

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 по курсу "Информатика"  
"Исследование протоколов, форматов обмена  
информацией и языков разметки документов"

Группа: **P3131**  
Студент: Друян Эдуард Андреевич  
Вариант: **10**

Санкт-Петербург  
2021 г.

# Цель работы

Исследование протоколов, форматов обмена информацией и языков разметки документов.

## Текст заданий

### Обязательное задание

Исходя из структуры расписания **вторника**, сформировать файл с расписанием в формате **yaml**.

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

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

**src.yaml**

```
1.  schedule:
2.    Вт:
3.      - time: 11:40-13:10
4.        evenWeek: false
5.        auditorium: 306
6.        building: Кронверкский пр., д.49, лит.А
7.        subject: ИНФОРМАТИКА(ЛАБ)
8.        teacher: Белозубов Александр Владимирович
9.        lessonType: Очно – дистанционный
10.     - time: 11:40-13:10
11.       evenWeek: true
12.       auditorium: 285
13.       building: Кронверкский пр., д.49, лит.А
14.       subject: ИНФОРМАТИКА(ЛЕК)
15.       teacher: Белозубов Александр Владимирович
16.       lessonType: Очно – дистанционный
17.     - time: 11:40-13:10
18.       evenWeek: false
19.       auditorium: 306
20.       building: Кронверкский пр., д.49, лит.А
21.       subject: ИНФОРМАТИКА(ЛАБ)
22.       teacher: Белозубов Александр Владимирович
23.       lessonType: Очно – дистанционный
24.     - time: 13:30-15:00
25.       evenWeek: true
26.       auditorium: 305
27.       building: Кронверкский пр., д.49, лит.А
28.       subject: ПРОГРАММИРОВАНИЕ(ЛАБ)
29.       teacher: Блохина Елена Николаевна
30.       lessonType: Очно – дистанционный
31.     - time: 15:20-16:50
32.       evenWeek: true
33.       auditorium: 305
34.       building: Кронверкский пр., д.49, лит.А
35.       subject: ПРОГРАММИРОВАНИЕ(ЛАБ)
36.       teacher: Блохина Елена Николаевна
37.       lessonType: Очно – дистанционный
```

## lab4\_0.py

```
1. import time
2.
3. def isInt(s: str):
4.     toCheck = True
5.     try:
6.         int(s)
7.     except ValueError:
8.         toCheck = False
9.     return toCheck
10. def isBool(s: str):
11.     if s=="true" or s=="True" or s=="false" or s=="False":
12.         return True
13.     return False
14. def parseVal(s: str):
15.     i = 0
16.     while i < len(s) and s[i] != ':':
17.         i += 1
18.     i += 2
19.     if i < len(s):
20.         val = s[i:]
21.         if isInt(val) or isBool(val):
22.             return val
23.         return '""' + val + '""'
24.     return ""
25. def parseName(s: str):
26.     i = 0
27.     while i < len(s) and s[i] != ':':
28.         i += 1
29.     if s[:2] == "- ":
30.         return '""' + s[2:i] + '""'
31.     return '""' + s[:i] + '""'
32. def countSpaces(s: str):
33.     cnt = 0
34.     while cnt < len(s) and s[cnt] == ' ':
35.         cnt += 1
36.     return cnt
```

```

37. def main():
38.     srcPath = "/Users/ed/Documents/itmo/linfa/lab4/src.yaml"
39.     dstPath = "/Users/ed/Documents/itmo/linfa/lab4/dst_0.json"
40.     # open source file
41.     with open(srcPath, mode='r', encoding='utf-8') as rFile:
42.         # open the file to write
43.         with open(dstPath, mode='w', encoding='utf-8') as wFile:
44.             lines = []
45.             spaces = []
46.             for line in rFile:
47.                 spaces += [countSpaces(line)]
48.                 lines += [line.strip()]
49.
50.             wFile.write("{\""+parseName(lines[0])+"\":")
51.             for i in range(1, len(lines)):
52.                 name = parseName(lines[i])
53.                 val = parseVal(lines[i])
54.                 if spaces[i] == 2: # 2 spaces
55.                     if lines[i][0]=='-' and spaces[i-1]==2: # open list
56.                         wFile.write("[\n    {"+name+": "+val)
57.                     elif lines[i][0]=='-': # continue list
58.                         wFile.write("},\n    {"+name+": "+val)
59.                     else: # day
60.                         wFile.write("\n    {"+name+":")
61.                 else: # 4 spaces
62.                     wFile.write(",\n    "+name+": "+val)
63.             wFile.write("}]\n    }\n}")
64.
65. if __name__ == "__main__":
66.     dstPath = "/Users/ed/Documents/itmo/linfa/lab4/lab4_3.txt"
67.     with open(dstPath, mode='w', encoding='utf=8') as wFile:
68.         start_time = time.time()
69.         main()
70.         d = (time.time() - start_time) * 1000
71.         wFile.write("lab4_0.py: %s (seconds*1000)\n" % d)
72.

```

## dst\_0.json

```
1.  {"shedule":  
2.    {"Br":[  
3.      {"time": "11:40-13:10",  
4.        "evenWeek": false,  
5.        "auditorium": 306,  
6.        "building": "Кронверкский пр., д.49, лит.А",  
7.        "subject": "ИНФОРМАТИКА(ЛАБ)",  
8.        "teacher": "Белозубов Александр Владимирович",  
9.        "lessonType": "Очно – дистанционный"},  
10.     {"time": "11:40-13:10",  
11.       "evenWeek": true,  
12.       "auditorium": 285,  
13.       "building": "Кронверкский пр., д.49, лит.А",  
14.       "subject": "ИНФОРМАТИКА(ЛЕК)",  
15.       "teacher": "Белозубов Александр Владимирович",  
16.       "lessonType": "Очно – дистанционный"},  
17.     {"time": "11:40-13:10",  
18.       "evenWeek": false,  
19.       "auditorium": 306,  
20.       "building": "Кронверкский пр., д.49, лит.А",  
21.       "subject": "ИНФОРМАТИКА(ЛАБ)",  
22.       "teacher": "Белозубов Александр Владимирович",  
23.       "lessonType": "Очно – дистанционный"},  
24.     {"time": "13:30-15:00",  
25.       "evenWeek": true,  
26.       "auditorium": 305,  
27.       "building": "Кронверкский пр., д.49, лит.А",  
28.       "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",  
29.       "teacher": "Блохина Елена Николаевна",  
30.       "lessonType": "Очно – дистанционный"},  
31.     {"time": "15:20-16:50",  
32.       "evenWeek": true,  
33.       "auditorium": 305,  
34.       "building": "Кронверкский пр., д.49, лит.А",  
35.       "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",  
36.       "teacher": "Блохина Елена Николаевна",  
37.       "lessonType": "Очно – дистанционный"}]  
38.   }  
39. }
```

## Дополнительное задание №1

- 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.

## lab4\_1.py

```
1. import time
2. import yaml
3. import json
4.
5. def main():
6.     srcPath = "/Users/ed/Documents/itmo/linfa/lab4/src.yaml"
7.     dstPath = "/Users/ed/Documents/itmo/linfa/lab4/dst_1.json"
8.     # open source file
9.     with open(srcPath, mode='r', encoding='utf-8') as rFile:
10.         # open the file to write
11.         with open(dstPath, mode='w', encoding='utf-8') as wFile:
12.             y = yaml.safe_load(rFile)
13.             json.dump(y, wFile, ensure_ascii=False)
14.
15. if __name__ == "__main__":
16.     dstPath = "/Users/ed/Documents/itmo/linfa/lab4/lab4_3.txt"
17.     with open(dstPath, mode='a', encoding='utf=8') as wFile:
18.         start_time = time.time()
19.         main()
20.         d = (time.time() - start_time) * 1000
21.         wFile.write("lab4_1.py: %s (seconds*1000)\n" % d)
```

## dst\_1.json

```
1. {"shedule": {"Br": [{"time": "11:40-13:10",
2.     "evenWeek": false, "auditorium": 306,
3.     "building": "Кронверкский пр., д.49, лит.А",
4.     "subject": "ИНФОРМАТИКА(ЛАБ)",
5.     "teacher": "Белозубов Александр Владимирович",
6.     "lessonType": "Очно - дистанционный"},
7.     {"time": "11:40-13:10",
8.     "evenWeek": true, "auditorium": 285,
9.     "building": "Кронверкский пр., д.49, лит.А",
10.    "subject": "ИНФОРМАТИКА(ЛЕК)",
11.    "teacher": "Белозубов Александр Владимирович",
12.    "lessonType": "Очно - дистанционный"},
13.    {"time": "11:40-13:10",
14.    "evenWeek": false, "auditorium": 306,
15.    "building": "Кронверкский пр., д.49, лит.А",
16.    "subject": "ИНФОРМАТИКА(ЛАБ)",
17.    "teacher": "Белозубов Александр Владимирович",
18.    "lessonType": "Очно - дистанционный"},
19.    {"time": "13:30-15:00",
20.    "evenWeek": true, "auditorium": 305,
21.    "building": "Кронверкский пр., д.49, лит.А",
22.    "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",
23.    "teacher": "Блохина Елена Николаевна",
24.    "lessonType": "Очно - дистанционный"},
25.    {"time": "15:20-16:50",
26.    "evenWeek": true, "auditorium": 305,
27.    "building": "Кронверкский пр., д.49, лит.А",
28.    "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",
29.    "teacher": "Блохина Елена Николаевна",
30.    "lessonType": "Очно - дистанционный"}]}
31. }}
```

---

## Сравнение lab4\_1.py и lab4\_0.py:

Процесс написания своего алгоритма парсинга файла расписания конкретного дня дольше и требует значительно большего количества строк кода, чем использование готовых решений, таких как библиотеки `yaml` и `json`.

## Дополнительное задание №2

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

### lab4\_2.py

```
1.  import time
2.  import re
3.
4.  def isInt(s: str):
5.      if re.match(r'[0-9]+$', s):
6.          return True
7.      return False
8.  def isBool(s: str):
9.      if re.match(r'([Tt]rue)|([Ff]alse)', s):
10.         return True
11.         return False
12.  def parseVal(s: str):
13.      if s[len(s)-1] == ':':
14.          return ""
15.      i = re.search(r':', s).start() + 2
16.      val = s[i:]
17.      if isInt(val) or isBool(val):
18.          return val
19.      return "" + val + ""
20.  def parseName(s: str):
21.      res = re.split(r':', s)
22.      if re.match(r'- ', res[0]):
23.          return "" + res[0][2:] + ""
24.      else:
25.          return "" + res[0] + ""
26.  def countSpaces(s: str):
27.      cnt = 0
28.      while cnt < len(s) and s[cnt] == ' ':
29.          cnt += 1
30.      return cnt
31.
```

## lab4\_2.py

```
32. def main():
33.     srcPath = "/Users/ed/Documents/itmo/1infa/lab4/src.yaml"
34.     dstPath = "/Users/ed/Documents/itmo/1infa/lab4/dst_2.json"
35.     # open source file
36.     with open(srcPath, mode='r', encoding='utf-8') as rFile:
37.         # open the file to write
38.         with open(dstPath, mode='w', encoding='utf-8') as wFile:
39.             lines = []
40.             spaces = []
41.             for line in rFile:
42.                 spaces += [countSpaces(line)]
43.                 lines += [line.strip()]
44.
45.             wFile.write("{"+parseName(lines[0])+":")
46.             for i in range(1, len(lines)):
47.                 name = parseName(lines[i])
48.                 val = parseVal(lines[i])
49.                 if spaces[i] == 2: # 2 spaces
50.                     if lines[i][0]=='-' and spaces[i-1]==2: # open list
51.                         wFile.write("\n    {"+name+": "+val)
52.                     elif lines[i][0]=='-': # continue list
53.                         wFile.write("},\n    {"+name+": "+val)
54.                     else: # day
55.                         wFile.write("\n    {"+name+":")
56.                 else: # 4 spaces
57.                     wFile.write(",\n    "+name+": "+val)
58.             wFile.write("}\n }\n}")
59.
60. if __name__ == "__main__":
61.     dstPath = "/Users/ed/Documents/itmo/1infa/lab4/lab4_3.txt"
62.     with open(dstPath, mode='a', encoding='utf=8') as wFile:
63.         start_time = time.time()
64.         main()
65.         d = (time.time() - start_time) * 1000
66.         wFile.write("lab4_2.py: %s (seconds*1000)\n" % d)
```



## dst\_2.json

```
1.  {"shedule":
2.    {"Br":[
3.      {"time": "11:40-13:10",
4.        "evenWeek": false,
5.        "auditorium": 306,
6.        "building": "Кронверкский пр., д.49, лит.А",
7.        "subject": "ИНФОРМАТИКА(ЛАБ)",
8.        "teacher": "Белозубов Александр Владимирович",
9.        "lessonType": "Очно – дистанционный"},
10.     {"time": "11:40-13:10",
11.       "evenWeek": true,
12.       "auditorium": 285,
13.       "building": "Кронверкский пр., д.49, лит.А",
14.       "subject": "ИНФОРМАТИКА(ЛЕК)",
15.       "teacher": "Белозубов Александр Владимирович",
16.       "lessonType": "Очно – дистанционный"},
17.     {"time": "11:40-13:10",
18.       "evenWeek": false,
19.       "auditorium": 306,
20.       "building": "Кронверкский пр., д.49, лит.А",
21.       "subject": "ИНФОРМАТИКА(ЛАБ)",
22.       "teacher": "Белозубов Александр Владимирович",
23.       "lessonType": "Очно – дистанционный"},
24.     {"time": "13:30-15:00",
25.       "evenWeek": true,
26.       "auditorium": 305,
27.       "building": "Кронверкский пр., д.49, лит.А",
28.       "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",
29.       "teacher": "Блохина Елена Николаевна",
30.       "lessonType": "Очно – дистанционный"},
31.     {"time": "15:20-16:50",
32.       "evenWeek": true,
33.       "auditorium": 305,
34.       "building": "Кронверкский пр., д.49, лит.А",
35.       "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",
36.       "teacher": "Блохина Елена Николаевна",
37.       "lessonType": "Очно – дистанционный"}]
38.   }
39. }
```

---

### Сравнение lab4\_2.py, lab4\_1.py и lab4\_0.py:

В **lab4\_2.py** регулярные выражения позволили сократить размер кода, а также сократили время написания программы, в сравнении с **lab4\_0.py**. Однако по прежнему по данным показателям лидирует **lab4\_1.py**.

### Дополнительное задание №3

- 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 цикл.
- Проанализировать полученные результаты и объяснить их сходство/различие.

В каждый из файлов `lab4_0.py`, `lab4_1.py`, `lab4_2.py` я дописал строки кода, позволяющие оценить время выполнения каждой программы. Также я создал shell-скрипт `lab4_3.sh`, выполняющий последовательно каждую из данных программ, каждая из которых выполняет соответствующую запись в файл `lab4_3.txt`.

## lab4\_3.sh

```
1. python3 lab4_0.py
2. python3 lab4_1.py
3. python3 lab4_2.py
```

## lab4\_3.txt

```
1. lab4_0.py: 0.64635 (seconds*1000)
2. lab4_1.py: 5.44905 (seconds*1000)
3. lab4_2.py: 0.97990 (seconds*1000)
```

В данном случае написание своих алгоритмов дает преимущество во времени выполнения программы, так как модульные функции, методы предназначены для общих случаев и могут выполняться дольше.

Следует отметить, что потеря времени при использовании регулярных выражений не столь существенна, чем при использовании модулей `yaml` и `json`.

## Выводы

Исследовал протоколы, форматы обмена информацией и языки разметки документов. Составил файл в формате `yaml` своего расписания для **вторника**. Написал несколько программ на языке Python 3.x, которые осуществляют парсинг исходного `yaml`-файла в формат `json`.