()').timeit(100))  
 print(Timer('from \_\_main\_\_ import task\_3; task\_3()').timeit(100))  
  
  
task\_1\_yaml = task\_1()  
print(task\_1\_yaml)  
task\_2\_yaml = task\_2()  
print(task\_2\_yaml)  
task\_3\_yaml = task\_3()  
print(task\_3\_yaml)  
task\_4()

**Вывод программы**

{'schedule': {'day': 'Среда', 'lessons': {'lesson': {'time': '13:30-15:00', 'week': 'чётная неделя', 'room': {'num': '285 ауд.', 'place': 'Кронверкский пр., д.49, лит.А'}, 'format': 'Очно - дистанционный', 'subject': 'Дискретная математика(Лек)', 'teacher': 'Поляков Владимир Иванович'}}}}

{'schedule': {'day': 'Среда', 'lessons': {'lesson': {'time': '13:30-15:00', 'week': 'чётная неделя', 'room': {'num': '285 ауд.', 'place': 'Кронверкский пр., д.49, лит.А'}, 'format': 'Очно - дистанционный', 'subject': 'Дискретная математика(Лек)', 'teacher': 'Поляков Владимир Иванович'}}}}

{'schedule': {'day': 'Среда', 'lessons': {'lesson': {'time': '13:30-15:00', 'week': 'чётная неделя', 'room': {'num': '285 ауд.', 'place': 'Кронверкский пр., д.49, лит.А', 'format': 'Очно - дистанционный', 'subject': 'Дискретная математика(Лек)', 'teacher': 'Поляков Владимир Иванович'}}}}}

0.20108190199971432

0.36511270600021817

0.22849370000039926

Содержимое файла schedule.xml

<schedule>  
 <day>Среда</day>  
 <lessons>  
 <lesson>  
 <time>13:30-15:00</time>  
 <week>чётная неделя</week>  
 <room>  
 <num>285 ауд.</num>  
 <place>Кронверкский пр., д.49, лит.А</place>  
 </room>  
 <format>Очно - дистанционный</format>  
 <subject>Дискретная математика(Лек)</subject>  
 <teacher>Поляков Владимир Иванович</teacher>  
 </lesson>  
 </lessons>  
</schedule>

Содержимое файла out после выполнения программы

schedule:   
 day: Среда  
 lessons:   
 lesson:   
 time: 13:30-15:00  
 week: чётная неделя  
 room:   
 num: 285 ауд.  
 place: Кронверкский пр., д.49, лит.А  
 format: Очно - дистанционный  
 subject: Дискретная математика(Лек)  
 teacher: Поляков Владимир Иванович

Теперь проанализируем различия в выполнении разных пунктов.

Видно, что пункт 2 состоит всего из 2 строк самого парсинга, за счёт использования уже готовой библиотеки по преобразованию.

Пункт 3 отличается от пункта 1 тем, что за счёт использования регулярных выражений, можно облегчить поиск нужных нам паттернов в исходной структуре.