

Федеральное государственное автономное образовательное учреждение высшего  
образования  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»  
(ФПИиКТ)

ОТЧЕТ ПО ЛАБАРОТОРНОЙ РАБОТЕ №4:  
“ИССЛЕДОВАНИЕ ПРОТОКОЛОВ, ФОРМАТОВ ОБМЕНА ИНФОРМАЦИЕЙ И  
ЯЗЫКОВ РАЗМЕТКИ ДОКУМЕНТОВ”

Вариант 2

Выполнил:

Студент группы Р3115

Зыков Иван Евгеньевич

Проверил:

Авксентьева Елена Юрьевна,

к.п.н., доцент факультета ПИиКТ

Санкт-Петербург, г. 2023

## Оглавление

|  |    |
|--|----|
| Оглавление .....                             | 1  |
| Задание .....                                | 2  |
| Выполнение .....                             | 4  |
| Обязательное задание .....                   | 5  |
| Расписание на JSON .....                     | 5  |
| Код программы .....                          | 6  |
| Результат работы .....                       | 8  |
| Дополнительное задание №1 .....              | 9  |
| Код программы .....                          | 10 |
| Результат работы .....                       | 10 |
| Дополнительное задание №2 .....              | 11 |
| Код программы .....                          | 12 |
| Результат работы .....                       | 13 |
| Дополнительное задание №3 .....              | 14 |
| Код программы .....                          | 15 |
| Результат работы с моим расписанием.....     | 16 |
| Результат работы со вторым расписанием ..... | 17 |
| Дополнительное задание №4 .....              | 19 |
| Вывод: .....                                 | 20 |

## Задание

Обязательное задание (позволяет набрать до 45 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 путём простой замены метасимволов исходного формата на метасимволы результирующего формата.

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).

- а) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
- б) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
- с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).

- а) Переписать исходный код, добавив в него использование регулярных выражений.
- б) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №3 (позволяет набрать +25 процентов от максимального числа баллов БаРС за данную лабораторную).

- а) Переписать исходный код таким образом, чтобы для решения задачи использовались формальные грамматики. То есть ваш код должен уметь осуществлять парсинг и конвертацию любых данных, представленных в исходном формате, в данные, представленные в результирующем формате: как с готовыми библиотеками из дополнительного задания №1.
- б) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
- с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

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

а) Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.

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

Объяснение должно быть отражено в отчёте.

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

а) 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.

б) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

## Выполнение

### Обязательное задание

#### Расписание на JSON

```
{
  "group_number": "P3115",
  "id": "408674",
  "day": "Monday",
  "schedule": [
    {
      "lesson": "Английский язык",
      "lesson_format": "Очно-дистанционный",
      "type": "Практика",
      "time": "11:40-13:10",
      "room": {
        "build": "ул.Ломоносова, д.9, лит. Е",
        "name": "ауд. 3301"
      },
      "teacher": "Левшинова Екатерина Андреевна"
    },
    {
      "lesson": "Английский язык",
      "lesson_format": "Очно-дистанционный",
      "type": "Практика",
      "time": "13:30-15:00",
      "room": {
        "build": "ул.Ломоносова, д.9, лит. Е",
        "name": "ауд. 3301"
      },
      "teacher": "Левшинова Екатерина Андреевна"
    },
    {
      "lesson": "Математический анализ (продвинутый уровень)",
      "lesson_format": "Очно-дистанционный",
```

```
"type": "Лекция",  
"time": "17:00-18:30",  
"room": {  
  "build": "Кронверкский пр., д.49, лит.А",  
  "name": "Ауд. Orange Classroom (1229)"  
},  
"teacher": "Трифанова Екатерина Станиславовна"  
}  
]  
}
```

### Код программы

```
fJSON = open('lab4JSON.json')
fYAML = open('lab4YAML.yaml', 'w')

lvlmas = False
countBracket = 0
countLine = 0
for line in fJSON:
    if line.find('[') != -1:
        line = line.replace('[', '"')
        lvlmas = True
    if line.find('{') != -1 and lvlmas == True and countBracket == 0:
        countBracket = 1
        continue
    if countLine == 0 and countBracket == 1:
        index = line.find('"')
        line = line[:index - 2] + '-' + line[index - 1:]
        countLine = 1
    if lvlmas == True and line.find('{') != -1:
        countBracket += 1
    if lvlmas == True and line.find('}') != -1:
        countBracket -= 1
        if countBracket == 0:
            countLine = 0
    if line.find(']') != -1:
        lvlmas = False
        continue
    if line.find('}') != -1:
        continue
    if line.find('{') != -1:
        if line.find(':') != -1:
            line = line.replace('{', '"')
        else:
            continue
```

```
if line.rfind(',') == len(line) - 2:
    line = line[:len(line) - 2] + line[-1:]
line = line[2:]
if line[0] == line[1] == ' ':
    line = line[2:]
line = line.replace("'", "")
fYAML.write(line)
fYAML.close()
```

### Результат работы

```
group_number: P3115
id: 408674
day: Monday
schedule:
- lesson: Английский язык
  lesson_format: Очно-дистанционный
  type: Практика
  time: 11:40-13:10
  room:
    build: ул.Ломоносова, д.9, лит. Е
    name: ауд. 3301
  teacher: Левшинова Екатерина Андреевна
- lesson: Английский язык
  lesson_format: Очно-дистанционный
  type: Практика
  time: 13:30-15:00
  room:
    build: ул.Ломоносова, д.9, лит. Е
    name: ауд. 3301
  teacher: Левшинова Екатерина Андреевна
- lesson: Математический анализ (продвинутый уровень)
  lesson_format: Очно-дистанционный
  type: Лекция
```



time: 17:00-18:30

room:

build: Кронверкский пр., д.49, лит.А

name: Ауд. Orange Classroom (1229)

teacher: Трифанова Екатерина Станиславовна

## Дополнительное задание №1

### Код программы

```
import yaml

fJSON = open('lab4JSON.json')
fYAML = open('lab4YAML.yaml', 'w')

parsed_json = json.load(fJSON)
yaml.dump(parsed_json, fYAML, allow_unicode=True)
fYAML.close()
```

### Результат работы

```
day: Monday
group_number: P3115
id: '408674'
schedule:
- lesson: Английский язык
  lesson_format: Очно-дистанционный
  room:
    build: ул.Ломоносова, д.9, лит. Е
    name: ауд. 3301
  teacher: Левшинова Екатерина Андреевна
  time: 11:40-13:10
  type: Практика
- lesson: Английский язык
  lesson_format: Очно-дистанционный
  room:
    build: ул.Ломоносова, д.9, лит. Е
    name: ауд. 3301
  teacher: Левшинова Екатерина Андреевна
  time: 13:30-15:00
  type: Практика
- lesson: Математический анализ (продвинутый уровень)
  lesson_format: Очно-дистанционный
```

room:

build: Кронверкский пр., д.49, лит.А

name: Ауд. Orange Classroom (1229)

teacher: Трифанова Екатерина Станиславовна

time: 17:00-18:30

type: Лекция

## Дополнительное задание №2

### Код программы

```
import re

fJSON = open('lab4JSON.json')
fYAML = open('lab4YAML.yaml', 'w')

lvlmas = False
countBracket = 0
countLine = 0
for line in fJSON:
    if re.match(r".*\[", line):
        line = line.replace('[', ' ')
        lvlmas = True
    if re.match(r".*\{", line) and lvlmas == True and countBracket == 0:
        countBracket = 1
        continue
    if countLine == 0 and countBracket == 1:
        index = line.find('"')
        line = line[:index - 2] + '-' + line[index - 1:]
        countLine = 1
    if lvlmas == True and re.match(r".*\{", line) :
        countBracket += 1
    if lvlmas == True and re.match(r".*\}", line) :
        countBracket -= 1
        if countBracket == 0:
            countLine = 0
    if re.match(r".*\]", line) :
        lvlmas = False
        continue
    if re.match(r".*\},?", line) or re.match(r"\s*\{", line):
        continue
    if re.match(r".*:\s*", line):
```

```
line = line.replace('{', '')
if re.match(r".*,$", line):
    line = line[:len(line) - 2] + line[-1:]
line = line[2:]
if line[0] == line[1] == ' ':
    line = line[2:]
line = line.replace("'", "")
fYAML.write(line)
fYAML.close()
```

#### Результат работы

group\_number: P3115

id: 408674

day: Monday

schedule:

- lesson: Английский язык

lesson\_format: Очно-дистанционный

type: Практика

time: 11:40-13:10

room:

build: ул.Ломоносова, д.9, лит. Е

name: ауд. 3301

teacher: Левшинова Екатерина Андреевна

- lesson: Английский язык

lesson\_format: Очно-дистанционный

type: Практика

time: 13:30-15:00

room:

build: ул.Ломоносова, д.9, лит. Е

name: ауд. 3301

teacher: Левшинова Екатерина Андреевна

- lesson: Математический анализ (продвинутый уровень)

lesson\_format: Очно-дистанционный

type: Лекция

time: 17:00-18:30

room:

build: Кронверкский пр., д.49, лит.А

name: Ауд. Orange Classroom (1229)

teacher: Трифанова Екатерина Станиславовна

### Дополнительное задание №3

Код программы

```
import re

fJSON = open('lab4JSON.json')
fYAML = open('lab4YAML.yaml', 'w')

lvlmas = [False] * 1000000
countBracket = [0] * 100000
countLine = [0] * 100000
index = -1
for line in fJSON:
    if re.match(r".*{", line) and lvlmas[index] == True and countBracket[index] == 0:
        countBracket[index] = 1
        continue
    if countLine[index] == 0 and countBracket[index] == 1:
        jindex = line.find("")
        line = line[:jindex - 2] + '-' + line[jindex - 1:]
        countLine[index] = 1
    if lvlmas[index] == True and re.match(r".*{", line):
        countBracket[index] += 1
    if lvlmas[index] == True and re.match(r".*}", line):
        countBracket[index] -= 1
        if countBracket[index] == 0:
            countLine[index] = 0
    if re.match(r".*]", line):
        lvlmas[index] = False
        index -= 1
        continue
    if re.match(r".*\[", line):
        line = line.replace('[', '"')
        index += 1
        lvlmas[index]=True
```

```

if re.match(r".*},?", line) or re.match(r"\s*{ ", line):
    continue
if re.match(r".*:{ ", line):
    line = line.replace('{', ')')
if re.match(r".*,$", line):
    line = line[:len(line) - 2] + line[-1:]
line = line[2:]
if not index == -1:
    if countBracket[index] == 0:
        if not index == 0:
            line = line[(2 * index):]
        if not countBracket[index] == 0:
            line = line[(2*(index+1)):]
    line = line.replace("'", "")
fYAML.write(line)
fYAML.close()

```

#### Результат работы с моим расписанием

group\_number: P3115

id: 408674

day: Monday

schedule:

- lesson: Английский язык

lesson\_format: Очно-дистанционный

type: Практика

time: 11:40-13:10

room:

build: ул.Ломоносова, д.9, лит. Е

name: ауд. 3301

teacher: Левшинова Екатерина Андреевна

- lesson: Английский язык

lesson\_format: Очно-дистанционный

type: Практика

time: 13:30-15:00



room:

build: ул.Ломоносова, д.9, лит. Е

name: ауд. 3301

teacher: Левшинова Екатерина Андреевна

- lesson: Математический анализ (продвинутый уровень)

lesson\_format: Очно-дистанционный

type: Лекция

time: 17:00-18:30

room:

build: Кронверкский пр., д.49, лит.А

name: Ауд. Orange Classroom (1229)

teacher: Трифанова Екатерина Станиславовна

Результат работы со вторым расписанием

JSON:

```
{
  "first": {
    "type": "Лекция",
    "time": "10:00-11:30",
    "sub": "Программирование",
    "lec": "Письмак Алексей Евгеньевич",
    "aud": "1216",
    "corp": "ул.Ломоносова, д.9, лит. М",
    "homework": [
      {
        "theory": "ООП",
        "test": "new theory"
      },
      {
        "labs": [
          {
            "number": "3",
            "deadline": "end of november"
          }
        ]
      }
    ]
  }
}
```

```
{
  "number": "4",
  "deadline": "end of december"
}
]
}
]
},
"second": {
  "type": "Лабораторная",
  "time": "13:30-15:00",
  "sub": "Информатика",
  "lec": "Рыбаков Степан Дмитриевич",
  "aud": "2112",
  "corp": "Кронверский пр., д.49, лит. А"
},
"third": {
  "type": "Лабораторная",
  "time": "15:20-16:50",
  "sub": "Информатика",
  "lec": "Рыбаков Степан Дмитриевич",
  "aud": "2112",
  "corp": "Кронверский пр., д.49, лит. А"
}
}
```

YAML:

first:

type: Лекция

time: 10:00-11:30

sub: Программирование

lec: Письмак Алексей Евгеньевич

aud: 1216

corp: ул.Ломоносова, д.9, лит. М

homework:

- theory: OOP

test: new theory

- labs:

- number: 3

deadline: end of november

- number: 4

deadline: end of december

second:

type: Лабораторная

time: 13:30-15:00

sub: Информатика

lec: Рыбаков Степан Дмитриевич

aud: 2112

corp: Кронверский пр., д.49, лит. А

third:

type: Лабораторная

time: 15:20-16:50

sub: Информатика

lec: Рыбаков Степан Дмитриевич

aud: 2112

corp: Кронверский пр., д.49, лит. А

#### Дополнительное задание №4

| Задание           | Время работы       |
|-------------------|--------------------|
| Обязательное      | 0.0002169609069824 |
| Дополнительное №1 | 0.0015954971313476 |
| Дополнительное №2 | 0.0008392333984375 |
| Дополнительное №3 | 0.0034472942352294 |

Проанализировав таблицу, мы видим, что дольше всего работает третье дополнительное задание, в котором используются регулярные выражения и вложенные циклы. Дополнительное задание №1 так же долго работает, так как там есть регулярные выражения. Быстрее всего работает обязательное задание, в котором не используются регулярные выражения, только базовые функции. Примерно так же быстро работает задание выполненное с использованием готовых библиотек. В итоге мы можем прийти к выводу, что регулярные хоть и упрощают работу с выражениями, но из-за этого мы теряем в скорости.

## Вывод:

В ходе выполнения лабораторной работы я научился пользоваться языками разметки и написал конвертатор между ними, закрепил использование регулярных выражений.