НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО

Факультет программной инженерии и компьютерной техники

Информатика

Лабораторная работа № 4. " Исследование протоколов,форматов обмена информацией и языков разметкидокументов"

Выполнил студент

Хвостова Ирина Леонидовна

Группа № Р3124

Преподаватель: Болдырева Елена Александровна

г. Санкт-Петербург

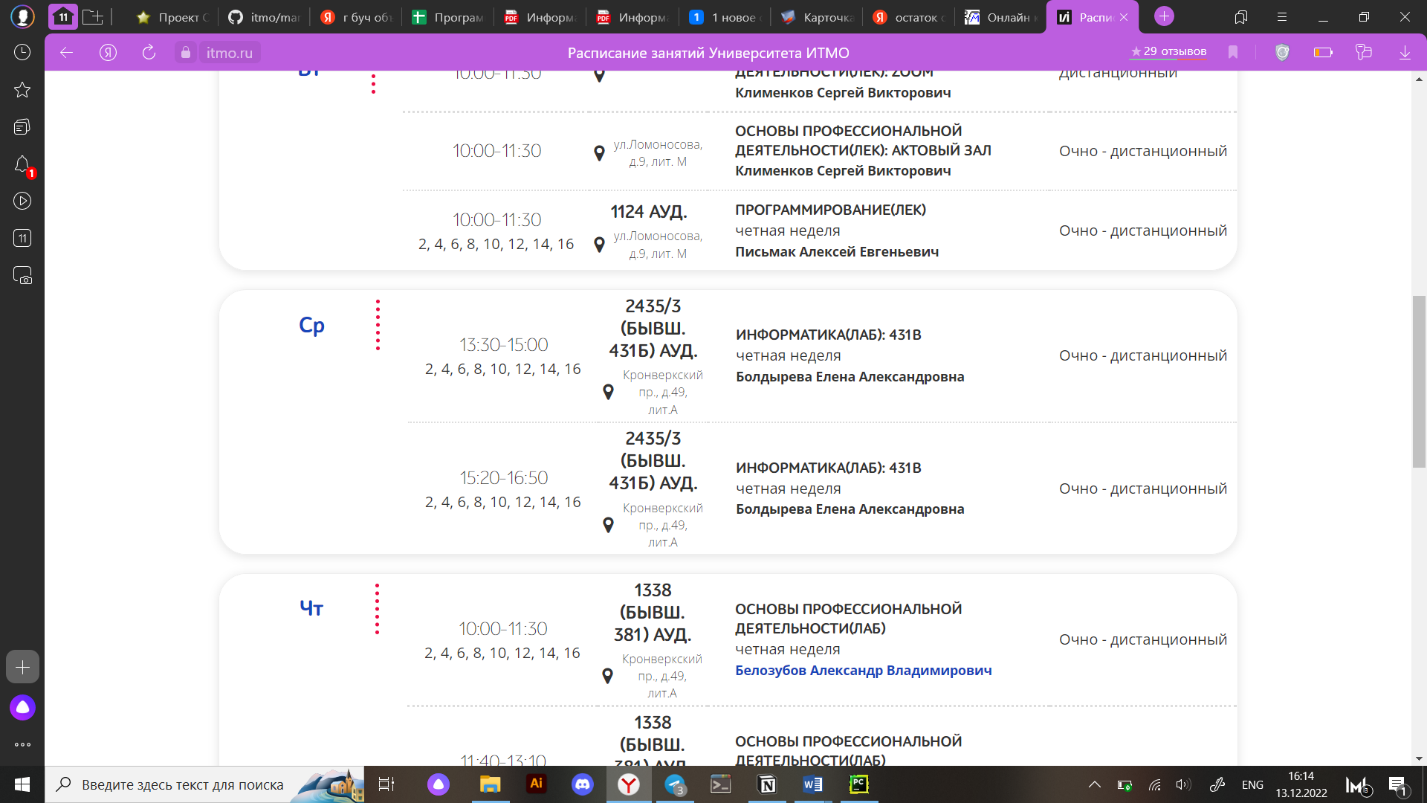
2022

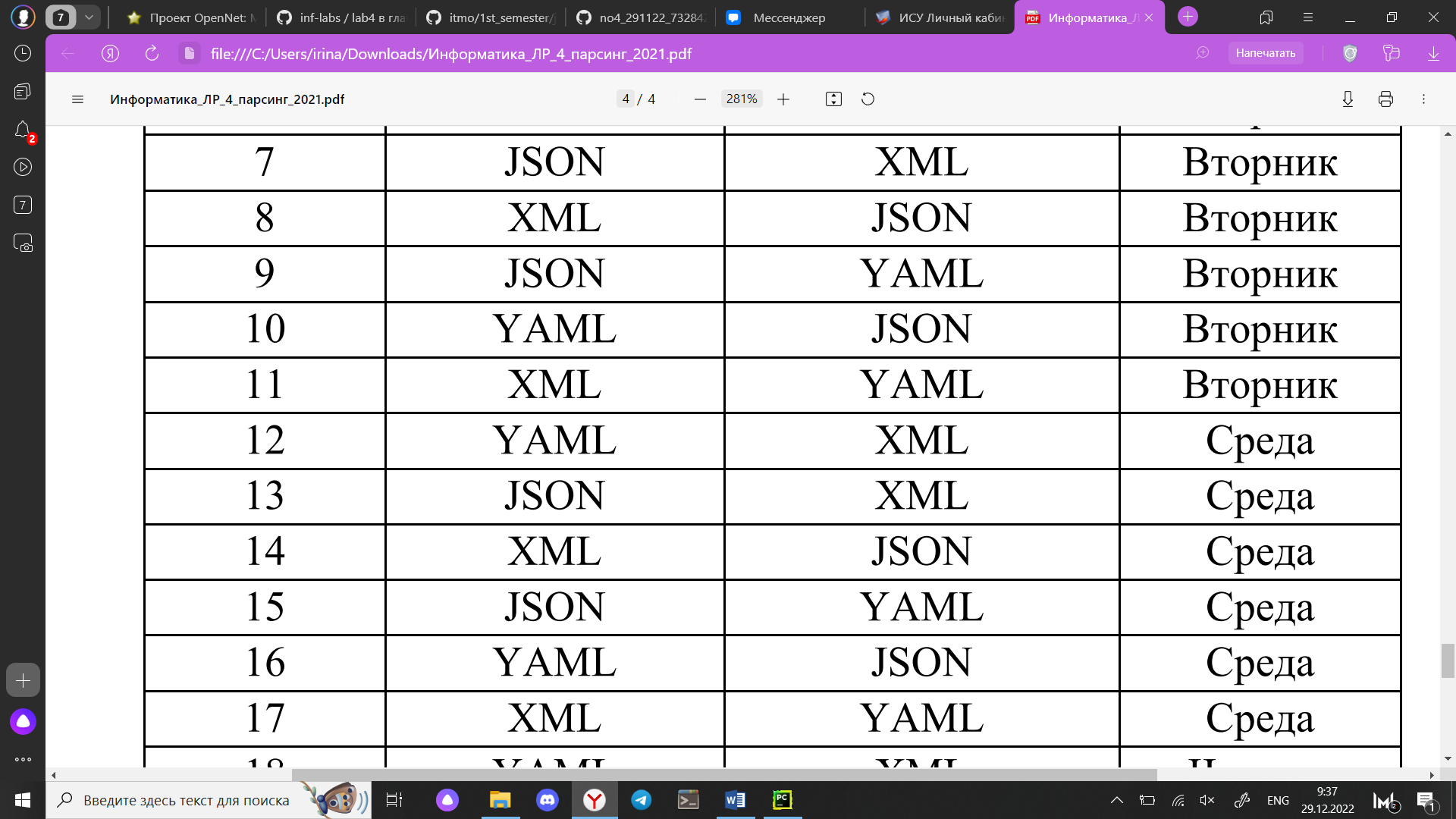
**Вариант: 17**

**Задание:**

1. **Задание на 60 баллов**

написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.

Расписание:



def parser(input):  
 while ('</' in input):  
 tr = False  
 start = 0  
 end = 0  
 for i in range(len(input)):  
 if tr and input[i] == '>':  
 end = i  
 tr = False  
 if tr:  
 continue  
 if (i < len(input) - 1):  
 if (input[i] + input[i + 1]) == '</':  
 tr = True  
 start = i  
  
 input = input[0:start] + input[end + 1:len(input)]  
 input = input.replace('<', '').replace('>', ": ")  
 return input  
  
  
with open('timetable.xml', 'r', encoding="utf-8") as f:  
 s = ''.join(f.readlines())  
  
res = parser(s)  
  
with open('output.yml', 'w', encoding="utf-8") as f:  
 f.write(res)

**xml**

<root>  
 <group>P3124</group>  
 <wednesday>  
 <lesson\_1>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <time>13:30-15:00</time>  
 <audience>2435/3 (БЫВШ. 431Б) АУД.</audience>  
 <place>Кронверкский пр., д.49, лит.А</place>  
 <weeks>2, 4, 6, 8, 10, 12, 14, 16</weeks>  
 <teacher>Болдырева Елена Александровна</teacher>  
 </lesson\_1>  
 <lesson\_4>  
 <name>ИНФОРМАТИКА(ЛАБ) - 431В</name>  
 <time>15:20-16:50</time>  
 <audience>2435/3 (БЫВШ. 431Б) АУД.</audience>  
 <place> Кронверкский пр., д.49, лит.А</place>  
 <weeks>2, 4, 6, 8, 10, 12, 14, 16</weeks>  
 <teacher>Болдырева Елена Александровна</teacher>  
 </lesson\_4>  
 </wednesday>  
</root>

**yml**

root:  
 group: P3124  
 wednesday:  
 lesson\_1:  
 name: ИНФОРМАТИКА(ЛАБ)  
 time: 13:30-15:00  
 audience: 2435/3 (БЫВШ. 431Б) АУД.  
 place: Кронверкский пр., д.49, лит.А  
 weeks: 2, 4, 6, 8, 10, 12, 14, 16  
 teacher: Болдырева Елена Александровна  
 lesson\_4:  
 name: ИНФОРМАТИКА(ЛАБ) - 431В  
 time: 15:20-16:50  
 audience: 2435/3 (БЫВШ. 431Б) АУД.  
 place: "\tКронверкский пр., д.49, лит.А"  
 weeks: 2, 4, 6, 8, 10, 12, 14, 16  
 teacher: Болдырева Елена Александровна

1. **Доп. задание №1 (+10 баллов)**

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.

import xmlplain  
  
with open("timetable.xml", encoding="utf-8") as f:  
 root = xmlplain.xml\_to\_obj(f, strip\_space=True, fold\_dict=True)  
  
with open("output\_with\_libs.yml", "w", encoding="utf-8") as outf:  
 xmlplain.obj\_to\_yaml(root, outf)

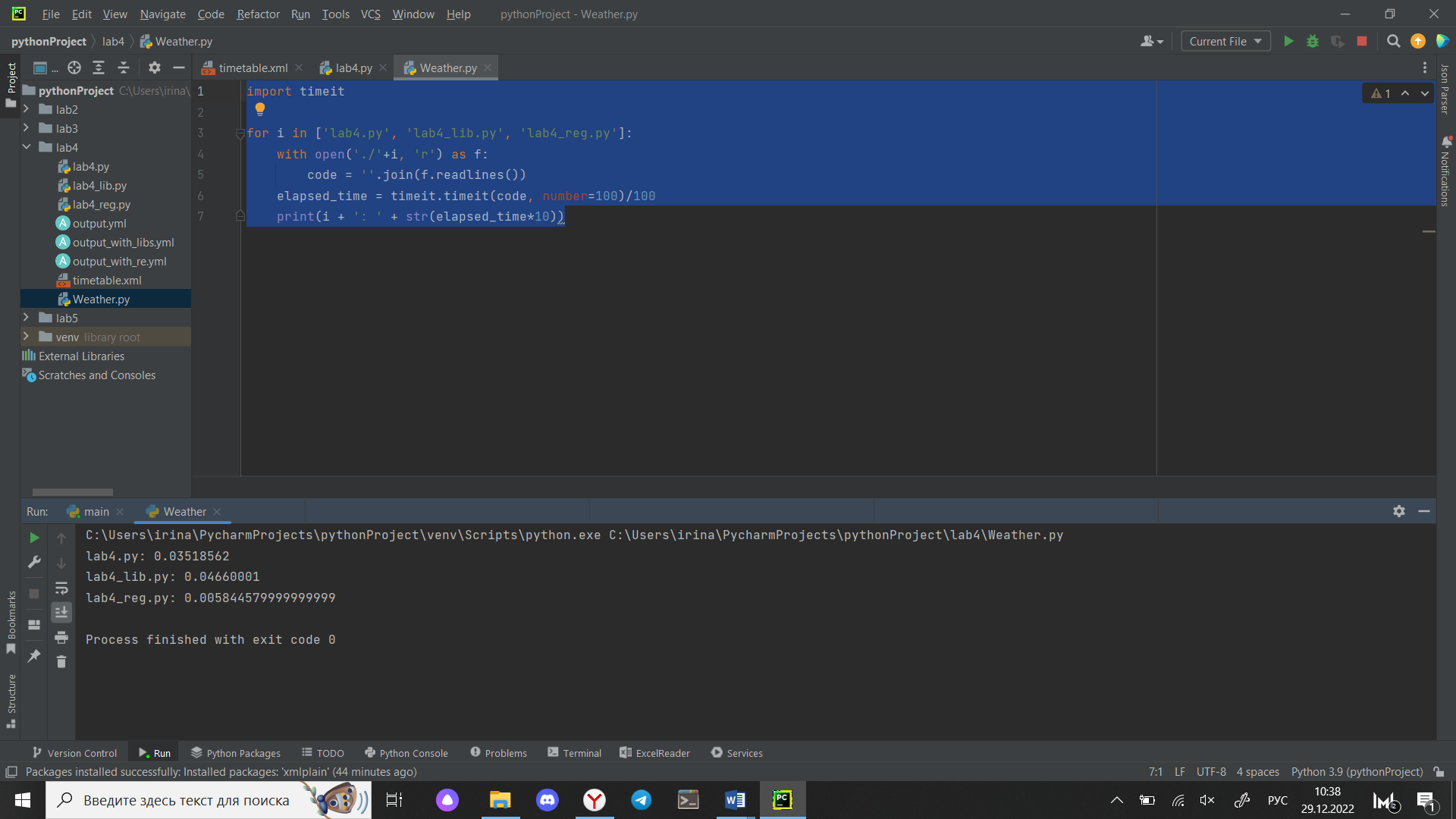
1. **Доп. задание №3 (+10 баллов)**

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.

import re  
  
def parser(input):  
 input = re.sub(r'<(\w+)>', r'\g<1>: ', input)  
 input = re.sub(r'</\w+>', r'', input)  
 return input  
  
with open('timetable.xml', 'r', encoding="utf-8") as f:  
 s = ''.join(f.readlines())  
  
res = parser(s)  
  
with open('output\_with\_re.yml', 'w', encoding="utf-8") as f:  
 f.write(res)

import timeit  
  
for i in ['lab4.py', 'lab4\_lib.py', 'lab4\_reg.py']:  
 with open('./'+i, 'r') as f:  
 code = ''.join(f.readlines())  
 elapsed\_time = timeit.timeit(code, number=100)/100  
 print(i + ': ' + str(elapsed\_time\*10))



**Вывод:**

Выполняя задания, я познакомилась языками разметки и научилась писать парсер на языке Python.

**Список литературы:**

1. Балакшин, П.В. Информатика / П.В. Балакшин, В.В. Соснин, Е.А. Машина. – СПб: Университет ИТМО, 2020. – 122 с.
2. https://docs.python.org/3/library/re.html