

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 дисциплине

«Информатика»

Вариант №17

Выполнил:

Студент группы Р3113

Султанов А.Р.

Проверил:

Белозубов А.В.

г. Санкт-Петербург

2022г.

## Оглавление

<b>Оглавление</b>	<b>2</b>
<b>Задание</b>	<b>3</b>
<b>Решение</b>	<b>6</b>
<b>Исходный код</b>	<b>8</b>
<b>Заключение</b>	<b>9</b>
<b>Литература</b>	<b>10</b>

## Задание

1. Изучить форму Бэкуса-Наура.
2. Изучить особенности языков разметки/форматов JSON, YAML, XML.
3. Понять устройство страницы с расписанием для своей группы:  
<http://itmo.ru/ru/schedule/0/P3110/schedule.htm>
4. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы в выбранном дне было не менее двух занятий (можно использовать своё персональное). В случае, если в данный день недели нет таких занятий, то увеличить номер варианта ещё на восемь.
5. 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.
6. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.
7. Дополнительное задание No1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  - a) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
  - b) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
  - c) Сравнить полученные результаты и объяснить их сходство/различие.
8. Дополнительное задание No2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
  - a) Переписать исходный код, добавив в него использование регулярных выражений.

b) Сравнить полученные результаты и объяснить их сходство/различие.

9. Дополнительное задание No3 (позволяет набрать +10 процентов от максимального числа баллов BaPC за данную лабораторную).

a) Используя свою исходную программу из обязательного задания, программу из дополнительного задания No1 и программу из дополнительного задания No2, сравнить стократное время выполнения парсинга + конвертации в цикле.

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

10. Дополнительное задание No4 (позволяет набрать +5 процентов от максимального числа баллов BaPC за данную лабораторную).

c) 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.

d) Проанализировать полученные результаты, объяснить особенности использования формата.

№ варианта	Исходный формат	Результирующий формат	День недели
0	JSON	XML	Понедельник
1	XML	JSON	Понедельник
2	JSON	YAML	Понедельник
3	YAML	JSON	Понедельник
4	YAML	XML	Понедельник
5	XML	YAML	Понедельник
6	JSON	XML	Вторник
7	XML	JSON	Вторник
8	JSON	YAML	Вторник
9	YAML	JSON	Вторник
10	YAML	XML	Вторник
11	XML	YAML	Вторник
12	JSON	XML	Среда
13	XML	JSON	Среда
14	JSON	YAML	Среда
15	YAML	JSON	Среда
16	YAML	XML	Среда
17	XML	YAML	Среда
18	JSON	XML	Четверг
19	XML	JSON	Четверг
20	JSON	YAML	Четверг
21	YAML	JSON	Четверг
22	YAML	XML	Четверг
23	XML	YAML	Четверг

*Рисунок 1. Вариант задания*

## Решение

Для начала необходимо сформировать XML-файл на основе структуры расписания (файл - **input/input.xml**):

```
<?xml version="1.0" encoding="UTF-8"?>
<timetable>
  <days>
    <day id="3day" name="wednesday">
      <lessons>
        <lesson>
          <time>11:40-13:10</time>
          <weeks>
            <week>2</week>
            <week>4</week>
            <week>6</week>
            <week>8</week>
            <week>10</week>
            <week>12</week>
            <week>14</week>
            <week>16</week>
          </weeks>
          <room>
            <address>Кронверкский пр., д.49, лит.А</address>
            <name>2308 (бывш. 306) ауд.</name>
          </room>
          <detail>
            <name>Информатика(Лаб)</name>
            <teacher>Белозубов Александр Владимирович</teacher>
            <format>Очно - дистанционный</format>
          </detail>
```

</lesson>

<lesson>

<time>13:30-15:00</time>

<weeks>

<week>2</week>

<week>4</week>

<week>6</week>

<week>8</week>

<week>10</week>

<week>12</week>

<week>14</week>

<week>16</week>

</weeks>

<room>

<address>Кронверкский пр., д.49, лит.А</address>

<name>2308 (бывш. 306) ауд.</name>

</room>

<detail>

<name>Информатика(Лаб)</name>

<teacher>Белозубов Александр Владимирович</teacher>

<format>Очно - дистанционный</format>

</detail>

</lesson>

<lesson>

<time>13:30-15:00</time>

<weeks>

<week>2</week>

<week>4</week>

<week>6</week>

<week>8</week>

```

    <week>10</week>
    <week>12</week>
    <week>14</week>
    <week>16</week>
  </weeks>
  <room>
    <address>Кронверкский пр., д.49, лит.А</address>
    <name>2308 (бывш. 306) ауд.</name>
  </room>
  <detail>
    <name>Информатика(Лаб)</name>
    <teacher>Белозубов Александр Владимирович</teacher>
    <format>Очно - дистанционный</format>
  </detail>
</lesson>
<lesson>
  <time>15:20-16:50</time>
  <weeks>
  </weeks>
  <room>
    <address>Кронверкский пр., д.49, лит.А</address>
    <name>2304 (бывш. 302) ауд.</name>
  </room>
  <detail>
    <name>Дискретная          математика          (базовый
уровень)(лек)</name>
    <teacher>Поляков Владимир Иванович</teacher>
    <format>Очно - дистанционный</format>
  </detail>
</lesson>

```



```

        </lessons>
    </day>
</days>
</timetable>

```

Далее, необходимо написать конвертер (XML в YAML) без использования python-библиотек (файл - **converter\_custom.py**). В нём используются:

1. Написанный мною XML-парсер (файл - **le\_xml.py**), построенный на рекуррентном посимвольном проходе. Он преобразует XML в python-словарь для промежуточного хранения
2. Написанный мною YAML-преобразователь (файл - **le\_yaml.py**). Он преобразует python-словарь в YAML.

Результат выполнения **converter\_custom.py** (**output/output\_custom.yaml**):

```

timetable:
  value:
    days:
      value:
        day:
          attrs:
            id: 3day
            name: wednesday
          value:
            lessons:
              value:
                - lesson:
                    value:
                      time:
                        value: 11:40-13:10

```

weeks:

value:

- week:

value: 2

- week:

value: 4

- week:

value: 6

- week:

value: 8

- week:

value: 10

- week:

value: 12

- week:

value: 14

- week:

value: 16

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2308 (бывш. 306) ауд.

detail:

value:

name:

value: Информатика(Лаб)

teacher:

value: Белозубов Александр Владимирович

format:

value: Очно - дистанционный

- lesson:

value:

time:

value: 13:30-15:00

weeks:

value:

- week:

value: 2

- week:

value: 4

- week:

value: 6

- week:

value: 8

- week:

value: 10

- week:

value: 12

- week:

value: 14

- week:

value: 16

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2308 (бывш. 306) ауд.

detail:

value:

name:

value: Информатика(Лаб)

teacher:

value: Белозубов Александр Владимирович

format:

value: Очно - дистанционный

- lesson:

value:

time:

value: 13:30-15:00

weeks:

value:

- week:

value: 2

- week:

value: 4

- week:

value: 6

- week:

value: 8

- week:

value: 10

- week:

value: 12

- week:

value: 14

- week:

value: 16

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2308 (бывш. 306) ауд.

detail:

value:

name:

value: Информатика(Лаб)

teacher:

value: Белозубов Александр Владимирович

format:

value: Очно - дистанционный

- lesson:

value:

time:

value: 15:20-16:50

weeks:

value: null

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2304 (бывш. 302) ауд.

detail:

value:

name:

value: Дискретная математика (базовый  
уровень)(лек)

teacher:

value: Поляков Владимир Иванович

format:

value: Очно - дистанционный

После этот конвертер был переписан (файл - **converter\_lib.py**) с использованием библиотек (**xmltodict. yaml**). Кол-во написанного кода значительно сократилось за счет того, что реализация конвертера находилась в библиотеках. Сравнение производительности будет приведено позже. Результат выполнения **converter\_lib.py** (**output/output\_lib.yaml**):

timetable:

days:

day:

'@id': 3day

'@name': wednesday

lessons:

lesson:

- detail:

format: Очно - дистанционный

name: Информатика(Лаб)

teacher: Белозубов Александр Владимирович

room:

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

name: 2308 (бывш. 306) ауд.

time: 11:40-13:10

weeks:

week:

- '2'
- '4'
- '6'
- '8'
- '10'
- '12'
- '14'
- '16'

- detail:

format: Очно - дистанционный

name: Информатика(Лаб)

teacher: Белозубов Александр Владимирович

room:

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

name: 2308 (бывш. 306) ауд.

time: 13:30-15:00

weeks:

week:

- '2'
- '4'
- '6'
- '8'
- '10'
- '12'
- '14'
- '16'

- detail:

format: Очно - дистанционный

name: Информатика(Лаб)

teacher: Белозубов Александр Владимирович

room:  
address: Кронверкский пр., д.49, лит.А  
name: 2308 (бывш. 306) ауд.  
time: 13:30-15:00  
weeks:  
week:  
- '2'  
- '4'  
- '6'  
- '8'  
- '10'  
- '12'  
- '14'  
- '16'  
- detail:  
format: Очно - дистанционный  
name: Дискретная математика (базовый уровень)(лек)  
teacher: Поляков Владимир Иванович  
room:  
address: Кронверкский пр., д.49, лит.А  
name: 2304 (бывш. 302) ауд.  
time: 15:20-16:50  
weeks: null

Реализация конвертера с использованием регулярных выражений представлена в файле **converter\_regex.py**. Объем кода так же значительно меньше, чем в изначальном варианте. Результат выполнения **converter\_regex.py (output/output\_regex.yaml)**:

timetable:  
value:



days:  
value:  
  day:  
  attrs:  
  id: 3day  
  name: wednesday  
  value:  
  lessons:  
  value:  
  - lesson:  
    value:  
    time:  
    value: 11:40-13:10  
    weeks:  
    value:  
    - week:  
      value: 2  
    - week:  
      value: 4  
    - week:  
      value: 6  
    - week:  
      value: 8  
    - week:  
      value: 10  
    - week:  
      value: 12  
    - week:  
      value: 14  
    - week:

value: 16

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2308 (бывш. 306) ауд.

detail:

value:

name:

value: Информатика(Лаб)

teacher:

value: Белозубов Александр Владимирович

format:

value: Очно - дистанционный

- lesson:

value:

time:

value: 13:30-15:00

weeks:

value:

- week:

value: 2

- week:

value: 4

- week:

value: 6

- week:

value: 8

- week:

value: 10

- week:

value: 12

- week:

value: 14

- week:

value: 16

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2308 (бывш. 306) ауд.

detail:

value:

name:

value: Информатика(Лаб)

teacher:

value: Белозубов Александр Владимирович

format:

value: Очно - дистанционный

- lesson:

value:

time:

value: 13:30-15:00

weeks:

value:

- week:

value: 2

- week:

value: 4  
- week:  
value: 6  
- week:  
value: 8  
- week:  
value: 10  
- week:  
value: 12  
- week:  
value: 14  
- week:  
value: 16  
room:  
value:  
address:  
value: Кронверкский пр., д.49, лит.А  
name:  
value: 2308 (бывш. 306) ауд.  
detail:  
value:  
name:  
value: Информатика(Лаб)  
teacher:  
value: Белозубов Александр Владимирович  
format:  
value: Очно - дистанционный  
- lesson:  
value:  
time:

value: 15:20-16:50

weeks:

value: null

room:

value:

address:

value: Кронверкский пр., д.49, лит.А

name:

value: 2304 (бывш. 302) ауд.

detail:

value:

name:

value: Дискретная математика (базовый

уровень)(лек)

teacher:

value: Поляков Владимир Иванович

format:

value: Очно - дистанционный

Сравнение производительности 3 конвертеров (100 тестов, входной файл - **input/large.xml**, 3.6MiB):

Тип	Время (сек.)
Самописный	71.83768124400012
Библиотечный	182.82278157099972
Самописный с regex	41.90337886100042

Детальное сравнение производительности 3 XML-парсеров (100 тестов, входной файл - **input/large.xml**, 3.6MiB):

Тип	Время (сек.)
Самописный	60.44801583399931
Библиотечный	16.9176900020002
Самописный с regex	29.75027050599965

Детальное сравнение производительности 2 YAML-конвертеров (100 тестов, входной файл - **input/large.xml**, 3.6MiB):

Тип	Время (сек.)
Самописный	9.17349793699941
Библиотечный	170.57052630999988

Самый медленный вариант парсинга - ручной, после - ручной с использованием регулярных выражений (примерно в 2 раза быстрее предыдущего). Библиотечный является лидером по скорости, он обогнал вариант с использованием регулярных выражений примерно в 2 раза. Можно сделать вывод, что использование регулярных выражений повысило скорость парсинга. Основная причина - работа с токенами, а не отдельными символами.

С YAML-преобразователями ситуация обратная. Библиотечная реализация сильно отстает по скорости в сравнении с самописной. Вероятно, это связано с тем, что библиотека помимо прямого перевода данных в YAML-строку занимается преобразованием python-словаря к некоторому промежуточному объекту.

## Исходный код

Исходный код доступен в прикрепленном архиве. Входной файл - **input/input.xml**. Самописный конвертер - **converter\_custom.py** (выходной файл - **output/output\_custom.yaml**). Самописный конвертер с использованием regex - **converter\_regex.py** (выходной файл - **output/output\_regex.yaml**). Библиотечный конвертер - **converter\_lib.py** (выходной файл - **output/output\_lib.yaml**).

## Заключение

В рамках данной лабораторной работы было необходимо изучить структуру YAML и XML и написать конвертеры с различными условиями, а после сравнить их производительность. Эта работа полезна необходимостью написания собственного парсера с учетом особенностей входных и выходных форматов данных, а также познакомиться с готовыми python-библиотеками, выполняющими поставленную задачу.



## Литература

- “Форма Бэкуса — Наура — Википедия.” *Википедия*,

[https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D1%80%D0%BC%D0%B0%D0%91%D1%8D%D0%BA%D1%83%D1%81%D0%B0\\_%E2%80%94%D0%9D%D0%B0%D1%83%D1%80%D0%B0](https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D1%80%D0%BC%D0%B0%D0%91%D1%8D%D0%BA%D1%83%D1%81%D0%B0_%E2%80%94%D0%9D%D0%B0%D1%83%D1%80%D0%B0).

- Балакшин, Павел. “Информатика 2021-2022 Лекция №4.” *Telegram*

*Web*, 2022, [https://t.me/balakshin\\_students/30](https://t.me/balakshin_students/30).