Федеральное государственное автономное образовательное учреждение высшего образования **«Национальный исследовательский университет ИТМО»**

Факультет Программной Инженерии и Компьютерной Техники

**Лабораторная работа по информатике №4**

вариант: 4

Преподаватель: Машина Екатерина Алексеевна

Выполнил: Онишков Валерий Вячеславович

Группа: Р3114

Санкт-Петербург, 2022г

Оглавление

[Задание 2](#_Toc86705323)

[Ход работы 3](#_Toc86705324)

[Обязательное задание 3](#_Toc86705325)

[Дополнительное задание №1 5](#_Toc86705326)

[Дополнительное задание №2 6](#_Toc86705327)

[Дополнительное задание №3 8](#_Toc86705328)

[Дополнительное задание №4 8](#_Toc86705329)

[Вывод 10](#_Toc86705330)

[Список литературы. 10](#_Toc86705331)

# Задание

* 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.
* Дополнительное задание задание №1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + 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.
* Дополнительное задание задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + Переписать исходный код, добавив в него использование регулярных выражений.
  + Сравнить полученные результаты и объяснить их сходство/различие.
* Дополнительное задание задание №3 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + 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.
  + Проанализировать полученные результаты и объяснить их сходство/различие.
* Дополнительное задание задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную.
  + 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.
  + Проанализировать полученные результаты, объяснить особенности использованного формата.

# Ход работы

Исходный файл:

timetable:

day: Пн

lesson1:

time:13:30-15:00

room: 2337

subject: Математика(продвинутый уровень)

teacher: Абжандадзе Заур Леванович

location: Кронверкский пр., д.49, лит.А

format: Очно-дистанционный

weeks: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17

lesson2:

time: 15:20-16:50

room: 2337

subject: Математика(продвинутый уровень)

teacher: Абжандадзе Заур Леванович

location: Кронверкский пр., д.49, лит.А

format: Очно-дистанционный

weeks: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17

lesson3:

time: 18:40-20:10

room: 1419

subject: Дискретная математика (продивинутый уровень)

teacher: Пастор Алексей Владимирович

location: Кронверкский пр., д.49, лит.А

format: Очно-дистанционный

weeks: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17

lesson4:

time: 20:20-21:50

room: 1419

subject: Дискретная математика (продивинутый уровень)

teacher: Пастор Алексей Владимирович

location: Кронверкский пр., д.49, лит.А

format: Очно-дистанционный

weeks: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17

# Обязательное задание

import time

f = open("Исходник.txt", encoding='utf-8')

d = f

def file\_to\_xml(file):

xml\_file\_start = ""

xml\_file\_end = ""

xml\_file = ''

timetablename = file.readline()[:-2]

day = file.readline()[:-1].strip()

day = day.split(':')

txt = file.read()

ms1 = txt.split('\n')

for i in range(len(ms1) - 1):

ms1[i] = ms1[i].strip()

for i in range(len(ms1) - 1):

if ms1[i] == '':

ms1.remove('')

ms1[i] = ms1[i].replace(':', '/', 1)

ms1[i] = ms1[i].split('/')

for i in range(len(ms1) - 1):

if 'lesson' in ms1[i][0]:

xml\_file += xml\_file\_start + xml\_file\_end

xml\_file\_start = ' '

xml\_file\_end = ' '

xml\_file\_start += f"<{ms1[i][0].strip()}>\n"

xml\_file\_end = f" </{ms1[i][0].strip()}>\n" + xml\_file\_end

else:

xml\_file\_start += f" <{ms1[i][0]}>{ms1[i][1].strip()}</{ms1[i][0]}>\n"

xml\_file = f"<{timetablename}>\n" + f" <{day[0]}>{day[1].strip()}</{day[0]}>\n" + xml\_file + xml\_file\_start + xml\_file\_end + f"</{timetablename}>\n"

return xml\_file

print(file\_to\_xml(f))

start\_time = time.time()

for n in range(100):

f = open("Исходник.txt", encoding='utf-8')

file\_to\_xml(f)

print("--- %s seconds ---" % (time.time() - start\_time))

**Результат:**

<timetable>

<day> Пн</day>

<lesson1>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 2337</room>

<subject> Математика(продвинутый уровень)</subject>

<teacher> Абжандадзе Заур Леванович</teacher>

<time>13:30-15:00</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson1>

<lesson2>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 2337</room>

<subject> Математика(продвинутый уровень)</subject>

<teacher> Абжандадзе Заур Леванович</teacher>

<time> 15:20-16:50</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson2>

<lesson3>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 1419</room>

<subject> Дискретная математика (продивинутый уровень)</subject>

<teacher> Пастор Алексей Владимирович</teacher>

<time> 18:40-20:10</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson3>

<lesson4>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 1419</room>

<subject> Дискретная математика (продивинутый уровень)</subject>

<teacher> Пастор Алексей Владимирович</teacher>

<time> 20:20-21:50</time>

</lesson4>

</timetable>

# Дополнительное задание №1

import dict2xml

import time

f = open("Исходник.txt", encoding='utf-8')

def file\_to\_dict(file):

dct = {}

timetablename = file.readline()[:-2]

day = file.readline()[:-1].strip()

day = day.split(':')

txt = file.read()

array = txt.split('\n')

dct[timetablename] = {}

dct[timetablename][day[0]] = day[1]

for i in range(len(array) - 1):

array[i] = array[i].strip()

for i in range(len(array) - 1):

if array[i] == '':

array.remove('')

array[i] = array[i].replace(':', '/', 1)

array[i] = array[i].split('/')

lesson\_number = ''

for i in range(len(array) - 1):

if 'lesson' in array[i][0]:

lesson\_number = array[i][0]

dct[timetablename][array[i][0]] = {}

else:

dct[timetablename][lesson\_number][array[i][0]] = array[i][1]

return dct

print(dict2xml.dict2xml(file\_to\_dict(f)))

start\_time = time.time()

for n in range(100):

f = open("Исходник.txt", encoding='utf-8')

file\_to\_dict(f)

print("--- %s seconds ---" %(time.time() - start\_time))

**Результат:**

# <timetable>

# <day> Пн</day>

# <lesson1>

# <format> Очно-дистанционный</format>

# <location> Кронверкский пр., д.49, лит.А</location>

# <room> 2337</room>

# <subject> Математика(продвинутый уровень)</subject>

# <teacher> Абжандадзе Заур Леванович</teacher>

# <time>13:30-15:00</time>

# <weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

# </lesson1>

# <lesson2>

# <format> Очно-дистанционный</format>

# <location> Кронверкский пр., д.49, лит.А</location>

# <room> 2337</room>

# <subject> Математика(продвинутый уровень)</subject>

# <teacher> Абжандадзе Заур Леванович</teacher>

# <time> 15:20-16:50</time>

# <weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

# </lesson2>

# <lesson3>

# <format> Очно-дистанционный</format>

# <location> Кронверкский пр., д.49, лит.А</location>

# <room> 1419</room>

# <subject> Дискретная математика (продивинутый уровень)</subject>

# <teacher> Пастор Алексей Владимирович</teacher>

# <time> 18:40-20:10</time>

# <weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

# </lesson3>

# <lesson4>

# <format> Очно-дистанционный</format>

# <location> Кронверкский пр., д.49, лит.А</location>

# <room> 1419</room>

# <subject> Дискретная математика (продивинутый уровень)</subject>

# <teacher> Пастор Алексей Владимирович</teacher>

# <time> 20:20-21:50</time>

# </lesson4>

# </timetable>

# --- 0.0965576171875 seconds ---

# Дополнительное задание №2

import re

import time

f = open("Исходник.txt", encoding='utf-8')

def file\_to\_xml(file):

xml\_file\_start = ""

xml\_file\_end = ""

xml\_file = ''

timetablename = file.readline()[:-2]

day = file.readline()[:-1].strip()

day = day.split(':')

txt = file.read()

ms1 = re.split('\n', txt)

for i in range(len(ms1) - 1):

ms1[i] = ms1[i].strip()

for i in range(len(ms1) - 1):

if ms1[i] == '':

ms1.remove('')

ms1[i] = re.sub(':', '/',ms1[i], 1)

ms1[i] = ms1[i].split('/')

for i in range(len(ms1) - 1):

if re.fullmatch(r"lesson\d+",ms1[i][0]):

xml\_file += xml\_file\_start + xml\_file\_end

xml\_file\_start = ''

xml\_file\_end = ' '

xml\_file\_start += f"<{ms1[i][0].strip()}>\n"

xml\_file\_end = f" </{ms1[i][0].strip()}>\n" + xml\_file\_end

else:

xml\_file\_start += f" <{ms1[i][0]}>{ms1[i][1].strip()}</{ms1[i][0]}>\n"

xml\_file = f"<{timetablename}>\n"+ f"<{day[0]}>{day[1].strip()}</{day[0]}>\n" + xml\_file +xml\_file\_start + xml\_file\_end + f"</{timetablename}>\n"

return xml\_file

print(file\_to\_xml(f))

start\_time = time.time()

for n in range(100):

f = open("Исходник.txt", encoding='utf-8')

file\_to\_xml(f)

print("--- %s seconds ---" %(time.time() - start\_time))

**Результат:**

<timetable>

<day> Пн</day>

<lesson1>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 2337</room>

<subject> Математика(продвинутый уровень)</subject>

<teacher> Абжандадзе Заур Леванович</teacher>

<time>13:30-15:00</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson1>

<lesson2>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 2337</room>

<subject> Математика(продвинутый уровень)</subject>

<teacher> Абжандадзе Заур Леванович</teacher>

<time> 15:20-16:50</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson2>

<lesson3>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 1419</room>

<subject> Дискретная математика (продивинутый уровень)</subject>

<teacher> Пастор Алексей Владимирович</teacher>

<time> 18:40-20:10</time>

<weeks> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17</weeks>

</lesson3>

<lesson4>

<format> Очно-дистанционный</format>

<location> Кронверкский пр., д.49, лит.А</location>

<room> 1419</room>

<subject> Дискретная математика (продивинутый уровень)</subject>

<teacher> Пастор Алексей Владимирович</teacher>

<time> 20:20-21:50</time>

</lesson4>

</timetable>

--- 0.014565706253051758 seconds ---

# Дополнительное задание №3

1) Время работы программы для обязательного задания:

**0.016557931900024414**

2) Время работы программы для доп. задания №1 (программа использует сторонние библиотека dict2xml):

**0.017239809036254883**

3) Время работы программы для доп. задания №2 (программа использует библиотеку для регулярных выражений):

**0.011173486709594727**

Возможное объяснение результатов:

Дольше всего работает программа (2), тк она производит полноценный рекурсивный парсинг YAML, в то время как две остальные программы выбирают только нужные свойства, и парсинг производится в один проход

Прогамма (3) на регулярных выражениях работает дольше программы (1) так как регулярные выражения более затратны из-за проверки исходной строки на соответствие шаблону, а так же из-за использования ленивых квантификаторов. Но так как размер файла невелик, разница несущественная.

# Дополнительное задание №4

f = open("Исходник.txt", encoding='utf-8')

def file\_to\_massive(file):

xml\_file\_start = ""

xml\_file\_end = ""

xml\_file = ''

TSV\_file = [['day', 'lesson']]

timetablename = file.readline()[:-2]

day = file.readline()[:-1].strip()

day = day.split(':')

txt = file.read()

ms1 = txt.split('\n')

for i in range(len(ms1) - 1):

ms1[i] = ms1[i].strip()

for i in range(len(ms1) - 1):

if ms1[i] == '':

ms1.remove('')

ms1[i] = ms1[i].replace(':', '/', 1)

ms1[i] = ms1[i].split('/')

dct\_number = {

'day': 0,

'lesson': 1

}

lessonnumber = 0

for i in ms1:

if 'lesson' in i[0]:

lessonnumber = int(i[0][-1])

TSV\_file.append([day[1]])

TSV\_file[lessonnumber].append(lessonnumber)

dct\_number['lesson'] = int(i[0][-1])

else:

if i[0] in TSV\_file[0]:

TSV\_file[lessonnumber].append(i[1])

else:

dct\_number[i[0]] = len(TSV\_file[0])

TSV\_file[0].append(i[0])

TSV\_file[lessonnumber].append(i[1])

return TSV\_file

massive = file\_to\_massive(f)

for i in massive:

for j in i:

print(j, end='\t')

print('\n')

**Результат в :**

**day lesson time room subject teacher location format weeks**

**Пн 1 13:30-15:00 2337 Математика(продвинутый уровень) Абжандадзе Заур Леванович Кронверкский пр., д.49, лит.А Очно-дистанционный 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17**

**Пн 2 15:20-16:50 2337 Математика(продвинутый уровень) Абжандадзе Заур Леванович Кронверкский пр., д.49, лит.А Очно-дистанционный 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17**

**Пн 3 18:40-20:10 1419 Дискретная математика (продивинутый уровень) Пастор Алексей Владимирович Кронверкский пр., д.49, лит.А Очно-дистанционный 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 17**

**Пн 4 20:20-21:50 1419 Дискретная математика (продивинутый уровень) Пастор Алексей Владимирович Кронверкский пр., д.49, лит.А Очно-дистанционный**

**--- 1.2 seconds ---**



Рисунок 1 – Создатель регулярных выражений Кен Томпсон

# Вывод

Во время выполнения лабораторной работы я узнал о языках разметки XML и YAML, научился с ними работать и переводить один тип в другой с помощью средств языка Python. Научился работать с библиотекой time для измерения времени, и некоторыми библиотеками для автоматического парсинга.

# Список литературы.

1. Балакшин П.В., Соснин В.В., Калинин И.В., Малышева Т.А., Раков С.В., Рущенко Н.Г., Дергачев А.М. Информатика: лабораторные работы и тесты: Учебно-методическое пособие / Рецензент: Поляков В.И. - Санкт-Петербург: Университет ИТМО, 2019. - 56 с. - экз. - Режим доступа:

<https://books.ifmo.ru/book/2248/informatika:_laboratornye_raboty_i_testy:_uchebno-metodicheskoe_posobie_/_recenzent:_polyakov_v.i..htm>

1. Грошев А.С. Г89 Информатика: Учебник для вузов / А.С. Грошев. – Архангельск, Арханг. гос. техн. ун-т, 2010. -470с. -Режим доступа <https://narfu.ru/university/library/books/0690.pdf>