Федеральное государственное автономное образовательное учреждение высшего образования

«Санкт-Петербургский национальный исследовательский

**университет ИТМО»**

**Институт ПКиКТ**

**Дисциплина: Информатика**

**Лабораторная работа №4**

**«Исследование протоколов, форматов обмена информацией и языков разметки документов»**

**Вариант 11.**

Выполнил: Сиразетдинов Азат Ниязович

Группа: Р3116

Преподователь: Машина Е.А.

2022г.

# Оглавление

Оглавление

[Оглавление 2](#_Toc118323504)

[Задание 3](#_Toc118323505)

[Выполнение работы 4](#_Toc118323506)

[Обязательное задание 5](#_Toc118323507)

[Дополнительное задание 1 6](#_Toc118323508)

[Дополнительное задание 2 7](#_Toc118323509)

[Дополнительное задание 3 8](#_Toc118323510)

[Дополнительное задание 4 9](#_Toc118323511)

[Вывод 10](#_Toc118323512)

[Приложение 1 11](#_Toc118323513)

[Литература 12](#_Toc118323514)

# Задание

1. Определить номер варианта как остаток деления на 36 порядкового номера в списке группы в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь.

2. Изучить форму Бэкуса-Наура. 3. Изучить особенности языков разметки/форматов JSON, YAML, XML. 4. Понять устройство страницы с расписанием для своей группы: http://itmo.ru/ru/schedule/0/P3110/schedule.htm

5. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы в выбранном дне 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.

6. 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.

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

8. 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.

c) Сравнить полученные результаты и объяснить их сходство/различие.

9. 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.

1b) Сравнить полученные результаты и объяснить их сходство/различие.

10. Дополнительное задание №3 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную). a) Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и программу из дополнительного задания №2, сравнить стократное время выполнения парсинга + конвертации в цикле.

b) Проанализировать полученные результаты и объяснить их сходство/различие.

11. Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную). c) 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.

d) Проанализировать полученные результаты, объяснить особенности использования формата.

12. Проверить, что все пункты задания выполнены и выполнены верно. 13. Написать отчёт о проделанной работе. 14. Подготовиться к устным вопросам на защите.

Исходный формат: XML

Результирующий формат: YAML

День недели: вторник

# Выполнение работы

Задание 1

Расписание вторника:

Изображение выглядит как текст

Автоматически созданное описание

XML:

<https://github.com/Azat2202/InformaticsLab4/blob/master/Myxml.xml>

# Обязательное задание

Код программы: <https://github.com/Azat2202/InformaticsLab4/blob/master/paser_xml_to_yaml.py>

Программа осуществляет перевод файла Myxml.xml в Myyaml.yaml

# Дополнительное задание 1

Код использующий библиотеки:

<https://github.com/Azat2202/InformaticsLab4/blob/master/paser_xml_to_yaml_with_libs.py>

Используемая библиотека: xmlplain

Обе программы генерируют правильный yaml файл, но в библиотеки все переменные-строки, а у меня реализовано определение типа

# Дополнительное задание 2

Код использующий регулярные выражения:

<https://github.com/Azat2202/InformaticsLab4/blob/master/parser_xml_to_yaml_regex.py>

Вывод получился тем же самым, потому что я просто заменил функции на регулярные выражения

# Дополнительное задание 3

Время стократного выполнения программы без регулярных выражений: 0.29 секунд

Время стократного выполнения программы с регулярными выражениями: 0.68 секунды

Время стократного выполнения программы с библиотекой: 3.27 секунд

Вывод: без регулярных выражений код работает достаточно близко. В моем классе при создании каждого объекта заново компилируются регулярные выражения, если их компилировать единожды, то время выполнения существенно сократилась. А библиотека, найденная в интернете, оказалась слишком медленно, хотя она в отличии от моего кода работает для общего случая, а не частного.

# Дополнительное задание 4

Я выбрал табличный язык данных csv

Код парсера: <https://github.com/Azat2202/InformaticsLab4/blob/master/parser_xml_to_csv.py>

Рисунок 1. Полученный csv файл в Excel

Странный формат данных, сложно читается человеком и не так удобен для машины. SQL в этих отношениях лучше.

# Вывод

В процессе выполнения задания лабораторной работы я научился делать парсеры, продвинулся в использовании регулярных выражений, узнал про форматы обмена информациией и языки разметки.

# Приложение 1

Изображение выглядит как текст, стол

Автоматически созданное описание

Рисунок . Полученный csv файл в Excel

# Литература

1. <https://tproger.ru/translations/yaml-za-5-minut-sintaksis-i-osnovnye-vozmozhnosti/>
2. <https://habr.com/ru/post/524288/>
3. Лекция 4 Балакшин П.В.