Logo

Description automatically generated

Факультет: ПИиКТ

Информатика

\_Лабораторная работа\_№4

Вариант: 333879%36 = 15

­

**Студент: Абрабоу Ахмед Елсаид А.И**

**№ группы: P3110**

**Номер студента: 333879**

**Преподаватель: Балакшин Павел Валерьевич**

**Санкт-Петербург**

**2021 г.**

Оглавление

[Задание 2](#_Toc86705323)

[Ход работы 3](#_Toc86705324)

[Обязательное задание 3](#_Toc86705325)

[Дополнительное задание №1 5](#_Toc86705326)

[Дополнительное задание №2 7](#_Toc86705327)

[Дополнительное задание №3 9](#_Toc86705328)

[Дополнительное задание №4 9](#_Toc86705329)

[Вывод 12](#_Toc86705330)

[Список литературы. 12](#_Toc86705331)

# 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.
* Дополнительное задание задание №1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
  + Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
  + Сравнить полученные результаты и объяснить их сходство/различие.
* Дополнительное задание задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + Переписать исходный код, добавив в него использование регулярных выражений.
  + Сравнить полученные результаты и объяснить их сходство/различие.
* Дополнительное задание задание №3 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  + 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.
  + Проанализировать полученные результаты и объяснить их сходство/различие.
* Дополнительное задание задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную.
  + Переписать исходную, чтобы она осуществляла парсинг и конвертацию исходного файла в любой другой формат (кроме JSON, YAML, XML, HTML): PROTOBUF, TSV, CSV, WML и т.п.
  + Проанализировать полученные результаты, объяснить особенности использованного формата.

# Ход работы

Исходный файл Парсинг.json

{"element":

[

{"day": "Ср",

"week": "2 4 6 8 10 12 14 16",

"time": "08:20-09:50",

"room": "306 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ПРОГРАММИРОВАНИЕ(ЛАБ)",

"teacher": "Сорокин Роман Борисович",

"format": "Очно - дистанционный"},

{"day": "Ср",

"week": "2 4 6 8 10 12 14 16",

"time": "10:00-11:30",

"room": "306 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ПРОГРАММИРОВАНИЕ(ЛАБ)",

"teacher": "Сорокин Роман Борисович",

"format": "Очно - дистанционный"},

{"day": "Ср",

"week": "2 4 6 8 10 12 14 16",

"time": "11:40-13:10",

"room": "304 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)",

"teacher": "Ларочкин Глеб Игоревич",

"format": "Очно - дистанционный"},

{"day": "Ср",

"week": "2 4 6 8 10 12 14 16",

"time": "13:30-15:00",

"room": "304 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)",

"teacher": "Ларочкин Глеб Игоревич",

"format": "Очно - дистанционный"},

{"day": "Ср",

"week": "3 5 7 9 11 13 15 17",

"time": "10:00-11:30",

"room": "371 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ИНФОРМАТИКА(ЛАБ)",

"teacher": "Рудникова Тамара Владимировна",

"format": "Очно - дистанционный"},

{"day": "Ср",

"week": "3 5 7 9 11 13 15 17",

"time": "11:40-13:10",

"room": "371 АУД.",

"address": "Кронверкский пр., д.49, лит.А",

"lesson": "ИНФОРМАТИКА(ЛАБ)",

"teacher": "Рудникова Тамара Владимировна",

"format": "Очно - дистанционный"}

]

}

# Обязательное задание

def newLine(result, depth, isList=False):

if not isList:

result += "\n" + " " \* depth

else:

result += "\n" + " " \* (depth - 1) + "- "

return result

def convert():

f = open("Парсинг.json", encoding="utf8")

text = f.read()

f.close()

writing = False

element = True

inList = False

depth = -1

result = ""

for i in text:

if writing:

if i == '"':

writing = not writing

if not element:

result = newLine(result, depth)

else:

result += ": "

element = not element

continue

result += i

else:

if i == '"':

writing = True

elif i == "{":

depth += 1

element = True

if inList:

result = newLine(result, depth, True)

elif i == "}":

depth -= 1

elif i == "[":

inList = True

elif i == "]":

inList = False

f = open("resultWithCode.yml", mode="w", encoding="utf8")

f.write(result)

f.close()

if \_\_name\_\_ == '\_\_main\_\_':

convert()

**Результат:**

Файл :resultWithCode.yml

element:

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 08:20-09:50

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 10:00-11:30

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 11:40-13:10

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 13:30-15:00

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 10:00-11:30

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 11:40-13:10

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

# Дополнительное задание №1

import yaml

import json

def convert():

f = open("Парсинг.json", mode="r", encoding="utf8")

text = json.loads(f.read())

f.close()

text1 = yaml.dump(text)

f1 = open("resultWithLibrary.yml", mode="w", encoding="utf8")

f1.write(text1)

f1.close()

if \_\_name\_\_ == '\_\_main\_\_':

convert()

**Результат:**

element:

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 08:20-09:50

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 10:00-11:30

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 11:40-13:10

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 13:30-15:00

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 10:00-11:30

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 11:40-13:10

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

# Дополнительное задание №2

import re

def newLine(result, depth, isList=False):

if not isList:

result += "\n" + " " \* depth

else:

result += "\n" + " " \* (depth - 1) + "- "

return result

def convert():

f = open("Парсинг.json", encoding="utf8")

text = f.read()

f.close()

writing = False

element = True

inList = False

depth = -1

result = ""

for i in text:

if writing:

if i == '"':

writing = not writing

if not element:

result = newLine(result, depth)

else:

result += ": "

element = not element

continue

result += i

else:

x = re.findall('"', i)

if len(x) > 0:

writing = True

elif i == "{":

depth += 1

element = True

if inList:

result = newLine(result, depth, True)

x1 = re.findall("}", i)

if len(x1) > 0:

depth -= 1

elif i == "[":

inList = True

x3 = re.findall("]", i)

if len(x3) > 0:

inList = False

f = open("resultWithRegulars.yml", mode="w", encoding="utf8")

f.write(result)

f.close()

xr = re.findall(\_\_name\_\_, '\_\_main\_\_')

if len(xr) > 0:

convert()

**Результат:**

Файл точно такой же, как тот который получается в результате выполнения обязательного задания

element:

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 08:20-09:50

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 10:00-11:30

room: 306 АУД.

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

lesson: ПРОГРАММИРОВАНИЕ(ЛАБ)

teacher: Сорокин Роман Борисович

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 11:40-13:10

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 2 4 6 8 10 12 14 16

time: 13:30-15:00

room: 304 АУД.

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

lesson: ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)

teacher: Ларочкин Глеб Игоревич

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 10:00-11:30

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

- day: Ср

week: 3 5 7 9 11 13 15 17

time: 11:40-13:10

room: 371 АУД.

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

lesson: ИНФОРМАТИКА(ЛАБ)

teacher: Рудникова Тамара Владимировна

format: Очно - дистанционный

# Дополнительное задание №3

**Code:**

import time

import taskOneWithCode

import taskTwoWithLibrary

print("Now testing my interpreter")

start\_time = time.time()

for i in range(10):

taskOneWithCode.convert()

print("--- %s seconds ---" % (time.time() - start\_time))

print("Now testing some other library")

start\_time = time.time()

for i in range(10):

taskTwoWithLibrary.convert()

print("--- %s seconds ---" % (time.time() - start\_time))

**Результат:**

"C:\Program Files\Python310\python.exe" "C:/Users/sabur/OneDrive/Рабочий стол/otherThinks/withAhmed/checkTime.py"

Now testing my interpreter

--- 0.01561117172241211 seconds ---

Now testing some other library

--- 0.15629959106445312 seconds ---

Now testing code with regulars

--- 0.06248354911804199 seconds ---

Возможное объяснение результатов:

Дольше всего работает программа (2), тк она производит полноценный рекурсивный парсинг json в yaml, в то время как две остальные программы выбирают только нужные свойства, и парсинг производится в один проход

Прогамма (3) на регулярных выражениях работает дольше программы (1) так как регулярные выражения более затратны из-за проверки исходной строки на соответствие шаблону, а так же из-за использования ленивых квантификаторов. Но так как размер файла невелик, разница несущественная.

Дополнительное задание №4

def convert():

f = open("Парсинг.json", encoding="utf8")

text = f.read()

f.close()

writing = False

element = True

result = {}

realResult = ""

keys = []

longest = 0

elementName = ""

value = ""

for i in text:

if writing:

if i == '"':

writing = not writing

if not element:

result[elementName].append(value)

if len(result[elementName]) > longest:

longest = len(result[elementName])

elementName = ""

else:

if elementName not in result:

result[elementName] = []

keys.append(elementName)

value = ""

element = not element

continue

if not element:

value += i

else:

elementName += i

else:

if i == '"':

writing = True

elif i == "{":

element = True

elementName = ""

realResult = "\t".join(keys) + "\n"

for i in range(longest):

for key in keys:

if i < len(result[key]):

realResult += result[key][i]

realResult += "\t"

realResult += "\n"

f = open("resultWithTsv.tsv", mode="w", encoding="utf8")

f.write(realResult)

f.close()

if \_\_name\_\_ == '\_\_main\_\_':

convert()

**Результат:**

element day week time room address lesson teacher format

Ср 2 4 6 8 10 12 14 16 08:20-09:50 306 АУД. Кронверкский пр., д.49, лит.А ПРОГРАММИРОВАНИЕ(ЛАБ) Сорокин Роман Борисович Очно - дистанционный

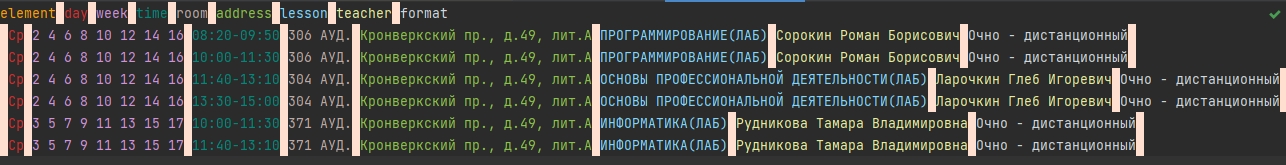
Ср 2 4 6 8 10 12 14 16 10:00-11:30 306 АУД. Кронверкский пр., д.49, лит.А ПРОГРАММИРОВАНИЕ(ЛАБ) Сорокин Роман Борисович Очно - дистанционный

Ср 2 4 6 8 10 12 14 16 11:40-13:10 304 АУД. Кронверкский пр., д.49, лит.А ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ) Ларочкин Глеб Игоревич Очно - дистанционный

Ср 2 4 6 8 10 12 14 16 13:30-15:00 304 АУД. Кронверкский пр., д.49, лит.А ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ) Ларочкин Глеб Игоревич Очно - дистанционный

Ср 3 5 7 9 11 13 15 17 10:00-11:30 371 АУД. Кронверкский пр., д.49, лит.А ИНФОРМАТИКА(ЛАБ) Рудникова Тамара Владимировна Очно - дистанционный

Ср 3 5 7 9 11 13 15 17 11:40-13:10 371 АУД. Кронверкский пр., д.49, лит.А ИНФОРМАТИКА(ЛАБ) Рудникова Тамара Владимировна Очно - дистанционный



# Вывод

Во время выполнения лабораторной работы я узнал о языках разметки YAML, JSON и PSV, научился с ними работать и переводить один тип в другой с помощью средств языка Python. Научился работать с библиотекой time для измерения времени, и некоторыми библиотеками для автоматического парсинга.

# Список литературы.

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. - экз. - Режим доступа:

<https://books.ifmo.ru/book/2248/informatika:_laboratornye_raboty_i_testy:_uchebno-metodicheskoe_posobie_/_recenzent:_polyakov_v.i..htm>

1. Грошев А.С. Г89 Информатика: Учебник для вузов / А.С. Грошев. – Архангельск, Арханг. гос. техн. ун-т, 2010. -470с. -Режим доступа <https://narfu.ru/university/library/books/0690.pdf>