Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Отчёт о лабораторной работе №4

Исследование протоколов, форматов обмена информацией и

языков разметки документов

Вариант №24

Выполнил:

Дядев Владислав Александрович, гр. P3131

Проверила:

Авксентьева Е. Ю., к.п.н., доцент

Санкт-Петербург 2024

Оглавление

[Задание 3](#_Toc183083064)

[Основные этапы вычисления 7](#_Toc183083065)

[Обязательное задание 7](#_Toc183083066)

[Дополнительное задание №1 7](#_Toc183083067)

[Дополнительное задание №2 8](#_Toc183083068)

[Дополнительное задание №3 8](#_Toc183083069)

[Дополнительное задание №4 8](#_Toc183083070)

[Дополнительное задание №5 9](#_Toc183083071)

[Заключение 10](#_Toc183083072)

[Список использованных источников 11](#_Toc183083073)

# **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

1. 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.
2. 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.
3. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.
4. Дополнительное задание №1 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
5. Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
6. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
7. Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
8. Дополнительное задание №2 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
9. Переписать исходный код, добавив в него использование регулярных выражений.
10. 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).
12. 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 No1.
13. Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
14. Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
15. Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).
16. Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.
17. Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
18. Дополнительное задание №5 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).
19. 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.
20. Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.
21. Проверить, что все пункты задания выполнены и выполнены верно.
22. Написать отчёт о проделанной работе.
23. Подготовиться к устным вопросам на защите.

Вариант задания представлен на Рисунок 1.

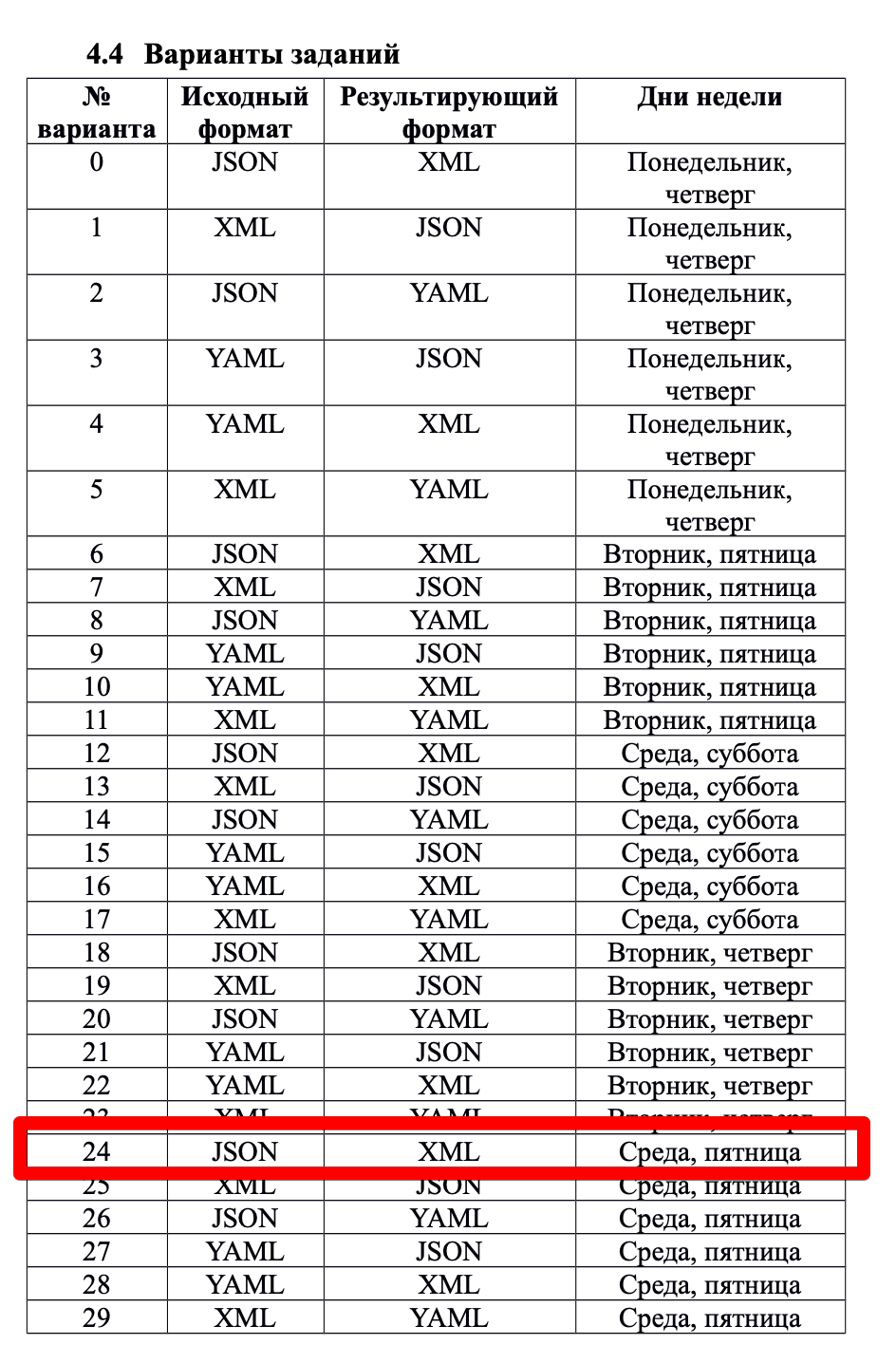


Рисунок 1 - Вариант задания

# **Основные этапы вычисления**

## **Обязательное задание**

Расписание, составленное на основе варианта, представлено в формате JSON в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/in.json>

Программа, осуществляющая парсинг из JSON и конвертацию в XML, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/main.py>

Итоговый файл с расписанием в формате XML представлен в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/out_main.xml>

## **Дополнительное задание №1**

Программа, осуществляющая парсинг из JSON и конвертацию в XML с помощью готовой библиотеки, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/add_task1.py>

Итоговый файл с расписанием в формате XML представлен в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/out1.xml>

Результаты работы этой программы отличается от результата работы исходной тем, что вместо тега <root></root> используется тег <all></all>, а также другим тегом обозначаются элементы списков (<item></item> вместо <{name\_of\_list}\_item></{name\_of\_list}\_item>).

## **Дополнительное задание №2**

Программа, осуществляющая парсинг из JSON и конвертацию в XML с использованием регулярных выражений, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/add_task2.py>

Итоговый файл с расписанием в формате XML представлен в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/out2.xml>

Результат работы этой программы совпадает с результатом работы исходной.

## **Дополнительное задание №3**

Программа, осуществляющая парсинг из JSON и конвертацию в XML с использованием формальных грамматик, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/add_task3.py>

Итоговый файл с расписанием в формате XML представлен в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/out3.xml>

Результат работы этой программы совпадает с результатом работы исходной.

## **Дополнительное задание №4**

Программа, осуществляющая сравнения стократного времени выполнения предыдущих программ, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/add_task4.py>

Результат работы этой программы представлен на *Рисунок 2*.

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Рисунок 2 - Время работы программ

Можно заметить, что быстрее всего работает программа с использованием формальных грамматик, затем идёт программа с использованием регулярных выражений, а после него исходная программа ручных парсинга и конвертации. У первых двух время примерно одинаковое. У исходной программы время также близко к этим двум, потому что все они выполняют ручные парсинг и конвертацию. Дольше же всего работает программа с использованием сторонней библиотеки, потому что в ней учтены все тонкости и случаи парсинга JSON и конвертации его в XML.

## **Дополнительное задание №5**

Программа, осуществляющая парсинг из JSON и конвертацию в CSV, представлена в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/add_task5.py>

Итоговый файл с расписанием в формате CSV представлен в репозитории на github:

<https://github.com/Alvas07/ITMO/blob/main/1%20Informatics/Labs/Lab4/data/out5.csv>

CSV – это формат таблицы, где в роли заголовков выступают ключи из словаря формат JSON, а в ячейки же записываются значения соответствующих ключей. Каждая ячейка таблицы в файле отделяется символ «;».

# **Заключение**

В ходе лабораторной работы мы изучили формальные грамматики, научились осуществлять парсинг из JSON и конвертирование в XML, а также сравнивать время работы программ.

# **Список использованных источников**

1. Лямин А.В., Череповская Е.Н. Объектно-ориентированное программирование. Компьютерный практикум. – СПб: Университет ИТМО, 2017. – 143 с. – Режим доступа: https://books.ifmo.ru/file/pdf/2256.pdf.
2. <https://habr.com/ru/post/309242/>