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 №4

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

По дисциплине

*Информатика*

Вариант №15

Выполнил студент группы P3112:

Степутенко Илья Сергеевич

Преподаватель:

Белозубов Александр Владимирович

**Оглавление**

[Вариант заданий 3](#_Toc118201140)

[Содержание файла schedule.yml 4](#_Toc118201141)

[Задание №1 5](#_Toc118201142)

[Выполнение задания №1 5](#_Toc118201143)

[Содержание файла schedule\_1.json 5](#_Toc118201144)

[Задание №2 6](#_Toc118201145)

[Выполнение задания №2 6](#_Toc118201146)

[Содержание файла schedule\_2.json 7](#_Toc118201147)

[Задание №3 8](#_Toc118201148)

[Выполнение задания №3 8](#_Toc118201149)

[Содержание файла schedule\_3.json 9](#_Toc118201150)

[Задание №4 10](#_Toc118201151)

[Выполнение задания №4 10](#_Toc118201152)

[Пример вывода программы из задания №4 10](#_Toc118201153)

[Задание №5 11](#_Toc118201154)

[Выполнение задания №5 11](#_Toc118201155)

[Содержание файла schedule\_5.csv 12](#_Toc118201156)

[Заключение 13](#_Toc118201157)

Вариант заданий

|  |  |  |
| --- | --- | --- |
| Исходный формат | Результирующий формат | День недели |
| YAML | JSON | Среда |

Содержание файла schedule.yml

day: Среда

group: P3112

lessons:

-

time: 11:40-13:10

weeks: 3, 5, 7, 9, 11, 13, 15, 17

room: 2308

subject: ИНФОРМАТИКА(ЛАБ)

teacher: Белозубов Александр Владимирович

format: Очно-дистанционный

-

time: 13:30-15:00

weeks: 3, 5, 7, 9, 11, 13, 15, 17

room: 2308

subject: ИНФОРМАТИКА(ЛАБ)

teacher: Белозубов Александр Владимирович

format: Очно-дистанционный

Задание №1

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

Выполнение задания №1

def yml\_parser(file: str) -> None:

with open(file) as inputfile:

d = {}

lkey = ""

while (s := inputfile.readline()):

s = s.strip()

if not s: continue

if s[-1] == ":":

lkey = s[:-1]

d[lkey.strip()] = []

elif s[0] == "-":

try:

d[lkey] += [{}]

except:

print("Неправильный формат файла"); quit()

s = s[1:].strip()

if ":" in s:

key, \*string = s.split(":")

d[lkey][-1][key.strip()] = ":".join(string).strip()

elif lkey:

key, \*string = s.split(":")

try: d[lkey.strip()][-1][key] = ":".join(string).strip()

except: print("Неправильный формат файла - не задана группа значений."); quit()

elif ":" in s:

key, \*string = s.split(":")

d[key.strip()] = ":".join(string).strip()

else:

print("Неправильный формат файла"); quit()

with open(f"{file.split('.')[0]}\_1.json", "w+") as output:

output.write(str(d).replace("'", '"'))

yml\_parser("schedule.yml")

Содержание файла schedule\_1.json

{"day": "Среда", "group": "P3112", "lessons": [{"time": "11:40-13:10", "weeks": "3, 5, 7, 9, 11, 13, 15, 17", "room": "2308", "subject": "ИНФОРМАТИКА(ЛАБ)", "teacher": "Белозубов Александр Владимирович", "format": "Очно-дистанционный"}, {"time": "13:30-15:00", "weeks": "3, 5, 7, 9, 11, 13, 15, 17", "room": "2308", "subject": "ИНФОРМАТИКА(ЛАБ)", "teacher": "Белозубов Александр Владимирович", "format": "Очно-дистанционный"}]}

Задание №2

1. Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
2. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
3. Сравнить полученные результаты и объяснить их сходство/различие.

Выполнение задания №2

import yaml, json

def yml\_parser(file: str) -> None:

with open(file, 'r') as inputstream:

try:

tempdict = yaml.safe\_load(inputstream)

except yaml.YAMLError as exc:

tempdict = {}

with open(f"{file.split('.')[0]}\_2.json", "w", encoding="utf-8") as output:

json.dump(tempdict, output, ensure\_ascii=False, indent=4)

yml\_parser("schedule.yml")

Различия:

1. Код больше не содержит циклы.
2. Исходный код не индентирует JSON-файл.
3. Исходный код не имеет зависимостей, так как не подключает дополнительные библиотеки.

Сходства:

1. Как исходный, так и переписанный код успешно выполняют конвертацию.
2. Оба кода требуют открытия файла.

Содержание файла schedule\_2.json

{

"day": "Среда",

"group": "P3112",

"lessons": [

{

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

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

"room": 2308,

"subject": "ИНФОРМАТИКА(ЛАБ)",

"teacher": "Белозубов Александр Владимирович",

"format": "Очно-дистанционный"

},

{

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

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

"room": 2308,

"subject": "ИНФОРМАТИКА(ЛАБ)",

"teacher": "Белозубов Александр Владимирович",

"format": "Очно-дистанционный"

}

]

}

Задание №3

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

Выполнение задания №3

import re

def yml\_parser(file: str) -> None:

with open(file) as inputfile:

d = {}

lkey = ""

while (s := inputfile.readline()):

s = s.strip()

if not s: continue

if re.match(r".\*:$", s):

lkey = s[:-1]

d[lkey.strip()] = []

elif re.match(r"^-.\*", s):

try:

d[lkey] += [{}]

except:

print("Неправильный формат файла"); quit()

if re.match(r"^-.\*:.\*", s):

key, \*string = s[1:].strip().split(":")

d[lkey][-1][key.strip()] = ":".join(string).strip()

elif lkey:

key, \*string = re.findall(r"^(\w\*\s\*):\s\*(.\*)", s)[0]

try: d[lkey.strip()][-1][key] = ":".join(string).strip()

except: print("Неправильный формат файла - не задана группа значений."); quit()

elif ":" in s:

key, \*string = s.split(":")

d[key.strip()] = ":".join(string).strip()

else:

print("Неправильный формат файла"); quit()

with open(f"{file.split('.')[0]}\_3.json", "w+") as output:

output.write(str(d).replace("'", '"'))

yml\_parser("schedule.yml")

Различия и сходства:

1. В отличие от кода из задания №2 и подобно коду из задания №1, код из задания №3 не индентирует JSON-файл.
2. В отличие от кода из задания №1 и подобно коду из задания №2, код из этого задания содержит зависимости в связи с использованием дополнительных библиотек.

Содержание файла schedule\_3.json

{"day": "Среда", "group": "P3112", "lessons": [{"time": "11:40-13:10", "weeks": "3, 5, 7, 9, 11, 13, 15, 17", "room": "2308", "subject": "ИНФОРМАТИКА(ЛАБ)", "teacher": "Белозубов Александр Владимирович", "format": "Очно-дистанционный"}, {"time": "13:30-15:00", "weeks": "3, 5, 7, 9, 11, 13, 15, 17", "room": "2308", "subject": "ИНФОРМАТИКА(ЛАБ)", "teacher": "Белозубов Александр Владимирович", "format": "Очно-дистанционный"}]}

Задание №4

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

Выполнение задания №4

from task1 import yml\_parser as parser\_1

from task2 import yml\_parser as parser\_2

from task3 import yml\_parser as parser\_3

from timeit import timeit

class Tests:

f = "schedule.yml"

def test\_1():

parser\_1(Tests.f)

def test\_2():

parser\_2(Tests.f)

def test\_3():

parser\_3(Tests.f)

print(timeit(Tests.test\_1, number=100))

print(timeit(Tests.test\_2, number=100))

print(timeit(Tests.test\_3, number=100))

Пример вывода программы из задания №4

0.03727060299999607

0.19510987600000362

0.03607519900000966

Подобный вывод связан с тем, что регулярные выражения обрабатывают строки на более низком уровне выполнения программы, чем обработка напрямую в Python. Кроме того, подключение дополнительных тяжелых библиотек вроде yaml и json так же замедляет время выполнения программы.

Задание №5

1. 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.
2. Проанализировать полученные результаты, объяснить особенности использования формата.

Выполнение задания №5

Выбранный формат: CSV.

import csv

def yml\_parser(file: str) -> None:

with open(file) as inputfile:

d = {}

lkey = ""

while (s := inputfile.readline()):

s = s.strip()

if not s: continue

if s[-1] == ":":

lkey = s[:-1]

d[lkey.strip()] = []

elif s[0] == "-":

try:

d[lkey] += [{}]

except:

print("Неправильный формат файла"); quit()

if ":" in s:

key, \*string = s[1:].strip().split(":")

d[lkey][-1][key.strip()] = ":".join(string).strip()

elif lkey:

key, \*string = s.split(":")

try: d[lkey.strip()][-1][key] = ":".join(string).strip()

except: print("Неправильный формат файла - не задана группа значений."); quit()

elif ":" in s:

key, \*string = s.split(":")

d[key.strip()] = ":".join(string).strip()

else:

print("Неправильный формат файла"); quit()

headstr = ",".join(key for key in d)

headstr += "\n"

headstr += ",".join(d[key] for key in d if isinstance(d[key], str))

lstr = ""

c = 0

for lesson in d["lessons"]:

c += 1

lstr += f"\nlesson №{c}\n"

for char in lesson:

lstr += f",,{char},{lesson[char].replace(', ', '; ')}\n"

with open('schedule\_5.csv', 'w') as csvstream:

csvstream.writelines([headstr,lstr])

yml\_parser("schedule.yml")

Содержание файла schedule\_5.csv

day,group,lessons

Среда,P3112

lesson №1

,,time,11:40-13:10

,,weeks,3; 5; 7; 9; 11; 13; 15; 17

,,room,2308

,,subject,ИНФОРМАТИКА(ЛАБ)

,,teacher,Белозубов Александр Владимирович

,,format,Очно-дистанционный

lesson №2

,,time,13:30-15:00

,,weeks,3; 5; 7; 9; 11; 13; 15; 17

,,room,2308

,,subject,ИНФОРМАТИКА(ЛАБ)

,,teacher,Белозубов Александр Владимирович

,,format,Очно-дистанционный

Заключение

Выполняя лабораторную работу, я разобрался в принципах работы с файлами различных форматов, научился парсить и конвертировать файлы, освоил разные способы взаимодействия с ними.