

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 18

Выполнил:

Шмунк Андрей Александрович

Группа Р3108

Проверил:

Доцент ПИиКТ, кандидат технических наук

Балакшин Павел Валерьевич

Санкт-Петербург 2023

Содержание

Задание:	3
Основные этапы вычислений	5
1. Обязательное задание	5
2. Дополнительное задание №1	5
3. Дополнительное задание №2	5
4. Дополнительное задание №3	5
5. Дополнительное задание №4	6
6. Дополнительное задание №5	6
Вывод:	6
Источники:	6

Задание:

1. Определить номер варианта как остаток деления на 36 последних двух цифр своего идентификационного номера в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь. **JSON->XML(четверг)**
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 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.
7. Обязательное задание (позволяет набрать до 45 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 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 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. Объяснение должно быть отражено в отчёте.
10. 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. Объяснение должно быть отражено в отчёте.

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

- a) 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.
- b) Проверку осуществить как минимум для расписания с двумя учебными днями по два занятия в каждом.
- c) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.

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

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

13.Дополнительное задание No5 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную).

- a) 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.
- b) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

14. Проверить, что все пункты задания выполнены и выполнены верно.

15. Написать отчёт о проделанной работе.

16. Подготовиться к устным вопросам на защите.

Основные этапы вычислений

Исходный файл: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/schedule.json>

Исходный файл 2 для дополнительного задания №4:

<https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/schedule2.json>

1. Обязательное задание

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/main.py>

Результат: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez.xml>

2. Дополнительное задание №1

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/dop1.py>

Результат: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez1.xml>

В получившемся файле вместо тега <root> пишется <all>. Также отсутствует определение кодировки в первой строчке результирующего XML-файла. В остальном файлы идентичны.

3. Дополнительное задание №2

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/dop2.py>

Результат: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez2.xml>

Получившийся файл идентичен файлу, который был получен в ходе работы программы обязательного задания. Это можно объяснить тем, что частично изменилась обработка строк файла, остальной функционал не поменялся.

4. Дополнительное задание №3

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/dop3.py>

Результат: https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez3_1.xml

Результат с другими входными данными:

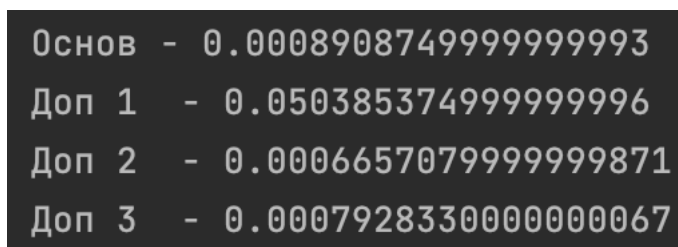
https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez3_2.xml

Получившийся файл идентичен файлу обязательного задания. Алгоритм обработки JSON-файла был изменен с построчного на посимвольный. Функционал был дополнен.

5. Дополнительное задание №4

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/dop4.py>

Результат представлен на Рисунок 1



```
Основ - 0.0008908749999999993
Доп 1 - 0.050385374999999996
Доп 2 - 0.0006657079999999871
Доп 3 - 0.0007928330000000067
```

Рисунок 1

Программа обязательного задания работает быстрее всего, потому что там реализована работа только исходного файла. Доп 2 работает чуть дольше, потому что применено использование регулярных выражений. Доп 3 расширяет функционал парсера, следовательно, увеличивает время работы программы. Доп 1 работает дольше всех, потому что является универсальным парсером, в библиотеке реализованы все тонкости перевода JSON в XML.

6. Дополнительное задание №5

Код: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/dop5.py>

Результат: <https://github.com/Gastozavr/itmo/blob/main/Информатика/lab4/rez5.csv>

Исходный JSON файл был преобразован в файл с расширением CSV. Формат CSV предназначен для представления исходных данных в виде таблицы. Столбцы разделены между собой разделителем, по умолчанию символом “;”.

Вывод:

В ходе выполнения лабораторной работы, я познакомился с форматами JSON, XML, CSV. Также был получен опыт преобразования форматов между собой с помощью библиотек и самописного алгоритма. Кроме того, был повторно изучен материал для работы с регулярными выражениями.

Источники:

1. Балакшин П.В., Соснин В.В., Калинин И.В., Малышева Т.А., Раков С.В., Руценко Н.Г., Дергачев А.М. Информатика: лабораторные работы и тесты: Учебно-методическое пособие / Рецензент: Поляков В.И. - Санкт-Петербург: Университет ИТМО, 2019. - 56 с.
2. Грошев А.С. Г89 Информатика: Учебник для вузов / А.С. Грошев. – Архангельск, Арханг. гос. техн. ун-т, 2010. -470с.