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: 4**

Выполнил:

студент 1 курса

Батманов Даниил Евгеньевич

Группа: Р3107

Принял:

Белозубов Александр Владимирович

Отчёт принят «\_\_»\_\_\_\_\_2022 г.

Оценка: \_\_\_\_\_\_\_\_\_\_\_

г. Санкт-Петербург, 2022

**Оглавление**

[Задание 3](#_Toc116942465)

[Основные этапы вычисления 4](#_Toc116942466)

[Заключение 36](#_Toc116942467)

[Список литературы 37](#_Toc116942469)

# 

# 

# 

# Задание

**Обязательное задание:**

Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый (yaml –> xml). Выполнить без использования библиотек и регулярных выражений.

**Дополнительное задание №1:**

Переписать исходный код с использованием библиотек и сравнить эти два кода.

**Необязательное задание №2:**

# Переписать исходный код с использованием регулярных выражений и сравнить эти два кода.

**Необязательное задание №3:**

Используя свою исходную программу из обязательного задания, программу из дополнительного задания No1 и программу из дополнительного задания No2, сравнить стократное время выполнения парсинга + конвертации в цикле. Проанализировать полученные результаты и объяснить их сходство/различие.

**Необязательное задание №4:**

Переписать исходную программу, чтобы она осуществляла парсинг и конвертацию исходного файла в любой другой формат (кроме JSON, YAML, XML, HTML): PROTOBUF, TSV, CSV, WML и т.п. Проанализировать полученные результаты, объяснить особенности использования формата.

# Основные этапы вычисления

**Обязательное задание:**

Исходный файл (yaml):

schedule:  
 doublet1:  
 time:  
 start:  
 hour: XI  
 minutes: XL  
 end:  
 hour: XIII  
 minutes: X  
 weeks: [II, IV, VI, VIII, X, XII, XIV, XVI]  
 place:  
 address: |  
 Кронверкский пр., д.49, лит.А  
 room: |  
 2306/1 (БЫВШ. 304) АУД.  
 name: |  
 ИНФОРМАТИКА(ЛАБ)  
 week: |  
 чётная неделя  
 teacher: |  
 Белозубов Александр Владимирович  
 kind: |  
 Очно-дистанционный  
 doublet2:  
 time:  
 start:  
 hour: XI  
 minutes: XL  
 end:  
 hour: XIII  
 minutes: X  
 weeks: [III, V, VII, IX, XI, XIII, XV, XVII]  
 place:  
 address: |  
 Кронверкский пр., д.49, лит.А  
 room: |  
 2306/1 (БЫВШ. 304) АУД.  
 name: |  
 ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)  
 week: |  
 нечётная неделя  
 teacher: |  
 Ткешелашвили Нино Мерабиевна  
 kind: |  
 Очно-дистанционный  
 doublet3:  
 time:  
 start:  
 hour: XIII  
 minutes: XXX  
 end:  
 hour: XV  
 minutes: I  
 weeks: [II, IV, VI, VIII, X, XII, XIV, XVI]  
 place:  
 address: |  
 Кронверкский пр., д.49, лит.А  
 room: |  
 2306/1 (БЫВШ. 304) АУД.  
 name: |  
 ИНФОРМАТИКА(ЛАБ)  
 week: |  
 чётная неделя  
 teacher: |  
 Белозубов Александр Владимирович  
 form: |  
 Очно-дистанционный  
 doublet4:  
 time:  
 start:  
 hour: XIII  
 minutes: XXX  
 end:  
 hour: XV  
 minutes: I  
 weeks: [III, V, VII, IX, XI, XIII, XV, XVII]  
 place:  
 address: |  
 Кронверкский пр., д.49, лит.А  
 room: |  
 2306/1 (БЫВШ. 304) АУД.  
 name: |  
 ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)  
 week: |  
 нечётная неделя  
 teacher: |  
 Ткешелашвили Нино Мерабиевна  
 kind: |  
 Очно-дистанционный

*Исходный код:*

# lab4, var4, yaml –> xml  
schedule = open('schedule.yaml')  
openedSchedule = schedule.readlines()  
openedResFile = open('resFile.xml', 'w')  
s = ""  
mas = [len(openedSchedule)]  
for l in range(0, len(openedSchedule)):  
 max = 0  
 k = 0  
 for j in range(0, len(openedSchedule[l])):  
 h = openedSchedule[l]  
 if h[j] == " ":  
 k += 1  
 else:  
 if max < k:  
 max = k  
 mas.append(max + max % 2)  
 k = 0  
mas[0] = 0  
openedResFile.write("<?xml version=" + '"' + "1.0" + '"' + "?>" + "\n" + "<result>" + "\n")  
arr = []  
for i in range(0, len(openedSchedule)):  
 part = openedSchedule[i]  
 if (mas[i]) // 2 <= (len(arr) - 1):  
 openedResFile.write(' ' \* mas[i] + "</" + arr.pop(-1) + '>' + "\n")  
 if part[-2] == ":":  
 arr.append(part.strip()[:-1])  
 openedResFile.write(' ' \* (part.count(' ') + 1) + '<' + part[:-2].strip() + '>' + '\n')  
 elif part.find('[') != -1:  
 for j in range(0, len(part)):  
 if part[j] != ":":  
 s += part[j]  
 else:  
 v = s  
 s = ""  
 break  
 for j in range(part.find('[') + 1, part.find(']') + 1):  
 if (part[j] != ",") and (part[j] != " ") and (part[j] != ']'):  
 s += part[j]  
 else:  
 if s != "":  
 openedResFile.write(' ' \* (v.count(' ') + 1) + '<' + v.strip() + '>' + s + '<' + '/' + v.strip() + '>' + '\n')  
 s = ""  
 elif (part[-2] != ":") and (part.find(':') != -1) and (part.find('|') == -1):  
 for j in range(0, len(part)):  
 if part[j] != ":":  
 s += part[j]  
 else:  
 v = s  
 openedResFile.write(' ' \* (part.count(' ')) + '<' + v.strip() + '>' + part[(part.find(':') + 2):][:-1] + '<' + '/' + v.strip() + '>' + '\n')  
 s = ""  
 break  
 elif (part[-2] != ":") and (part.find(':') != -1) and (part.find('|') != -1):  
 mx = 0  
 k = 0  
 for j in range(0, len(openedSchedule[i + 1])):  
 h = openedSchedule[i + 1]  
 if h[j] == " ":  
 k += 1  
 else:  
 if mx < k:  
 mx = k  
 k = 0  
 for j in range(0, len(part)):  
 if part[j] != ":":  
 s += part[j]  
 else:  
 v = s  
 openedResFile.write(' ' \* (part.count(' ') + 1) + '<' + v.replace(' ', '') + '>' + openedSchedule[i + 1][mx:][:-1] + '<' + '/' + v.replace(' ', '') + '>' + '\n')  
 s = ""  
 break  
for i in range((len(arr) - 1), -1, -1):  
 openedResFile.write(' ' \* (i + 1) + "</" + arr.pop(-1) + ">" + "\n")  
openedResFile.write("</result>")  
openedResFile.close()

*Результат выполнения кода:*

<?xml version="1.0"?>  
<result>  
 <schedule>  
 <doublet1>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet1>  
 <doublet2>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet2>  
 <doublet3>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <form>Очно-дистанционный</form>  
 </doublet3>  
 <doublet4>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционны</kind>  
 </doublet4>  
 </schedule>  
</result>

**Необязательное задание №1:**

Исходный код:

# lab4, var4, yaml –> xml  
import xmlplain  
with open("schedule.yaml") as inf:  
 root = xmlplain.obj\_from\_yaml(inf)  
with open("resFile.xml", 'w') as outf:  
 xmlplain.xml\_from\_obj(root, outf, pretty=True)

*Результат выполнения кода:*

<?xml version="1.0"?>  
<result>  
 <schedule>  
 <doublet1>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet1>  
 <doublet2>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet2>  
 <doublet3>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <form>Очно-дистанционный</form>  
 </doublet3>  
 <doublet4>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционны</kind>  
 </doublet4>  
 </schedule>  
</result>

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

**Необязательное задание №2:**

*Исходный код:*

# lab4, var4, yaml –> xml  
import re  
schedule = open('schedule.yaml')  
openedSchedule = schedule.readlines()  
openedResFile = open('resFile.xml', 'w')  
mas = [len(openedSchedule)]  
for l in range(1, len(openedSchedule)):  
 mas.append(re.search(r"^\s+", openedSchedule[l]).end() + re.search(r"^\s+", openedSchedule[l]).end() % 2)  
mas[0] = 0  
openedResFile.write("<?xml version=" + '"' + "1.0" + '"' + "?>" + '\n' + "<result>" + '\n')  
arr = []  
for i in range(0, len(openedSchedule)):  
 part = openedSchedule[i]  
 if (mas[i]) // 2 <= (len(arr) - 1):  
 openedResFile.write(' ' \* mas[i] + "</" + arr.pop(-1) + '>' + "\n")  
 if re.search(r".+:\n", part) != None:  
 arr.append(part.strip()[:-1])  
 openedResFile.write(' ' \* (part.count(' ') + 1) + '<' + re.search(r"\w+", part).group() + '>' + '\n')  
 elif re.search(r"\[", part) != None:  
 s = ""  
 for j in re.search(r"\[[\w+, ]+\]", part).group():  
 if j != "[" and j != "]" and j != " " and j != ",":  
 s += j  
 else:  
 if s != "":  
 openedResFile.write(re.search(r"^\s+", part).group() + '<' + re.search(r"\w+", part).group() + '>' + s + '<' + '/' + re.search(r"\w+", part).group() + '>' + '\n')  
 s = ""  
 elif re.search(r"\w+: \w+", part) != None:  
 openedResFile.write(re.search(r"^\s+", part).group() + '<' + re.search(r"\w+", part).group() + '>' + part[(part.find(':') + 2):][:-1] + '<' + '/' + re.search(r"\w+", part).group() + '>' + '\n')  
 elif re.search(r"\w+: \|", part) != None:  
 openedResFile.write(re.search(r"^\s+", part).group() + '<' + re.search(r"\w+", part).group() + '>' + openedSchedule[i + 1][:-1][re.search(r"^\s+", openedSchedule[i + 1]).end():] + '</' + re.search(r"\w+", part).group() + '>' + '\n')  
for i in range((len(arr) - 1), -1, -1):  
 openedResFile.write(' ' \* (i + 1) + "</" + arr.pop(-1) + ">" + "\n")  
openedResFile.write("</result>")  
openedResFile.close()

*Результат выполнения программы:*

<?xml version="1.0"?>  
<result>  
 <schedule>  
 <doublet1>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet1>  
 <doublet2>  
 <time>  
 <start>  
 <hour>XI</hour>  
 <minutes>XL</minutes>  
 </start>  
 <end>  
 <hour>XIII</hour>  
 <minutes>X</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционный</kind>  
 </doublet2>  
 <doublet3>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>II</weeks>  
 <weeks>IV</weeks>  
 <weeks>VI</weeks>  
 <weeks>VIII</weeks>  
 <weeks>X</weeks>  
 <weeks>XII</weeks>  
 <weeks>XIV</weeks>  
 <weeks>XVI</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ИНФОРМАТИКА(ЛАБ)</name>  
 <week>чётная неделя</week>  
 <teacher>Белозубов Александр Владимирович</teacher>  
 <form>Очно-дистанционный</form>  
 </doublet3>  
 <doublet4>  
 <time>  
 <start>  
 <hour>XIII</hour>  
 <minutes>XXX</minutes>  
 </start>  
 <end>  
 <hour>XV</hour>  
 <minutes>I</minutes>  
 </end>  
 <weeks>III</weeks>  
 <weeks>V</weeks>  
 <weeks>VII</weeks>  
 <weeks>IX</weeks>  
 <weeks>XI</weeks>  
 <weeks>XIII</weeks>  
 <weeks>XV</weeks>  
 <weeks>XVII</weeks>  
 </time>  
 <place>  
 <address>Кронверкский пр., д.49, лит.А</address>  
 <room>2306/1 (БЫВШ. 304) АУД.</room>  
 </place>  
 <name>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</name>  
 <week>нечётная неделя</week>  
 <teacher>Ткешелашвили Нино Мерабиевна</teacher>  
 <kind>Очно-дистанционны</kind>  
 </doublet4>  
 </schedule>  
</result>

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

**Необязательное задание №3:**

При помощи библиотеки timeit сравним стократное время выполнения парсинга + конвертации в цикле.

Результаты замеров:

* Код без библиотек и регулярных выражений: 0.0009987295299652033 с.;
* Код с библиотекой: 0.007624935949861538 с.;
* Код с регулярными выражениями: 0.0009934758594534565 с.

Таким образом, можно сделать вывод, что код написаны с использованием регулярных выражений и без использования регулярных выражений выполняется приблизительно за одинаковое время, но при этом, написать код с регулярными выражениями быстрее, чем без них. Самым медлительным оказался код с использованием библиотеки, что неудивительно.

# Заключение

В ходе выполнения данной лабораторной работы я научился работать с библиотеками языка программирования Python: xmlplain и timeit. Более того, познакомился с языками разметки yaml, xml, json. Реализовал программный продукт, который выполняет конвертацию из yaml в xml. Научился открывать и записывать данные на Python.