Федеральное государственное автономное образовательное учреждение высшего образования

Национальный исследовательский университет ИТМО

Факультет программной инженерии и компьютерной техники

# Отчет по лабораторной работе № 4

**«Исследование протоколов, форматов обмена информацией и**

**языков разметки документов»**

**По дисциплине**

**«Информатика»**

**Вариант 2**

Выполнил: студент гр. P3115

Бусыгин Д.А.

Проверилa:

Малышева Т.А.

Санкт-Петербург

2021 г.

Оглавление:

[Задание: 3](#_Toc86777216)

[Основные этапы работы: 4](#_Toc86777217)

[Основное задание. 4](#_Toc86777218)

[Дополнительное задание 1. 4](#_Toc86777219)

[Дополнительное задание 2. 4](#_Toc86777220)

[Дополнительное задание 3. 4](#_Toc86777221)

[Дополнительное задание 4. 5](#_Toc86777222)

[Вывод: 5](#_Toc86777223)

[Репозиторий с исходным кодом: 5](#_Toc86777224)

[Список литературы: 5](#_Toc86777226)

## Задание:

1. Определить номер варианта как остаток деления на 36 порядкового номера в списке группы в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь.
2. Изучить форму Бэкуса-Наура.
3. Изучить особенности протоколов и форматов обмена информацией между системами: JSON, YAML, XML.
4. Понять устройство страницы с расписанием для своей группы: <https://itmo.ru/ru/schedule/0/P3110/schedule.htm>
5. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного.
6. **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.
7. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.
8. **Дополнительное задание №1** (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
   1. Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
   2. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
   3. Сравнить полученные результаты и объяснить их сходство/различие.
9. **Дополнительное задание №2** (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
   1. Переписать исходный код, добавив в него использование регулярных выражений.
   2. Сравнить полученные результаты и объяснить их сходство/различие.
10. **Дополнительное задание №3** (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную).
    1. 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 цикле.
    2. Проанализировать полученные результаты и объяснить их сходство/различие.
11. **Дополнительное задание №4** (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную.
    1. Переписать исходную, чтобы она осуществляла парсинг и конвертацию исходного файла в любой другой формат (кроме JSON, YAML, XML, HTML): PROTOBUF, TSV, CSV, WML и т.п.

## Основные этапы работы:

## Основное задание.

Для выполнения обязательного задания я использовал работу со строками. Стоит отметить, что мой парсер подразумевает знание изначальной структуры XML-файла, необходимо, чтобы он был оформлен по всем канонам (с использованием табуляций). Для облегчения задачи вынесем в отдельные функции проверку строки на ее содержание: только открывающий тег, только закрывающий тег или строка с контентом. На вход программа получает XML-файл, а на выход выдает JSON-файл.

## Дополнительное задание 1.

При возможности использования библиотек задача очень облегчается. Подключив библиотеки xmltodict и json, с легкостью переводим XML-текст в словарь, а после, переводим словарь в json-формат и вывод перенаправляем в файл.

## Дополнительное задание 2.

Внедрение регулярных выражений в исходный код основного задания я выполнил заменой функций поиска тегов и контента на функции с RE. В остальном же, код остался тем же, для «справедливой» проверки на время выполнения программы в доп. задании 3.

## Дополнительное задание 3.

Сравнение времени выполнения программ реализуется с помощью библиотеки time, фиксированием времени начала программы и ее окончания (start\_time и end\_time). Перенаправляем вычисленное время в файл time.txt.

По времени получился такой вывод: мой код, в силу отсутствия сложной и нагроможденной логики, работает чуть быстрее кода с регулярными выражениями, т.к. в них, все-таки, подгружается дополнительная логика, программа задействует больше необязательных элементов. Программа с библиотеками для парсинга же работает еще медленнее, т.к. в ней заложены более сложные классы и методы, которые призваны решать не маленькие и известные файлы, как у меня, а все-таки, универсально решать все задачи этого типа.

## Дополнительное задание 4.

Для задания я выбрал формат CSV, т.к. он «красиво» интерпретирует табличные данные. С помощью библиотек pandas и xml.etree.ElementTree задаем названия столбцов и их наполнение, вывод перенаправляем в файл csv\_schedule.csv.

Полученная таблица на ресурсе GitHub имеет вид, как на рис. 1.

Изображение выглядит как текст

Автоматически созданное описание

Рис.1. Таблица с расписанием в формате csv.

## Вывод:

В процессе выполнения лабораторной работы я ознакомился с понятием парсинга, попрактиковался в извлечении информации с интернет-ресурса, а так же её интерпретации в другие языки разметки текста. Также, это помогло улучшить мои умения в применении языка программирования Python.

## Репозиторий с исходным кодом:

## <https://github.com/Busygind/LaboratoryWork4--INFORMATICS--ITMO-UNIVERSITY>

## Список литературы:

# Статья «Редактируем CSV-файлы, чтобы не сломать данные», режим доступа URL: <https://habr.com/ru/company/hflabs/blog/432906/>

# Статья «Пишем изящный парсер на Python», режим доступа URL: <https://habr.com/ru/post/309242/>

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>