

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

Факультет Программной Инженерии и Компьютерной Техники

Лабораторная работа № 4

Исследование протоколов, форматов обмена информации и языков разметки
документов

Вариант 11

Выполнил:

Петров Вячеслав Маркович Р3108

Проверил:

Балакшин Павел Валерьевич

Кандидат технических наук, ординарный доцент факультета ПИиКТ

Санкт-Петербург 2023

Содержание

Задание	3
Основные этапы вычисления.....	5
1. Обязательное задание	5
2. Дополнительное задание №1	5
3. Дополнительное задание №2	5
4. Дополнительное задание №3	6
5. Дополнительное задание №4	6
Заключение	8
Список использованных источников	8

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. **XML -> YAML (вторник)**
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 №1 (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 №2 (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 №3 (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 №1.
 - b) 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. Проверить, что все пункты задания выполнены и выполнены верно.

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.

Основные этапы вычисления

1. Обязательное задание

Код:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/stupid_pars_xml_to_yaml/stupid_xml_to_yaml.py

Результат:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/stupid_pars_xml_to_yaml/finish_stupid_pars.yaml

2. Дополнительное задание №1

Код:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/stupid_reg/stupid_xml_to_yaml_reg.py

Результат:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/stupid_reg/finish_stupid_pars.yaml

Различие с основным заданием в том, что при простой замене метасимволов невозможно реализовать выделение каждой структуры, находящейся под тегом «pair», знаком тире вместо написания «pair:» каждый раз. Всё остальное совпадает.

3. Дополнительное задание №2

Код:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/stupid_reg/stupid_xml_to_yaml_reg.py

https://github.com/8B5/lab4/xml_to_yaml_libs.py

Результат:

<https://github.com/8B5/lab4/finish.yaml>

В сравнении с библиотеками осталась та же проблема. Однако, если сравнивать с обязательным заданием, то различий нет, поскольку я заменил условия и циклы на регулярные выражения.

4. Дополнительное задание №3

Код:

https://github.com/8B5/lab4/genious_pars_xml_to_yaml/xml_to_yaml.py

Результат:

https://github.com/8B5/lab4/genious_pars_xml_to_yaml/finish_xml_to_yaml.yaml

Итоговый yaml-файл совпадает с тем, что было получено при использовании библиотек.

5. Дополнительное задание №4

Время 100-кратного повторения:

1. Обязательное задание: 0.1257483959197998
2. Дополнительное задание №1: 0.2025153636932373
3. Дополнительное задание №2: 1.0138137340545654
4. Дополнительное задание №3: 0.0580599308013916

Самый быстрый результат у кода, написанного через формальные грамматики, так как его принцип работы – перевод в единую строку всего кода. Самый медленный код – через регулярные выражения, так как при каждом запуске заново компилируются.

6. Дополнительное задание №5

Код:

https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/xml_to_tsv.py

Результат:

<https://github.com/petrovviacheslav/myitmo/blob/main/%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/Viacheslav/%D0%9B%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B5/lab4/finish.tsv>

TSV - это текстовый формат файла, в котором данные в столбцах разделены знаками табуляции, а строки — знаками перевода строки. Идея по переводу похожа на перевод в yaml, поэтому написание кода не было слишком сложным.

Заключение

В процессе выполнения лабораторной работы я изучил языки разметки XML, YAML, научился обрабатывать данные из них, переводить их из XML в YAML, также изучил библиотеку для парсинга этих языков - xmlplain.

Список использованных источников

1. Балакшин П.В., Соснин В.В., Калинин И.В., Малышева Т.А., Раков С.В., Рущенко 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.