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

Вариант № 23

Лабораторная работа № 4

«Исследование протоколов, форматов обмена информацией и языков разметки документов»

по дисциплине

“ **Информатика**”

Выполнил

Студент группы P3110

Чжун Цзяцзюнь : 407959

Преподаватель:

Рыбаков Степан Дмитриевич

Оглавление

[Задание 3](#_Toc182777473)

[Основные этапы вычисления 4](#_Toc182777474)

[1. Обязательное задание 4](#_Toc182777475)

[2. Дополнительное задание №1 4](#_Toc182777476)

[3. Дополнительное задание №2 5](#_Toc182777477)

[4. Дополнительное задание №3 5](#_Toc182777478)

[5. Дополнительное задание №4 5](#_Toc182777479)

[6. Дополнительное задание №5 6](#_Toc182777480)

[Вывод 7](#_Toc182777481)

[Список литературы 8](#_Toc182777482)

# Задание

1. Обязательное задание (позволяет набрать до 45 процентов от максимального числа баллов БаРС за данную лабораторную): написать программу на языке Python 3.x или любом другом, 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.
2. 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. 1 c) Сравнить полученные результаты и объяснить их сходство/различие. Объяснение должно быть отражено в отчёте.
3. 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.
4. 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. 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.
5. Дополнительное задание №4 (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.
6. Дополнительное задание №5 (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. 2 b) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

|  |  |  |  |
| --- | --- | --- | --- |
| 23 | XML | YAML | Вторник, четверг |

# Основные этапы вычисления

# Обязательное задание

Исходный файл xml:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myXML/input_schedule.xml>

Исходный код:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/task/task1.py>

Результат YAML:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myYAML/output_schedule.yaml>

# Дополнительное задание №1

Исходный код:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/task/task2.py>

Результат YAML:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myYAML/output_schedule_ex1.yaml>

# Дополнительное задание №2

Исходный код:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/task/task3.py>

Результат YAML:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myYAML/output_schedule_ex2.yaml>

# Дополнительное задание №3

Исходный код:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/task/task4.py>

Результат YAML:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myYAML/output_schedule_ex3.yaml>

# Дополнительное задание №4

1. Время работы программы для обязательного задания:

0.01520848274230957 seconds

1. Время работы программы для доп задания №1:

0.003412961959838867 seconds

1. Время работы программы для доп задания №2:

0.002999544143676758 seconds

1. Время работы программы для доп задания №3:

0.002971172332763672 seconds

1. Время работы программы для доп задания №5:

0.0029981136322021484 seconds

Возможное объяснение результатов:

Использование специализированных библиотек, таких как xml.etree.ElementTree (например, в task2.py, task4.py и task5.py), обычно более эффективно, чем манипулирование строками или разбор регулярных выражений. Манипулирование строками хорошо подходит для разрабора простого текста, но имеет низкую производительность и читаемость для структурированных данных, таких как XML. Регулярные выражения обеспечивают гибкий подход к XML, но они менее эффективны и удобны в обслуживании, чем специализированные XML-парсеры, из-за их способности обрабатывать вложенные и сложные структуры.

# Дополнительное задание №5

Исходный код:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/task/task5.py>

Результат CSV:

<https://github.com/ivanstylish/ITMO-Study-Note/blob/main/Informatics/Labwork/lab4/myYAML/output_schedule_ex5.csv>

# Вывод

Я изучал разные языки разметки, например, JSON, XML,CSV,YAML, и.т.к. Научился как писать программу на языке Python, которая бы осуществляла парсинг и конвертацию исходного файла в новый путём.Научился работать с некоторыми библиотеками для парсинга, такими как pyyaml, xml.

# Список литературы

1. ГОСТ 12.11.2024 форма Бэкуса-Наура

<https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D1%80%D0%BC%D0%B0_%D0%91%D1%8D%D0%BA%D1%83%D1%81%D0%B0_%E2%80%94_%D0%9D%D0%B0%D1%83%D1%80%D0%B0>

1. ГОСТ 12.11.2024 основные принципы организации формальных грамматик

<https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0>

1. ГОСТ 13.11.2024 особенности языков разметки/форматов JSON, YAML, XML

<https://aws.amazon.com/ru/compare/the-difference-between-json-xml/>

<https://habr.com/ru/articles/248147/>

1. ГОСТ 13.11.2024 Чтение и запись файлов CSV

<https://docs.python.org/zh-cn/3/library/csv.html>