Федеральное государственное автономное образовательное учреждение

высшего образования «Национальный исследовательский университет

ИТМО»

*Факультет программной инженерии и компьютерной техники*

*Направление подготовки: 09.03.01 - Информатика и вычислительная техника, Компьютерные системы и технологии*

*Дисциплина «Информатика»*

**Отчет**

**По лабораторной работе №4**

**Исследование протоколов, форматов обмена информацией и языков разметки документов**

**Вариант №10**

Выполнил:

Миронов Иван Николаевич

Группа: Р3132

Принял:

Балакшин Павел Валерьевич

Санкт-Петербург, 2023 год

**Оглавление**

[Задание: 3](#_Toc149381992)

[Выполнение работы: 5](#_Toc149381993)

[Обязательное задание: 5](#_Toc149381994)

[Дополнительное задание 1: 5](#_Toc149381995)

[Дополнительное задание 2: 5](#_Toc149381996)

[Дополнительное задание 3: 5](#_Toc149381997)

[Дополнительное задание 4: 5](#_Toc149381998)

[Дополнительное задание 5: 5](#_Toc149381999)

[Заключение: 6](#_Toc149382000)

[Список литературы: 7](#_Toc149382001)

# 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.
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). 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. 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. 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.

Исходный формат: YAML

Результирующий формат: XML

# Выполнение работы:

## Обязательное задание:

Программа, которая производит простую замену метасимволов.

## Дополнительное задание 1:

Программа, которая использует сторонние библиотеки для парсинга. В отличие от программы для обязательного задания, получается правильный XML файл.

## Дополнительное задание 2:

Программа, которая использует регулярные выражения для парсинга. Результат работы программы идентичен с результатом работы программы из обязательного задания

## Дополнительное задание 3:

Парсер выполняет конвертацию любых данных из YAML в XML. В отличие от программы из прошлого пункта, получается правильный XML файл.

## Дополнительное задание 4:

Парсер, который использует регулярные выражения работает немного медленнее, чем парсер, который не использует их, это связано с тем, что в программе не использованы сложные фильтрующие конструкции.

Настоящий парсер работает немного медленнее, чем обычный парсер и парсер, использующий регулярные выражения, потому что он делает правильный XML файл

Парсер, который использует библиотеки работает значительно дольше всех, потому что сначала создает python словарь, а затем парсит его в XML файл

|  |  |
| --- | --- |
| Тип парсера | Скорость выполнения |
| Обязательное задание | 0.14302 с. |
| Дополнительное задание 1 | 4.23394 с. |
| Дополнительное задание 2 | 0.17436 с. |
| Дополнительное задание 3 | 0.22191 с. |

Таблица 1. Скорость выполнения парсинга

## Дополнительное задание 5:

Парсер выполняет конвертацию любых данных из YAML в Markdown. Самый лаконичный формат для просмотра.

# Заключение:

В ходе выполнения лабораторной работы я научился работать с различными языками разметки, изучил их особенности и применение.

# Список литературы:

1. Форма Бэкуса-Наура: <https://ru.wikipedia.org/wiki/Форма_Бэкуса_—_Наура>
2. Балакшин П.В, доцент факультета ПИиКТ «Презентация по информатике» <https://t.me/balakshin_students>
3. Пишем изящный парсер на Python: <https://habr.com/ru/post/309242/>