Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной Инженерии и Компьютерной Техники

**Лабораторная работа по информатике №4**

Вариант: 14

Практик: Машина Екатерина Алексеевна

Выполнил: Состанов Тимур Айратович

Группа: P3114

Санкт-Петербург, 2022

**Оглавление**

[**Задание** 3](#_Toc121572030)

[**Ход работы** 4](#_Toc121572031)

[**Основное задание** 4](#_Toc121572032)

[**Дополнительное задание №1** 4](#_Toc121572033)

[**Дополнительное задание №2** 4](#_Toc121572034)

[**Дополнительное задание №3** 5](#_Toc121572035)

[**Дополнительное задание №4** 6](#_Toc121572036)

[**Вывод** 7](#_Toc121572037)

[**Список литературы** 7](#_Toc121572038)

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

**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.

**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.

**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.

**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.

# **Ход работы**

Все файлы лабораторной работы:

<https://github.com/tsostanov/lab4_inf>

Из-за некорректного отображения файлов (из-за переносов строк) буду вставлять ссылки на файлы

**Исходные файлы:**

Записанный в одну строку (timetable.json):

<https://github.com/tsostanov/lab4_inf/blob/main/timetable.json>

Записанный с учетом вложенности (re\_timetable.json):

<https://github.com/tsostanov/lab4_inf/blob/main/re_timetable.json>

## **Основное задание**

<https://github.com/tsostanov/lab4_inf/blob/main/main.py>

Данную программу я делал под работу с данными, записанными в одну строку (в такой формат довольно просто преобразовать файл, записанный с учетом вложенности).

## **Дополнительное задание №1**

<https://github.com/tsostanov/lab4_inf/blob/main/library.py>

Я использовал две библиотеки: JSON и YAML, так как они являются самыми популярными для работы с файлами в форматах JSON и YAML, а еще они общего характера, то есть не заточены только под парсер из JSON в YAML, поэтому могут применятся и в других целях.

## **Дополнительное задание №2**

<https://github.com/tsostanov/lab4_inf/blob/main/main_re.py>

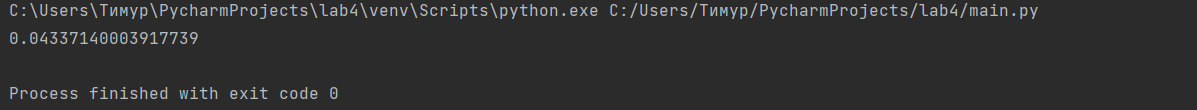
Так как код из основного задания сложно оптимизировать под использование регулярных выражений (они просто не будут содержательными), я решил написать эту программу для JSON-файла, записанного с учетом вложенности

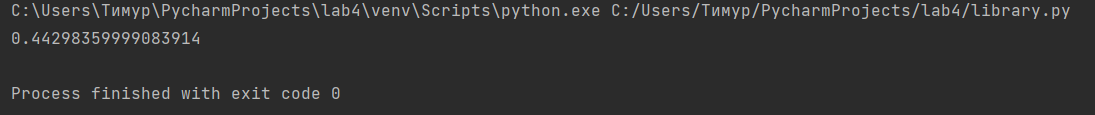
Результаты

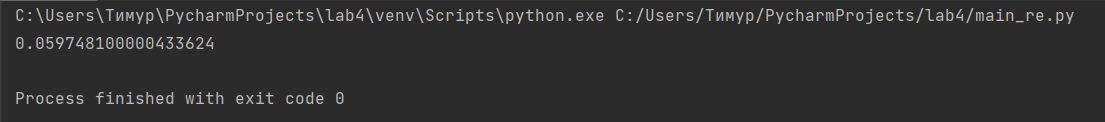
<https://github.com/tsostanov/lab4_inf/blob/main/timetable.yaml>

Все программы дают одинаковый результат

## **Дополнительное задание №3**

 *Рисунок 1 - результат работы основной программы*

 *Рисунок 2 – результат работы программы с библиотеками*

* Рисунок 3 – результат работы программы с регулярными выражениями*

Дольше всего работала программа с готовыми библиотеками, так как она реализует полноценный парсинг с полной обработкой входных и выходных файлов. И еще немного костыльная обработка символов, записанных в кодировке UTF-8, не добавляет скорости программе.

Код с регулярными выражениями работает немного дольше основного кода, так как сами регулярные выражения довольно медленные. Но разница незначительна из-за малых размеров читаемого файла.

## **Дополнительное задание №4**

Я сделал парсер из JSON в CSV, только для удобности дальнейшего отображения пришлось заменить запятные-разделители на символ “;” (так как в исходном файле была информация с запятыми)

<https://github.com/tsostanov/lab4_inf/blob/main/tocsv.py>

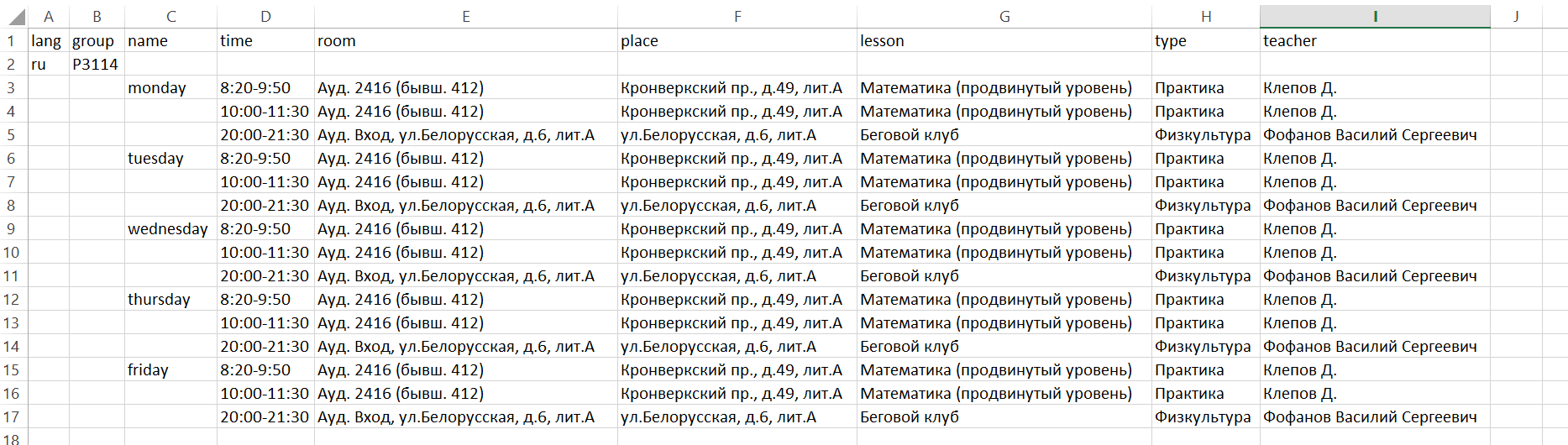
За основу я брал код из основного задания, так как он уже умеет получать информацию из JSON файла. Поэтому данная программа работает с JSON файлами, записанными в одну строку.

**Результат работы дополнительного задания №4**

<https://github.com/tsostanov/lab4_inf/blob/main/to_csv.csv>

Этот же файл, открытый в Excel (для удобства анализа данных человеком)

Разделитель – “;”



# **Вывод**

Во время выполнения лабораторной работы я узнал о языках разметки JSON, XML и YAML, научился с ними работать и переводить один тип данных в другой с помощью средств языка Python. Научился работать с библиотекой time для измерения времени и некоторыми библиотеками для анализа этих форматов и автоматического парсинга между ними. Для дополнительного задания №4 немного поразбирался с другими языками разметки и сделал парсер из JSON в CSV.

# **Список литературы**

1. Лямин А.В., Череповская Е.Н. Объектно-ориентированное программирование. Компьютерный практикум. – СПб: Университет ИТМО, 2017. – 143 с. – Режим доступа:

<https://books.ifmo.ru/file/pdf/2256.pdf>

1. «Пишем изящный парсер на Питоне»:

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