A close up of a logo

Description automatically generated

Факультет ПИиКТ

Лабораторная работа №4 по Информатике

Исследование протоколов, форматов обмена информацей и языков разметки документов

Вариант 23

**Выполнил:**

Давааням Баясгалан

группа P3111

**Преподаватель:**

Малышева Татьяна Алексеевна

г. Санкт-Петербург

2021 год

Оглавление

[Вариант 3](#_Toc88527346)

[Исходный файл (XML) 3](#_Toc88527347)

[Обязательное задание 4](#_Toc88527348)

[Код программы на Python 4](#_Toc88527349)

[Результат программы 5](#_Toc88527350)

[Допольнительное задание №1 5](#_Toc88527351)

[Код программы на Python 5](#_Toc88527352)

[Результат программы 6](#_Toc88527353)

[Допольнителное задание №2 6](#_Toc88527354)

[Код программы на Python 7](#_Toc88527355)

[Результат программы 7](#_Toc88527356)

[Допольнительное задание №3 8](#_Toc88527357)

[Код программы на Python 8](#_Toc88527358)

[Результат программы 9](#_Toc88527359)

[Вывод 9](#_Toc88527360)

[Список литературы 9](#_Toc88527361)

# Вариант

|  |  |  |  |
| --- | --- | --- | --- |
| 23 | XML | YAML | Четверг |

# Исходный файл (XML)

<?xml version="1.0" encoding="UTF-8"?>

<timetable>

   <subject\_1>

      <lesson>Математика (Практика)</lesson>

      <location>ул.Ломоносова д.9ь лит.М </location>

      <room>1512</room>

      <teacher>Холодова Светлана Евгеньевна</teacher>

      <time>8:20 - 9:50</time>

      <week>Все</week>

      <weekday>Четверг</weekday>

   </subject\_1>

   <subject\_2>

      <lesson>Математика (Практика)</lesson>

      <location>ул.Ломоносова д.9ь лит.М</location>

      <room>1329</room>

      <teacher>Холодова Светлана Евгеньевна</teacher>

      <time>11:40-13:10</time>

      <week>Все</week>

      <weekday>Четверг</weekday>

   </subject\_2>

   <subject\_3>

      <lesson>Математика (Лекция)</lesson>

      <location>ул.Ломоносова д.9ь лит.М</location>

      <room>1222</room>

      <teacher>Холодова Светлана Евгеньевна</teacher>

      <time>13:30-15:00</time>

      <week>Все</week>

      <weekday>Четверг</weekday>

   </subject\_3>

</timetable>

# Обязательное задание

1. Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.
2. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

## **Код программы на Python**

def parser():

    address\_input = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_xml.xml"

    address\_output = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_yaml.yml"

    input\_file = open(address\_input,'r',encoding='utf-8')

    output\_file = open(address\_output,'w',encoding='utf-8')

    lines = input\_file.readlines()

    countTab = 0

    for line in lines:

        line = line.strip()

        if "<?" in line[:2]:

            continue

        elif "</" in line[:2]:

            countTab -= 1

        elif "<" in line[0]:

            start = line.index("<")

            end = line.index(">")

            cline = line[end+1:]

            if len(cline)<1:

                output\_file.write(countTab \* '  ' + line[start+1:end] + ":\n" )

                countTab += 1

            else:

                if cline.find("<") > -1:

                    close\_tag = cline.index("<")

                    output\_file.write(countTab \* '  ' + line[start+1:end] + ": " + cline[:close\_tag] + "\n")

                else:

                    output\_file.write(countTab \* '  ' + line[start+1:end] + ": " + cline + "\n")

                    countTab += 1

        else:

            if "<" in line:

                end = line.index("<")

                output\_file.write(countTab \* '  ' + line[:end] + '\n')

                countTab -= 1

            else:

                output\_file.write(countTab \* '  ' + line + '\n')

    input\_file.close()

    output\_file.close()

## **Результат программы**

timetable:

  subject\_1:

    lesson: Математика (Практика)

    location: ул.Ломоносова д.9ь лит.М

    room: 1512

    teacher: Холодова Светлана Евгеньевна

    time: 8:20 - 9:50

    week: Все

    weekday: Четверг

  subject\_2:

    lesson: Математика (Практика)

    location: ул.Ломоносова д.9ь лит.М

    room: 1329

    teacher: Холодова Светлана Евгеньевна

    time: 11:40-13:10

    week: Все

    weekday: Четверг

  subject\_3:

    lesson: Математика (Лекция)

    location: ул.Ломоносова д.9ь лит.М

    room: 1222

    teacher: Холодова Светлана Евгеньевна

    time: 13:30-15:00

    week: Все

    weekday: Четверг

# Допольнительное задание №1

1. Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
2. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
3. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.

## **Код программы на Python**

import xmlplain

def parser\_with\_library():

    address\_input = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_xml.xml"

    address\_output = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_with\_library\_yaml.yml"

    input\_file = open(address\_input,'r',encoding='utf-8')

    output\_file = open(address\_output,'w',encoding='utf-8')

    root = xmlplain.xml\_to\_obj(input\_file, strip\_space=True, fold\_dict=True)

    xmlplain.obj\_to\_yaml(root, output\_file)

    input\_file.close()

    output\_file.close()

## **Результат программы**

timetable:

  subject\_1:

    lesson: Математика (Практика)

    location: 'ул.Ломоносова д.9ь лит.М '

    room: '1512'

    teacher: Холодова Светлана Евгеньевна

    time: 8:20 - 9:50

    week: Все

    weekday: Четверг

  subject\_2:

    lesson: Математика (Практика)

    location: ул.Ломоносова д.9ь лит.М

    room: '1329'

    teacher: Холодова Светлана Евгеньевна

    time: 11:40-13:10

    week: Все

    weekday: Четверг

  subject\_3:

    lesson: Математика (Лекция)

    location: ул.Ломоносова д.9ь лит.М

    room: '1222'

    teacher: Холодова Светлана Евгеньевна

    time: 13:30-15:00

    week: Все

    weekday: Четверг

# Допольнителное задание №2

1. Переписать исходный код, добавив в него использование регулярных выражений.
2. Сравнить полученные результаты и объяснить их сходство/различие.

## **Код программы на Python**

import re

def parser\_with\_re():

    address\_input = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_xml.xml"

    address\_output = "D:\ITMO\\1st Semester\Computer Science (Basics)\lab4\Timetable\_with\_re\_yaml.yml"

    input\_file = open(address\_input,'r',encoding='utf-8')

    output\_file = open(address\_output,'w',encoding='utf-8')

    xml\_lines = input\_file.readlines()

    countTab = 0

    for xml\_line in xml\_lines:

        line = xml\_line.strip()

        if re.search(r'<[а-я\w\d]\*>[а-я\W\w\d]\*<\/[а-я\w\d]\*>',line,re.I):

            str = ' '.join(re.findall(r'<[а-я\w\d]\*>',line))[1:-1]

            str = str + ": "+ ' '.join(re.findall(r'>[а-я\W\w\d]\*<\/',line))[1:-2]

            output\_file.write(countTab \* '  ' + str + "\n")

        elif re.search(r'<[а-я\w\d]\*>',line,re.I):

            str\_1 = ''.join(re.findall(r'<[а-я\w\d]\*>',line))[1:-1] + ':'

            output\_file.write(countTab \* '  ' + str\_1 + "\n")

            countTab += 1

        else:

            if re.search(r'^[а-я\w\d]+',line,re.I):

                str\_2 = ''.join(re.findall(r'[а-я\w\d\D]+',line,re.I))

                print(str\_2)

                output\_file.write(countTab \* '  ' + str\_2 + '\n')

                if re.search(r'<\/[а-я\w\d]\*>',line):

                    countTab -=1

            elif re.search(r'<\?',line):

                continue

            else:

                countTab -=1

    input\_file.close()

    output\_file.close()

## **Результат программы**

timetable:

  subject\_1:

    lesson: Математика (Практика)

    location: ул.Ломоносова д.9ь лит.М

    room: 1512

    teacher: Холодова Светлана Евгеньевна

    time: 8:20 - 9:50

    week: Все

    weekday: Четверг

  subject\_2:

    lesson: Математика (Практика)

    location: ул.Ломоносова д.9ь лит.М

    room: 1329

    teacher: Холодова Светлана Евгеньевна

    time: 11:40-13:10

    week: Все

    weekday: Четверг

  subject\_3:

    lesson: Математика (Лекция)

    location: ул.Ломоносова д.9ь лит.М

    room: 1222

    teacher: Холодова Светлана Евгеньевна

    time: 13:30-15:00

    week: Все

    weekday: Четверг

# Допольнительное задание №3

1. 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.
2. Проанализировать полученные результаты и объяснить их сходство/различие.

## **Код программы на Python**

import time

parser\_time = time.time()

for i in range(10):

    parser()

parser\_time = time.time() - parser\_time

parser\_with\_library\_time = time.time()

for i in range(10):

    parser\_with\_library()

parser\_with\_library\_time = time.time() - parser\_with\_library\_time

parser\_with\_re\_time = time.time()

for i in range(10):

    parser\_with\_re()

parser\_with\_re\_time = time.time() - parser\_with\_re\_time

print('Parser time: ' + str(parser\_time))

print('Parser time with library: ' + str(parser\_with\_library\_time))

print('Parser time with re: ' + str(parser\_with\_re\_time))

## **Результат программы**

Text

Description automatically generated

# Вывод

При выполнении лабораторной работы я ознакомился с XML и YAML и научил их синтаксис. В языке Python я научил использование готовых библиотек и заметил разницу времени выполнений между готовым библиотекой и сам написанным конвертором.

# Список литературы

<https://habr.com/ru/post/309242/>

<https://ru.wikipedia.org/wiki/Форма_Бэкуса_—_Наура>