

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

Вариант №22 % 36 = 14

Выполнил:

Девярых Павел
Леонидович

Группа №Р3110

Проверил:

Рыбаков Степан
Дмитриевич

Преподаватель практики

Оглавление

Задание.....	3
Основные этапы вычисления.....	5
Обязательное задание:.....	7
Дополнительное задание 1:.....	8
Дополнительное задание 2:.....	8
Дополнительное задание 3:.....	9
Дополнительное задание 4:.....	9
Дополнительное задание 5:.....	9
Заключение.....	11
Список использованных источников.....	12

Задание

1. Определить номер варианта как остаток деления на 36 последних двух цифр своего идентификационного номера в ISU: например, $125598 / 36 = 26$. В случае, если в оба указанных дня недели нет занятий, то увеличить номер варианта на восемь. В случае, если занятий нет и в новом наборе дней, то продолжать увеличивать на восемь.
2. Изучить форму Бэкуса-Наура.
3. Изучить основные принципы организации формальных грамматик.
4. Изучить особенности языков разметки/форматов JSON, YAML, XML.
5. Понять устройство страницы с расписанием на примере расписания лектора:
https://itmo.ru/ru/schedule/3/125598/raspisanie_zanyatiy.htm
6. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. При этом необходимо, чтобы хотя бы в одной из выбранных дней 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.
7. Обязательное задание (позволяет набрать до 45 процентов от максимального числа баллов БаРС за данную лабораторную):
написать программу на языке Python 3.x или любом другом, которая бы осуществляла парсинг и конвертацию исходного файла в новый путём простой замены метасимволов исходного формата на метасимволы результирующего формата.
8. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.
9. 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.
- 1с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
10. 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.

11.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.

12.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.

2б) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

Основные этапы вычисления

Ср	08:20-09:50 2, 4, 6, 8, 10, 12, 14, 16	Информ 1 (Р3106, Р3107, Р3108, Р3109, Р3110, Р3111, Р3112, Р3113, Р3114, Р3115, Р3116, Р3117, Р3118, Р3119)	1216/0 (УСЛ) АУД. ул.Ломоносова, д.9, лит. М	ИНФОРМАТИКА (ЛЕК) : АКТОВЫЙ ЗАЛ	Очно - дистанционный
Сб	13:30-15:00 2, 4, 6, 8, 10, 12, 14, 16	Информ 1.2 (Р3107)	1415 АУД. Кронверкский пр., д.49, лит.А	ИНФОРМАТИКА (ЛАБ)	Очно - дистанционный
	13:30-15:00 1, 3, 5, 7, 9, 11, 13, 15	Информ 1.1 (Р3106)	1415 АУД. Кронверкский пр., д.49, лит.А	ИНФОРМАТИКА (ЛАБ)	Очно - дистанционный
	15:20-16:50 2, 4, 6, 8, 10, 12, 14, 16	Информ 1.2 (Р3107)	1415 АУД. Кронверкский пр., д.49, лит.А	ИНФОРМАТИКА (ЛАБ)	Очно - дистанционный
	15:20-16:50 1, 3, 5, 7, 9, 11, 13, 15	Информ 1.1 (Р3106)	1415 АУД. Кронверкский пр., д.49, лит.А	ИНФОРМАТИКА (ЛАБ)	Очно - дистанционный

Рисунок 1: (расписание лектора)

Пример файла с расписанием в формате JSON:

```
{
  "расписание_Балакшина_Павла_Валерьевича": {
    "чётные_недели": {
      "номер_недели": [
        0,
        2,
        4,
        6,
        8
      ],
      "день_недели": {
```

```
"среда": {
  "пара_1": {
    "предмет": "информатика",
    "тип": "лекция",
    "время_начала": "08:20",
    "время_окончания": "09:50",
    "тип_пары": "очно-дистанционная"
  }
},
"суббота": {
  "пара_1": {
    "предмет": "информатика",
    "тип": "лабораторная",
    "время_начала": "13:30",
    "время_окончания": "15:00",
    "тип_пары": "очно-дистанционная"
  },
  "пара_2": {
    "предмет": "информатика",
    "тип": "лабораторная",
    "время_начала": "15:20",
    "время_окончания": "16:50",
    "тип_пары": "очно-дистанционная"
  }
}
},
"нечётные_недели": {
  "номер_недели": [
    1,
    3,
    5,
```

```

],
"День_недели": {
  "суббота": {
    "пара_1": {
      "предмет": "информатика",
      "тип": "лабораторная",
      "время_начала": "13:30",
      "время_окончания": "15:00",
      "тип_пары": "очно-дистанционная"
    },
    "пара_2": {
      "предмет": "информатика",
      "тип": "лабораторная",
      "время_начала": "15:20",
      "время_окончания": "16:50",
      "тип_пары": "очно-дистанционная"
    }
  }
}
}
}
}
}
}

```

Обязательное задание:

Исходный файл JSON:

<https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/myJson.json>

Исходный код:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/main_task/json_to_yaml_main.py

Результат:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/resultFinal_main.yaml

Дополнительное задание 1:

Исходный код:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/additional_task1/json_to_yaml_dop_1.py

Результат:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/resultFinal_dop_1.yaml

Готовые библиотеки: стандартная библиотека Python json для парсинга JSON и yaml для дампа словаря в файл yaml.

Результаты не отличаются кроме отличия отступов у элементов списка на одну табуляцию, из-за недоработки библиотеки.

Дополнительное задание 2:

Исходный код:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/additional_task2/json_to_yaml_dop_2.py

Результат:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/resultFinal_dop_2.yaml

Результаты не отличаются ничем, отличия в коде лишь в замене одного цикла на регулярные выражения.

Дополнительное задание 3:

Исходный код:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/additional_task3/json_to_yaml_dop_3.py

Результат:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/resultFinal_dop_3.yaml

результаты не отличаются кроме того, что теперь файлы будут проверены на формальную грамматику перед началом парсинга и по его завершению.

Дополнительное задание 4:

Исходный код:

https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/additional_task3/json_to_yaml_dop_3.py

1) Время работы программы для обязательного задания:

0.0012090420059394091

2) Время работы программы для доп. Задания 1:

0.008681299004820175

3) Время работы программы для доп. Задания 2:

0.0008553720035706647

4) Время работы программы для доп. Задания 3:

0.0015365759973064996

Медленнее всего оказался вариант с использованием библиотеки, но самый надёжный так как там идёт множество проверок для парсинга. Чуть быстрее вариант с регулярными выражениями, но это было ожидаемо ведь они основаны на посимвольном переборе. Вариант с использованием формальных грамматик оказался чуть медленнее из-за проверок основного задания.

Дополнительное задание 5:

Исходный код:

[https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/
additional_task5/json_to_yaml_dop_5.py](https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/additional_task5/json_to_yaml_dop_5.py)

Результат:

[https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/
resultFinal_dop_5.csv](https://github.com/MAZER-shadow/VT-Labs/blob/main/informatics/lab4/data/resultFinal_dop_5.csv)

CSV — текстовый формат, предназначенный для представления табличных данных. Строка таблицы соответствует строке текста, которая содержит одно или несколько полей, разделенных запятыми.

Заключение

Во время выполнения лабораторной работы я познакомился с разными форматами файлов. Узнал о существовании формальных грамматик. Написал парсер конвертирующий json-формат в yaml. Осознал насколько тяжелый труд использовать формальные грамматики.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Казарин, Лев Сергеевич. Введение в теорию кодирования, сжатия и восстановления информации : учебно-методическое пособие: Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2020. – 112 с.
2. Основы цифровой радиосвязи. Помехоустойчивое кодирование: метод. указания / сост. Д. В. Пьянзин. – Саранск : Изд-во Мордов. ун-та, 2009. – 16 с