Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной Инженерии и Компьютерной Техники

Лабораторная работа №4

Исследование протоколов, форматов обмена информацией и языков разметки документов

Вариант 17

Выполнил:

Некрутенко Максим Владимирович

Группа P3106

Проверил:

Доцент ПИиКТ, кандидат технических наук

Балакшин Павел Валерьевич

Санкт-Петербург 2024

Оглавление

[Задание. 3](#__RefHeading___Toc1130_3594372212)

[Основные этапы вычислений. 5](#__RefHeading___Toc1132_3594372212)

[1. Обязательное задание 6](#__RefHeading___Toc1134_3594372212)

[2. Дополнительное задание №1 7](#__RefHeading___Toc1136_3594372212)

[3. Дополнительное задание №2 9](#__RefHeading___Toc1138_3594372212)

[4. Дополнительное задание №3 11](#__RefHeading___Toc1140_3594372212)

[5. Дополнительное задание №4 13](#__RefHeading___Toc1142_3594372212)

[6. Дополнительное задание №5 15](#__RefHeading___Toc1144_3594372212)

[Вывод: 17](#__RefHeading___Toc1146_3594372212)

[Источники: 18](#__RefHeading___Toc1148_3594372212)

# Задание.

1. Определить номер варианта как остаток деления на 36 последних двух цифр своего идентификационного номера в ISU: например, 125598 / 36 = 26. В случае, если в оба указнных дня недели нет занятий, то увеличить номер варианта на восемь. В случае, если занятий нет и в новом наборе дней, то продолжать увеличивать на восемь.

2. Изучить форму Бэкуса-Наура.

3. Изучить основные принципы организации формальных грамматик.

4. Изучить особенности языков разметки/форматов JSON, YAML, XML.

5. Понять устройство страницы с расписанием на примере расписания лектора: <https://itmo.ru/ru/schedule/3/125598/raspisanie_zanyatiy.htm>

6. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы хотя бы в одной из выбранных дней было не менее двух занятий (можно использовать своё персональное). В случае, если в данный день недели нет таких занятий, то увеличить номер варианта ещё на восемь.

7. Обязательное задание (позволяет набрать до 45 процентов от максимального числа баллов БаРС за данную лабораторную): написать программу на языке Python 3.x или любом другом, которая бы осуществляла парсинг и конвертацию исходного файла в новый путём простой замены метасимволов исходного формата на метасимволы результирующего формата.

8. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

9. Дополнительное 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. 1 c) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

10.Дополнительное 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. Объяснение должно быть отражено в отчёте.

11. Дополнительное задание №3 (позволяет набрать +25 процентов от максимального числа баллов БаРС за данную лабораторную). а) Переписать исходный код таким образом, чтобы для решения задачи использовались формальные грамматики. То есть ваш код должен уметь осуществлять парсинг и конвертацию любых данных, представленных в исходном формате, в данные, представленные в результирующем формате: как с готовыми библиотеками из дополнительного задания №1. b) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом. с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

12. Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную). a) Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле. b) Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

13. Дополнительное задание №5 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную). a) 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. 2 b) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

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

15.Написать отчёт о проделанной работе.

16.Подготовиться к устным вопросам на защите.

# Основные этапы вычислений.

Исходный файл:

<root>

<nedeli>

<weeks>[2, 4, 6, 8, 10, 12, 14, 16]</weeks>

</nedeli>

<pair1>

<day>Среда</day>

<first>

<start>8:20</start>

<end>9:50</end>

<group>P3106</group>

<audience>1216</audience>

<adress>ул. Ломоносова, д. 9, лит. М</adress>

<name>Информатика</name>

<type>Лекция</type>

<teacher>Балакшин Павел Валерьевич</teacher>

<format>Очно</format>

</first>

<second>

<start>10:00</start>

<end>11:30</end>

<group>P3106</group>

<audience>1216</audience>

<adress>ул. Ломоносова, д. 9, лит. М</adress>

<name>Основы профессиональной деятельности</name>

<type>Лекция</type>

<teacher>Клименков Сергей Викторович</teacher>

<format>Очно</format>

</second>

</pair1>

<pair2>

<day>Суббота</day>

<first>

<start>10:00</start>

<end>11:30</end>

<group>P3106</group>

<name>Математический анализ</name>

<type>Лекция</type>

<teacher>Блейхер Оксана Владимировна</teacher>

<format>Дистанционно</format>

</first>

<second>

<start>13:30</start>

<end>15:00</end>

<group>P3106</group>

<audience>1415</audience>

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

<name>Информатика</name>

<type>Лабораторная</type>

<teacher>Балакшин Павел Валерьевич</teacher>

<format>Очно</format>

</second>

<third>

<start>17:00</start>

<end>18:30</end>

<group>P3106</group>

<audience>2416</audience>

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

<name>История Российской науки и техники</name>

<type>Лекция</type>

<teacher>Васильев Андрей Владимирович</teacher>

<format>Очно</format>

</third>

</pair2>

</root>

# 1. Обязательное задание

Код:

# вариант 17

# 17 XML YAML Среда, суббота

xml\_file = open('файл с расписанием.xml', 'r', encoding='utf-8')

yaml\_file = open('yaml расписание.txt', 'w', encoding='utf-8')

for line in xml\_file:

xml\_string = line

yaml\_string = ''

for i in range(len(xml\_string)):

symbol = xml\_string[i]

if symbol == '\t':

yaml\_string += '\t'

elif symbol == ' ':

yaml\_string += ' '

elif symbol == '<':

if xml\_string[i + 1] != '/':

yaml\_string += '-'

else:

break

elif symbol.isalpha() or symbol.isdigit() or symbol in '.:-':

yaml\_string += symbol

elif symbol == '>':

yaml\_string += ': '

yaml\_file.write(yaml\_string + '\n')

Результат:

-root:

-nedeli:

-weeks: 2 4 6 8 10 12 14 16

-pair1:

-day: Среда

-first:

-start: 8:20

-end: 9:50

-group: P3106

-audience: 1216

-adress: ул. Ломоносова д. 9 лит. М

-name: Информатика

-type: Лекция

-teacher: Балакшин Павел Валерьевич

-format: Очно

-second:

-start: 10:00

-end: 11:30

-group: P3106

-audience: 1216

-adress: ул. Ломоносова д. 9 лит. М

-name: Основы профессиональной деятельности

-type: Лекция

-teacher: Клименков Сергей Викторович

-format: Очно

-pair2:

-day: Суббота

-first:

-start: 10:00

-end: 11:30

-group: P3106

-name: Математический анализ

-type: Лекция

-teacher: Блейхер Оксана Владимировна

-format: Дистанционно

-second:

-start: 13:30

-end: 15:00

-group: P3106

-audience: 1415

-adress: Кронверкский пр. д. 49 лит. А

-name: Информатика

-type: Лабораторная

-teacher: Балакшин Павел Валерьевич

-format: Очно

-third:

-start: 17:00

-end: 18:30

-group: P3106

-audience: 2416

-adress: Кронверкский пр. д. 49 лит. А

-name: История Российской науки и техники

-type: Лекция

-teacher: Васильев Андрей Владимирович

-format: Очно

# 2. Дополнительное задание №1

Код:

import xmltodict

import yaml

with open('файл с расписанием.xml', encoding='utf-8') as xml\_file:

data = xmltodict.parse(xml\_file.read())

with open('yaml расписание.txt', 'w', encoding='utf-8') as yaml\_file:

yaml.dump(data, yaml\_file, allow\_unicode=True, sort\_keys=False)

Результат:

root:

nedeli:

weeks: '[2, 4, 6, 8, 10, 12, 14, 16]'

pair1:

day: Среда

first:

start: '8:20'

end: '9:50'

group: P3106

audience: '1216'

adress: ул. Ломоносова, д. 9, лит. М

name: Информатика

type: Лекция

teacher: Балакшин Павел Валерьевич

format: Очно

second:

start: '10:00'

end: '11:30'

group: P3106

audience: '1216'

adress: ул. Ломоносова, д. 9, лит. М

name: Основы профессиональной деятельности

type: Лекция

teacher: Клименков Сергей Викторович

format: Очно

pair2:

day: Суббота

first:

start: '10:00'

end: '11:30'

group: P3106

name: Математический анализ

type: Лекция

teacher: Блейхер Оксана Владимировна

format: Дистанционно

second:

start: '13:30'

end: '15:00'

group: P3106

audience: '1415'

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

name: Информатика

type: Лабораторная

teacher: Балакшин Павел Валерьевич

format: Очно

third:

start: '17:00'

end: '18:30'

group: P3106

audience: '2416'

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

name: История Российской науки и техники

type: Лекция

teacher: Васильев Андрей Владимирович

format: Очно

Оба файла имеют одинаковую иерархию и структуру, включая корневой элемент root, подэлементы nedeli, pair1, и pair2. В обоих файлах указаны одни и те же значения для weeks, day в pair1 и pair2.

В первом файле адреса содержат "д." и "лит.", разделенные пробелом, тогда как во втором файле адреса имеют запятые.

В первом файле weeks представлено как строка с пробелами, а во втором — как строка с квадратными скобками.

В первом файле значения времени представлены без кавычек, а во втором — в кавычках.

# 3. Дополнительное задание №2

Код:

# вариант 17

# 17 XML YAML Среда, суббота

import re

xml\_file = open('файл с расписанием.xml', 'r', encoding='utf-8')

yaml\_file = open('yaml расписание.txt', 'w', encoding='utf-8')

for line in xml\_file:

xml\_string = line

yaml\_string = ''

patt = re.findall(r'\s{0,}<\w{1,}>', xml\_string)

if patt:

name = patt[0]

name = name.replace('<', '')

name = name.replace('>', '')

if name[0] == '\t':

num = name.count('\t')

name = num \* '\t' + '-' + name[num:]

yaml\_string += name + ': '

patt = re.findall(r'>(.\*)<', xml\_string)

yaml\_string += (patt[0] if patt else '')

yaml\_file.write(yaml\_string + '\n' if yaml\_string else '')

Результат:

root:

nedeli:

weeks: [2, 4, 6, 8, 10, 12, 14, 16]

pair1:

day: Среда

first:

start: 8:20

end: 9:50

group: P3106

audience: 1216

adress: ул. Ломоносова, д. 9, лит. М

name: Информатика

type: Лекция

teacher: Балакшин Павел Валерьевич

format: Очно

second:

start: 10:00

end: 11:30

group: P3106

audience: 1216

adress: ул. Ломоносова, д. 9, лит. М

name: Основы профессиональной деятельности

type: Лекция

teacher: Клименков Сергей Викторович

format: Очно

pair2:

day: Суббота

first:

start: 10:00

end: 11:30

group: P3106

name: Математический анализ

type: Лекция

teacher: Блейхер Оксана Владимировна

format: Дистанционно

second:

start: 13:30

end: 15:00

group: P3106

audience: 1415

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

name: Информатика

type: Лабораторная

teacher: Балакшин Павел Валерьевич

format: Очно

third:

start: 17:00

end: 18:30

group: P3106

audience: 2416

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

name: История Российской науки и техники

type: Лекция

teacher: Васильев Андрей Владимирович

format: Очно

Оба файла имеют одинаковую иерархию, включая корневой элемент root, подэлементы nedeli, pair1 и pair2.

В первом файле weeks представлен как строка с пробелами, а во втором — как массив с квадратными скобками.

В первом файле адреса содержат пробелы между частями (например, "д. 9 лит. М"), а во втором — запятые (например, "д. 9, лит. М")

# 4. Дополнительное задание №3

Код:

import xml.etree.ElementTree as ET

import yaml

def xml\_to\_yaml(xml\_string):

root = ET.fromstring(xml\_string)

schedule = []

for n in root.findall('nedeli'):

course\_data={

'weeks': n.find('weeks').text,

}

schedule.append(course\_data)

for course in root.findall('pair1'):

course\_data = {

'day': course.find('day').text,

}

schedule.append(course\_data)

for d1 in course.findall('first'):

course\_data = {

'start': d1.find('start').text,

'end': d1.find('end').text,

'group': d1.find('group').text,

'audience': d1.find('audience').text,

'adress': d1.find('adress').text,

'name': d1.find('name').text,

'type': d1.find('type').text,

'teacher': d1.find('teacher').text,

'format': d1.find('format').text,

}

schedule.append({'first': course\_data})

for d2 in course.findall('second'):

course\_data = {

'start': d2.find('start').text,

'end': d2.find('end').text,

'group': d2.find('group').text,

'audience': d2.find('audience').text,

'adress': d2.find('adress').text,

'name': d2.find('name').text,

'type': d2.find('type').text,

'teacher': d2.find('teacher').text,

'format': d2.find('format').text,

}

schedule.append({'second': course\_data})

for course in root.findall('pair2'):

course\_data = {

'day': course.find('day').text,

}

schedule.append(course\_data)

for d1 in course.findall('first'):

course\_data = {

'start': d1.find('start').text,

'end': d1.find('end').text,

'group': d1.find('group').text,

'name': d1.find('name').text,

'type': d1.find('type').text,

'teacher': d1.find('teacher').text,

'format': d1.find('format').text,

}

schedule.append({'first': course\_data})

for d2 in course.findall('second'):

course\_data = {

'start': d2.find('start').text,

'end': d2.find('end').text,

'group': d2.find('group').text,

'audience': d2.find('audience').text,

'adress': d2.find('adress').text,

'name': d2.find('name').text,

'type': d2.find('type').text,

'teacher': d2.find('teacher').text,

'format': d2.find('format').text,

}

schedule.append({'second': course\_data})

for d3 in course.findall('third'):

course\_data = {

'start': d3.find('start').text,

'end': d3.find('end').text,

'group': d3.find('group').text,

'audience': d3.find('audience').text,

'adress': d3.find('adress').text,

'name': d3.find('name').text,

'type': d3.find('type').text,

'teacher': d3.find('teacher').text,

'format': d3.find('format').text,

}

schedule.append({'third': course\_data})

return schedule

# Чтение XML файла и запись в YAML файл

xml\_file\_path = 'файл с расписанием.xml' # Замените на ваш XML файл

yaml\_file\_path = 'yaml расписание.txt' # Замените на желаемый путь для YAML файла

with open(xml\_file\_path, 'r', encoding='utf-8') as xml\_file:

xml\_content = xml\_file.read()

yaml\_content = xml\_to\_yaml(xml\_content)

with open(yaml\_file\_path, 'w', encoding='utf-8') as yaml\_file:

yaml.dump(yaml\_content, yaml\_file, allow\_unicode=True, sort\_keys=False)

Результат:

- weeks: '[2, 4, 6, 8, 10, 12, 14, 16]'

- day: Среда

- first:

start: '8:20'

end: '9:50'

group: P3106

audience: '1216'

adress: ул. Ломоносова, д. 9, лит. М

name: Информатика

type: Лекция

teacher: Балакшин Павел Валерьевич

format: Очно

- second:

start: '10:00'

end: '11:30'

group: P3106

audience: '1216'

adress: ул. Ломоносова, д. 9, лит. М

name: Основы профессиональной деятельности

type: Лекция

teacher: Клименков Сергей Викторович

format: Очно

- day: Суббота

- first:

start: '10:00'

end: '11:30'

group: P3106

name: Математический анализ

type: Лекция

teacher: Блейхер Оксана Владимировна

format: Дистанционно

- second:

start: '13:30'

end: '15:00'

group: P3106

audience: '1415'

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

name: Информатика

type: Лабораторная

teacher: Балакшин Павел Валерьевич

format: Очно

- third:

start: '17:00'

end: '18:30'

group: P3106

audience: '2416'

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

name: История Российской науки и техники

type: Лекция

teacher: Васильев Андрей Владимирович

format: Очно

Оба файла содержат информацию о неделях, днях и занятиях, включая время, группу, аудиторию, название предмета, тип занятия, преподавателя и формат.

В первом файле weeks представлен как строка с пробелами, а во втором — как массив с квадратными скобками.

Во втором файле время и аудитории обрамлены кавычками, тогда как в первом файле они представлены без кавычек.

В первом файле адреса записаны без запятых между частями, а во втором — с запятыми.

# 5. Дополнительное задание №4

Код:

import time

start\_time = time.perf\_counter()

for i in range(100):

import main

end\_time = time.perf\_counter()

d=end\_time-start\_time

print(f"main - {end\_time - start\_time}")

start\_time = time.perf\_counter()

for i in range(100):

import dop1

end\_time = time.perf\_counter()

d=end\_time-start\_time

print(f"Доп 1 - {end\_time - start\_time}")

start\_time = time.perf\_counter()

for i in range(100):

import dop2

end\_time = time.perf\_counter()

print(f"Доп 2 - {end\_time - start\_time}")

start\_time = time.perf\_counter()

for i in range(100):

import dop3

end\_time = time.perf\_counter()

print(f"Доп 3 - {end\_time - start\_time}")

Результат:

Результат представлен на Рисунок

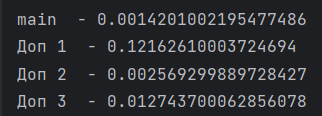


Рисунок 1

Программа обязательного задания работает быстрее всего, потому что там реализована работа только исходного файла. Доп 2 работает чуть дольше, потому что применено использование регулярных выражений. Доп 3 расширяет функционал парсера, следовательно, увеличивает время работы программы. Доп 1 работает дольше всех, потому что Создает полное представление XML в виде словаря, что может занимать больше времени и памяти, особенно если XML-файл большой.

# 6. Дополнительное задание №5

Код:

# xml -> wml

xml\_file = open('файл с расписанием.xml', 'r', encoding='utf-8')

wml\_file = open('расписание.txt', 'w', encoding='utf-8')

wml\_file.write('<wml>\n')

wml\_file.write(' <card>\n')

for line in xml\_file:

xml\_string = line

wml\_string = ''

if xml\_string.strip() == '<?xml version="1.0" encoding="UTF-8"?>' or xml\_string.strip() == '<root>':

continue

else:

for i in range(len(xml\_string)):

symbol = xml\_string[i]

if symbol == '\t':

wml\_string += '\t'

elif symbol == ' ':

wml\_string += ' '

elif symbol == '<':

if xml\_string[i + 1] != '/':

wml\_string += '<p>'

else:

break

elif symbol.isalpha() or symbol.isdigit() or symbol in '.:-':

wml\_string += symbol

elif symbol == '>':

wml\_string += ': '

if len(wml\_string.strip())==0:

continue

else:

wml\_file.write(f' {wml\_string}</p>\n')

wml\_file.write(' </card>\n')

wml\_file.write('</wml>')

Результат:

<wml>

<card>

<p>nedeli: </p>

<p>weeks: 2 4 6 8 10 12 14 16</p>

<p>pair1: </p>

<p>day: Среда</p>

<p>first: </p>

<p>start: 8:20</p>

<p>end: 9:50</p>

<p>group: P3106</p>

<p>audience: 1216</p>

<p>adress: ул. Ломоносова д. 9 лит. М</p>

<p>name: Информатика</p>

<p>type: Лекция</p>

<p>teacher: Балакшин Павел Валерьевич</p>

<p>format: Очно</p>

<p>second: </p>

<p>start: 10:00</p>

<p>end: 11:30</p>

<p>group: P3106</p>

<p>audience: 1216</p>

<p>adress: ул. Ломоносова д. 9 лит. М</p>

<p>name: Основы профессиональной деятельности</p>

<p>type: Лекция</p>

<p>teacher: Клименков Сергей Викторович</p>

<p>format: Очно</p>

<p>pair2: </p>

<p>day: Суббота</p>

<p>first: </p>

<p>start: 10:00</p>

<p>end: 11:30</p>

<p>group: P3106</p>

<p>name: Математический анализ</p>

<p>type: Лекция</p>

<p>teacher: Блейхер Оксана Владимировна</p>

<p>format: Дистанционно</p>

<p>second: </p>

<p>start: 13:30</p>

<p>end: 15:00</p>

<p>group: P3106</p>

<p>audience: 1415</p>

<p>adress: Кронверкский пр. д. 49 лит. А</p>

<p>name: Информатика</p>

<p>type: Лабораторная</p>

<p>teacher: Балакшин Павел Валерьевич</p>

<p>format: Очно</p>

<p>third: </p>

<p>start: 17:00</p>

<p>end: 18:30</p>

<p>group: P3106</p>

<p>audience: 2416</p>

<p>adress: Кронверкский пр. д. 49 лит. А</p>

<p>name: История Российской науки и техники</p>

<p>type: Лекция</p>

<p>teacher: Васильев Андрей Владимирович</p>

<p>format: Очно</p>

</card>

</wml>

Исходный XML файл был преобразован в файл с расширением WML. WML — это язык разметки, созданный для отображения контента на мобильных устройствах. Вместо тега <root> появился тег <wml>. WML основан на XML.

# Вывод:

В ходе выполнения лабораторной работы, я познакомился с форматами XML, YAML, WML. Я научился преобразовывать форматы между собой с помощью библиотек и кода. Также я повторил регулярные выражения.

# Источники:

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.
2. Грошев А.С. Г89 Информатика: Учебник для вузов / А.С. Грошев. – Архангельск, Арханг. гос. техн. ун-т, 2010. -470с.