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 № 6**

Выполнил:

Полищенко Николай Николаевич

Группа: Р3112

Проверил:

Малышева Татьяна Алексеевна

г. Санкт-Петербург

2024

**Содержание**

[**Содержание** 2](#_Toc178779731)

[**Задание** 3](#_Toc178779732)

[**Основные этапы вычисления** 4](#_Toc178779733)

[**Заключение** 8](#_Toc178779735)

[**Список использованных источников** 9](#_Toc178779736)

**Задание**

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

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. Объяснение должно быть отражено в отчёте.

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. Объяснение должно быть отражено в отчёте.

Дополнительное задание No3 (позволяет набрать +25 процентов от максимального числа баллов БаРС за данную лабораторную).

1. 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 No1.
2. Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
3. с) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).

1. Используя свою исходную программу из обязательного задания и программы из дополнительных заданий, сравнить стократное время выполнения парсинга + конвертации в цикле.
2. Проанализировать полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

Дополнительное задание №5 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).

1. 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.
2. Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

**Основные этапы решения**

**Обязательное задание**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/task.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule.json), [xml-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule.xml)).

**Дополнительное задание №1**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/dop1.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop1.json), [xml-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop1.xml)).

Различия в выводе программ заключаются в том, как они представляют списки в xml файле. Код основного задания оборачивает в тег каждый элемент списка отдельно, а код со встроенной библиотекой все элементы списка оборачивает в один тег. Написание конвертатора через встроенную библиотеку занимает намного меньше строк.

**Дополнительное задание №2**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/dop2.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop2.json), [xml- файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop2.xml)).

Различий в выводе программ нет. Регулярные выражения заменяют встроенные методы и частично упрощают код. Логика конвертации не изменена.

**Дополнительное задание №3**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/dop3.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop3.json), [xml- файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop3.xml)).

Различия в выводе программ из-за того, что в расписание добавилась одна строка с текстовым значением и одна строка с числовым числом. В код была добавлена обработка числовых значений, изменен цикл для добавления новых строк в xml-файл, добавлен общий тег для обертки всего остального xml-файла.

**Дополнительное задание №4**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/dop4.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop4.json), [xml- файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop4.xml)).

Среднее стократное время выполнения обязательного задания: 0.062378954887390134.  
Среднее стократное время выполнения дополнительного задания №1: 0.46595642566680906.

Среднее стократное время выполнения дополнительного задания №2: 0.1429898500442505.

Среднее стократное время выполнения дополнительного задания №3: 0.06476473808288574.

Обязательное задание выполняется примерно в 7.5 раз быстрее дополнительного задания №1, так как код обязательного задания конвертируют построчно и не обладает сложной логикой, в отличии от кода дополнительного задания №1 со встроенной библиотекой.

Обязательное задание выполняется примерно в 2.3 раза быстрее дополнительного задания №2, так как в коде дополнительного задания используются регулярные выражения, которые замедляют работу кода.

Обязательное задание и дополнительное задание №3 выполняются практически одинакового, так как логика кода одинакова.

**Дополнительное задание №5**

Код программы находится на Github([код программы](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/dop5.py), [json-файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop5.json), [csv- файл](https://github.com/Banbobr/ITMO/blob/main/labs_sem1/informatics/lab4/schedule_dop5.csv)).

CSV - простой текстовый формат для хранения табличных данных, где значения разделяются запятыми (или другими символами-разделителями). Формат CSV удобен для работы с большими объемами данных и легко читается как человеком, так и машиной.

**Заключение**

В ходе выполнения работы я вспомнил структуру json­-файла и csv-файла и узнал о языке разметки XML. Научился конвертировать json-файл в xml-файл без использования библиотек, повторил регулярные выражения.

**Список использованных источников**

1. А.М. Балакшин П.В. Соснин В.В. Калинин И.В. Малышева Т.А. Раков С.В. Рущенко Н.Г. Дергачев. Информатика: лабораторные работы и тесты: учебно-методическое пособие. — Университет ИТМО, 2019. — С. 56.
2. А.С. Грошев. Информатика: Учебник для вузов. — Архангельский Государственный технический Университет, 2016. — С. 470.