Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет Программной инженерии и компьютерной техники

**Лабораторная работа** №**5**

Работа с электронными таблицами

Вариант №16

Группа: P3112

Выполнил:

Соколов А. В.

Проверил:

к.т.н.; преподаватель Белозубов А.В.

ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 3](#_Toc118245340)

[РЕШЕНИЕ ЗАДАНИЙ 4](#_Toc118245341)

[ЗАКЛЮЧЕНИЕ 12](#_Toc118245342)

[СПИСОК ЛИТЕРАТУРЫ 13](#_Toc118245343)

# ВВЕДЕНИЕ

Работа заключается в конвертации файлов разных языков разметки. Инструменты; используемые для решения описанных выше задач: ноутбук.

Цель работы: изучить особенности языков разметки/форматов JSON; YAML; XML.

Задача: сформировать файл с расписанием в формате; указанном в задании в качестве исходного.

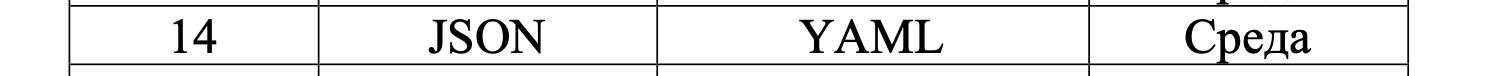


Рисунок 1. Вариант задания.

Содержимое файла input.json с расписанием на среду:

{

"Информатика (лаб) первая пара": {

"Формат обучения": "Очно - дистанционный";

"Четность недели": "нечетная неделя";

"Время": "11:40-13:10";

"Неделя": [3; 5; 7; 9; 11; 13; 15; 17];

"Аудитория": "2308 (бывш. 306) ауд.";

"Корпус": "Кронверкский пр.; д.49; лит.A";

"Преподаватель": "Белозубов Александр Владимирович"

};

"Информатика (лаб) вторая пара": {

"Формат обучения": "Очно - дистанционный";

"Четность недели": "нечетная неделя";

"Время": "13:30-15:00";

"Неделя": [3; 5; 7; 9; 11; 13; 15; 17];

"Аудитория": "2308 (бывш. 306) ауд.";

"Корпус": "Кронверкский пр.; д.49; лит.A";

"Преподаватель": "Белозубов Александр Владимирович"

};

"Математика (продвинутый уровень) (практика) первая пара": {

"Формат обучения": "Очный";

"Четность недели": "Четная и нечетная";

"Время": "17:00-18:30";

"Неделя": [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17];

"Аудитория": "1316 (бывш. 359) ауд.";

"Корпус": "Кронверкский пр.; д.49; лит.A";

"Преподаватель": "Резцова Ангелина Максимовна"

};

"Математика (продвинутый уровень) (практика) вторая пара": {

"Формат обучения": "Очный";

"Четность недели": "Четная и нечетная";

"Время": "18:40-20:10";

"Неделя": [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17];

"Аудитория": "1316 (бывш. 359) ауд.";

"Корпус": "Кронверкский пр.; д.49; лит.A";

"Преподаватель": "Резцова Ангелина Максимовна"

}

}

# РЕШЕНИЕ ЗАДАНИЙ

**Задание:**

Написать программу на языке Python 3.x; которая бы осуществляла парсинг и конвертацию исходного файла в новый.

**Код программы:**

f1 = open('input.json'; 'r'; encoding="utf8")

f2 = open('output1.yml'; 'w'; encoding="utf8")

buffer = ""

s = ''

arr = []

for line in f1:

s = line

for i in range(1; len(s)):

if '[' in s:

while ']' in s:

arr.append(s[i])

break

if s[i] != '"' and s[i] != '{' and s[i] != '}' and '};' not in s :

buffer = buffer + s[i]

buffer = buffer.replace('['; '').replace(']'; '')

f2.write(buffer)

f1.close()

f2.close()

**Результат работы программы:**

Информатика (лаб) первая пара:

Формат обучения: Очно - дистанционный;

Четность недели: нечетная неделя;

Время: 11:40-13:10;

Неделя: 3; 5; 7; 9; 11; 13; 15; 17;

Аудитория: 2308 (бывш. 306) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Белозубов Александр Владимирович

Информатика (лаб) вторая пара:

Формат обучения: Очно - дистанционный;

Четность недели: нечетная неделя;

Время: 13:30-15:00;

Неделя: 3; 5; 7; 9; 11; 13; 15; 17;

Аудитория: 2308 (бывш. 306) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Белозубов Александр Владимирович

Математика (продвинутый уровень) (практика) первая пара:

Формат обучения: Очный;

Четность недели: Четная и нечетная;

Время: 17:00-18:30;

Неделя: 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17;

Аудитория: 1316 (бывш. 359) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Резцова Ангелина Максимовна

Математика (продвинутый уровень) (практика) вторая пара:

Формат обучения: Очный;

Четность недели: Четная и нечетная;

Время: 18:40-20:10;

Неделя: 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17;

Аудитория: 1316 (бывш. 359) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Резцова Ангелина Максимовна

**Задание:**

Найти готовые библиотеки; осуществляющие аналогичный парсинг и конвертацию файлов.

Переписать исходный код; применив найденные библиотеки. Регулярные выражения также нельзя использовать.

**Код программы:**

import json

import yaml

with open('input.json'; 'r'; encoding="utf8") as f:

data = json.load(f)

with open('output2.yml'; 'w'; encoding="utf8") as f:

yaml.dump(data; f; default\_flow\_style=False; allow\_unicode=True)

**Результат работы программы:**

Информатика (лаб) вторая пара:

Аудитория: 2308 (бывш. 306) ауд.

Время: 13:30-15:00

Корпус: Кронверкский пр.; д.49; лит.A

Неделя:

- 3

- 5

- 7

- 9

- 11

- 13

- 15

- 17

Преподаватель: Белозубов Александр Владимирович

Формат обучения: Очно - дистанционный

Четность недели: нечетная неделя

Информатика (лаб) первая пара:

Аудитория: 2308 (бывш. 306) ауд.

Время: 11:40-13:10

Корпус: Кронверкский пр.; д.49; лит.A

Неделя:

- 3

- 5

- 7

- 9

- 11

- 13

- 15

- 17

Преподаватель: Белозубов Александр Владимирович

Формат обучения: Очно - дистанционный

Четность недели: нечетная неделя

Математика (продвинутый уровень) (практика) вторая пара:

Аудитория: 1316 (бывш. 359) ауд.

Время: 18:40-20:10

Корпус: Кронверкский пр.; д.49; лит.A

Неделя:

- 1

- 2

- 3

- 4

- 5

- 6

- 7

- 8

- 9

- 10

- 11

- 12

- 13

- 14

- 15

- 16

- 17

Преподаватель: Резцова Ангелина Максимовна

Формат обучения: Очный

Четность недели: Четная и нечетная

Математика (продвинутый уровень) (практика) первая пара:

Аудитория: 1316 (бывш. 359) ауд.

Время: 17:00-18:30

Корпус: Кронверкский пр.; д.49; лит.A

Неделя:

- 1

- 2

- 3

- 4

- 5

- 6

- 7

- 8

- 9

- 10

- 11

- 12

- 13

- 14

- 15

- 16

- 17

Преподаватель: Резцова Ангелина Максимовна

Формат обучения: Очный

Четность недели: Четная и нечетная

**Задание:**

Сравнить полученные результаты и объяснить их сходство/различие.

**Решение задания:**

Результаты работы программы из задания 1 и задания 2 отличаются реализацией списков. В моем подходе списки интерпретируются как строки для компактности данных; в уже готовой библиотеке каждый элемент списка переносится на новую строку с символом «-»

**Задание:**

Переписать исходный код; добавив в него использование регулярных выражений.

**Код программы:**

import re

jsonFile = (open('input.json'; 'r'; encoding="utf8").read())

outputYml = open('output3.yml'; 'w'; encoding="utf8")

jsonMatch = r'{|};|\[|\];|]|}|\"'

jsonFile\_new = re.sub(jsonMatch; ''; str(jsonFile))

outputYml.write(jsonFile\_new)

outputYml.close()

**Результат работы программы:**

Информатика (лаб) первая пара:

Формат обучения: Очно - дистанционный;

Четность недели: нечетная неделя;

Время: 11:40-13:10;

Неделя: 3; 5; 7; 9; 11; 13; 15; 17

Аудитория: 2308 (бывш. 306) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Белозубов Александр Владимирович

Информатика (лаб) вторая пара:

Формат обучения: Очно - дистанционный;

Четность недели: нечетная неделя;

Время: 13:30-15:00;

Неделя: 3; 5; 7; 9; 11; 13; 15; 17

Аудитория: 2308 (бывш. 306) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Белозубов Александр Владимирович

Математика (продвинутый уровень) (практика) первая пара:

Формат обучения: Очный;

Четность недели: Четная и нечетная;

Время: 17:00-18:30;

Неделя: 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17

Аудитория: 1316 (бывш. 359) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Резцова Ангелина Максимовна

Математика (продвинутый уровень) (практика) вторая пара:

Формат обучения: Очный;

Четность недели: Четная и нечетная;

Время: 18:40-20:10;

Неделя: 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17

Аудитория: 1316 (бывш. 359) ауд.;

Корпус: Кронверкский пр.; д.49; лит.A;

Преподаватель: Резцова Ангелина Максимовна

**Задание:**

Сравнить полученные результаты и объяснить их сходство/различие.

**Решение задания:**

Результаты работы программы из задания 1 и задания 2 и задания 3 отличаются реализацией списков. В моем подходе в задании 1 и 3 списки интерпретируются как строки для компактности данных; в уже готовой библиотеке каждый элемент списка переносится на новую строку с символом «-».

**Задание:**

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.

**Код программы:**

import sys; os; timeit

sys.path.append(os.path.abspath("/Users/isofinly/VSC-Projects/CS-lab-4"))

print('================= TEST 2 =================')

start2 = timeit.default\_timer()

for i in range(100):

import json

import yaml

with open('input.json'; 'r'; encoding="utf8") as f:

data = json.load(f)

with open('output2.yml'; 'w'; encoding="utf8") as f:

yaml.dump(data; f; default\_flow\_style=False; allow\_unicode=True)

stop2 = timeit.default\_timer()

print('Test two successful')

print('Time: '; stop2 - start2)

print('================= TEST 3 =================')

start3 = timeit.default\_timer()

for i in range(100):

import re

jsonFile = (open('input.json'; 'r'; encoding="utf8").read())

outputYml = open('output3.yml'; 'w'; encoding="utf8")

jsonMatch = r'{|};|\[|\];|]|}|\"'

jsonFile\_new = re.sub(jsonMatch; ''; str(jsonFile))

outputYml.write(jsonFile\_new)

outputYml.close()

stop3 = timeit.default\_timer()

print('Test three successful')

print('Time: '; stop3 - start3)

**Результат работы программы:**

================= TEST 2 =================

Test two successful

Time: 0.22404104100132827

================= TEST 3 =================

Test three successful

Time: 0.009978500005672686

**Задание:**

Проанализировать полученные результаты и объяснить их сходство/различие.

**Решение задания:**

Python долго импортирует внешние библиотеки. Встроенные функции работы с регулярными выражениями оптимизированы под более быструю работу.

**Задание:**

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.

import pandas as pd

with open('input.json'; encoding='utf-8-sig') as f\_input:

df = pd.read\_json(f\_input)

df.to\_csv('output4.csv'; encoding='utf-16'; sep='\t'; index=False)

**Задание:**

Проанализировать полученные результаты; объяснить особенности использования формата.

**Решение задания:**

При использовании других видов кодировок может возникать искажение данных при открытии в excel. А также csv — это текстовый файл; в котором содержится информация. Каждая строка — это отдельная строка таблицы; а столбцы отделены один от другого специальными символами - разделителями (например; запятой). В последнее время разделителем может быть не только запятая; но и другие символы (пробел; точка с запятой; табуляция; другое).

# ЗАКЛЮЧЕНИЕ

Задача выполнена полностью. Цель достигнута. Лабораторная работа сделана. Применил навыки программирования для проверки результата и создания программ для проверки решения примеров.

# СПИСОК ЛИТЕРАТУРЫ

**Andress Jason и Linn Ryan** Introduction to Python [В Интернете]. - 2017 г.. - 28 9 2022 г.. - http://tdc-www.harvard.edu/python.pdf.

**Bendersky E.** JSON is YAML; but YAML is not JSON [В Интернете]. - 2 11 2022 г.. - https://eli.thegreenplace.net/2008/11/22/json-yaml-but-yaml-json.

**Benjamin Dan** Top 10 Programming Fonts [В Интернете] // Hivelogic. - 2 11 2022 г.. - http://hivelogic.com/articles/top-10-programming-fonts/.

**Crockford Douglas** The application/json Media Type for JavaScript Object Notation (JSON) [В Интернете]. - 2006 г.. - 2 11 2022 г.. - https://tools.ietf.org/html/rfc4627.

**Ferguson Andrew** A History of Computer Programming Languages [В Интернете]. - 2 11 2022 г.. - http://cs.brown.edu/~adf/programming\_languages.html.