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

Отчёт по информатике

Лабораторная работа № 4

Вариант 2 + 8 = 10

Выполнил: Бутвин.М.П, студент группы P3130

Проверил: Балакшин П.В, доцент факультета ПИиКТ

Санкт-Петербург – 2022

Оглавление

[Задание 2](#_Toc117672166)

[Основные этапы выполнения 3](#_Toc117672167)

[Сравнение нативного решения и библиотеки yaml 3](#_Toc117672168)

[Сравнение нативного решение и решения с регулярными выражениями 3](#_Toc117672169)

[Сравнение быстродействия всех решений 3](#_Toc117672170)

[Конвертации в markdown 3](#_Toc117672171)

[Список литературы 4](#_Toc117672172)

## 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 языков разметки/форматов JSON, YAML, XML.

4. Понять устройство страницы с расписанием для своей группы: <http://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. 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.

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). 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.

9. 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.

10.Дополнительное задание №3 (позволяет набрать +10 процентов от максимального числа баллов БаРС за данную лабораторную). a) Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и программу из дополнительного задания №2, сравнить стократное время выполнения парсинга + конвертации в цикле. b) Проанализировать полученные результаты и объяснить их сходство/различие.

11.Дополнительное задание №4 (позволяет набрать +5 процентов от максимального числа баллов БаРС за данную лабораторную). c) 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. d) Проанализировать полученные результаты, объяснить особенности использования формата.

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

## Основные этапы выполнения

### Сравнение нативного решения и библиотеки yaml

В библиотеке, в отличии от моего решения, учтены все нюансы синтаксиса, производится приведение типов и строится сложное дерево объектов. Мое решение обладает лучшим быстродействием, но не поддерживает некоторые конструкции YAML, а также обладает менее функциональным интерфейсом.

### Сравнение нативного решение и решения с регулярными выражениями

Решение с использованием регулярных выражений работает немного медленнее. Это связано с тем, что регулярные выражения эффективны на большом объеме данных, но в парсинге маленьких строк проигрывают простым преобразованиям. С другой стороны, регулярные выражения позволяют выражать более сложные выражения, что может быть полезным при расширении проекта.

### Сравнение быстродействия всех решений

Время выполнения нативного решения: 0.2655

Время выполнения кода с использованием re: 0.6404

Время выполнения кода с использованием yaml: 5.9610

Самое быстрое решение – без использования каких-либо библиотек. Это связано с тем, что функционал данного решения ограниченнее чем у библиотеки yaml, как уже было сказано ранее. Регулярки же проигрывают из-за того, что парсинг происходит в пределах тегов и значений маленькой длины, где эффективность регулярных выражений не может быть задействована.

### Конвертации в markdown

В качестве еще одного целевого языка разметки я выбрал Markdown. В отличие от YAML или XML он предназначен не для хранения и передачи данных, а для их визуализации. Он часто используется для написания статей и документаций. При этом интерпретация интерфейса остается свободной и при правильной настройке можно получить очень наглядное представление данных.



Джон Бэкус - американский учёный в области информатики



Петер Наур - датский учёный в области информатики

## Список литературы

1. Балакшин П.В, доцент факультета ПИиКТ «Презентация по информатике» https://t.me/balakshin\_students
2. <https://ru.wikipedia.org/wiki/Форма_Бэкуса_—_Наура>
3. https://yaml.org/