

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 №25

Выполнил:
Решетников Сергей Евгеньевич
Группа Р3108
Проверил:
Преподаватель ПИиКТ
Бострикова Дарья Константиновна

Санкт-Петербург 2024

Оглавление

1. Задание.....	3
2. Основные этапы работы.....	5
3. Вывод.....	7
4. Источники.....	7

1. 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). а) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.

б) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.

с) 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.

- б) 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.

2. Основные этапы работы

Репозиторий с кодом:

https://github.com/NF-coder/ITMO_repo/tree/main/informatics/sem1/lab4

Результат для основного задания:

https://github.com/NF-coder/ITMO_repo/blob/main/informatics/sem1/lab4/out0.json

Результат работы скрипта идентичен результату работы библиотеки, однако он очень чувствителен к входным данным, даже корректный файл со схожей структурой будет обработан некорректно.

Результат для дополнительного задания 3:

https://github.com/NF-coder/ITMO_repo/blob/main/informatics/sem1/lab4/out1.json

Результат работы скрипта идентичен результату работы библиотеки.

Результат для дополнительного задания 2:

https://github.com/NF-coder/ITMO_repo/blob/main/informatics/sem1/lab4/out2.json

Результат работы варианта, использующего регулярные выражения на данном наборе идентичен результату работы скрипта, не использующего их. Однако вариант, использующий регулярные выражения, не умеет обрабатывать атрибуты.

Результат для дополнительного задания 1:

https://github.com/NF-coder/ITMO_repo/blob/main/informatics/sem1/lab4/out3.json

Результат работы библиотеки идентичен результату работы самописного скрипта.

Результат для дополнительного задания 4:

```
/home/nf/Рабочий стол/ИИТQ/ITMO_repo/informatics/sem1/lab4/converters/task1.py 0.055156с (39579100нс)
/home/nf/Рабочий стол/ИИТQ/ITMO_repo/informatics/sem1/lab4/converters/task1.py 0.033840с (31987900нс)
/home/nf/Рабочий стол/ИИТQ/ITMO_repo/informatics/sem1/lab4/converters/task2.py 0.058666с (58038600нс)
/home/nf/Рабочий стол/ИИТQ/ITMO_repo/informatics/sem1/lab4/converters/task3.py 0.037200с (3736400нс)
```

Рисунок 1 - Результат измерений

task1 — Время работы для основного задания и задания №3

task2 — Время работы для задания №2

task3 — Время работы для задания №1

Решение с регулярными выражениями скорее всего тратит много времени на компиляцию регулярного выражения, что замедляет работу на маленьких файлах, но может положительно сказываться на работе с большими файлами

Решение с библиотекой может быть быстрым т. к. внутри может быть код на C + хорошая оптимизация со стороны разработчиков

Результат для дополнительного задания 5:

https://github.com/NF-coder/ITMO_repo/blob/main/informatics/sem1/lab4/out4.proto

XML файл был преобразован в PROTOBUF, который часто используется для эффективного обмена данными между различными частями приложения (gRPC)

3. Вывод

В ходе выполнения лабораторной работы, я познакомился с форматами JSON, XML, PROTOBUF. 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.

4. Источники

1. Информатика [Электронный ресурс]: YouTube. – URL: <https://youtu.be/X8jsijhlIA>
(дата обращения: 05.10.2024).ко

2. Алексеев Е. Г., Богатырев С. Д. Информатика: Мультимедийный электронный учебник. Саранск: 2009.