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 8

Выполнила:

Абдуллаева София Улугбековна

Группа P3108

Проверила:

Бострикова Дарья Константиновна

**Оглавление**

[Задание 3](#_Toc182661693)

[Основные этапы выполнения 4](#_Toc182661694)

[Обязательное задание 4](#_Toc182661695)

[Дополнительное задание №1 4](#_Toc182661696)

[Дополнительное задание №2 4](#_Toc182661697)

[Дополнительное задание №3 5](#_Toc182661698)

[Дополнительное задание №4 5](#_Toc182661699)

[Дополнительно задание №5 6](#_Toc182661700)

[Вывод: 6](#_Toc182661701)

[Список литературы: 6](#_Toc182661702)

Задание

*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.

Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

*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.

*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.

*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.

*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.

*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.

b) Проанализировать полученные результаты, объяснить особенности использования формата. Объяснение должно быть отражено в отчёте.

Проверить, что все пункты задания выполнены и выполнены верно.

Написать отчёт о проделанной работе.

Подготовиться к устным вопросам на защите.

Основные этапы выполнения

Обязательное задание

Расписание в формате JSON, составленное для парсинга можно посмотреть по ссылке:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/schedule.json>

Исходный код программы – парсер из JSON в YAML:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/maintask_and_extratask3.py>

Результат работы программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/maintask_and_extratask3_output.yml>

Дополнительное задание №1

Исходный код программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask1.py>

Результат работы программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask1_output.yml>

Используются такие готовые библиотеки, как стандартная библиотека json, которая была нужна для парсинга, и PyYAML для преобразования данных в YAML через dump. Файл, полученный в результате конвертации, отличается только тем, что остаются кавычки в тех случаях, когда строковые значения содержат символы, которые требуют экранирования (например, пробел, символ новой строки). Более того, данный способ значительно сокращает объём кода.

Дополнительное задание №2

Исходный код программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask2.py>

Результат работы программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask2_output.yml>

Файл результата полностью совпадает с файлом, полученным в результате обязательного задания, так как функция, которая конвертирует из словаря в YAML, осталась прежней, и регулярные выражения использовались в парсинге из JSON в словарь. Регулярные выражения позволили упростить парсинг строк, чисел, null и булевых значений.

Дополнительное задание №3

Исходный код программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/maintask_and_extratask3.py>

Результат работы программы:

<https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/maintask_and_extratask3_output.yml>

В работе кода отсутствуют различия, потому что код для обязательного задания был написан с использованием знаний формальных грамматик.

Дополнительное задание №4

Исходный код программы: <https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask4.py>

Результат работы программы:

Время выполнения ручного перевода из JSON в YAML: 0.06285715103149414 секунд

Время выполнения перевода из JSON в YAML с помощью библиотеки: 0.2409529685974121 секунд

Время выполнения перевода из JSON в YAML с помощью регулярных выражений: 0.06518888473510742 секунд

Код из дополнительного задания №1 выполняется дольше всех. Это связано с тем, что библиотеки включают дополнительные проверки типов, то есть при обработке данных они анализируют каждый элемент, чтобы корректно определить и сохранить его тип. Данные проверки требуют времени, поэтому программа работает медленнее. 2 программы, в которых использовались правила грамматики и регулярные выражения, показали лучшие результаты, так как происходит обработка строк без каких-либо дополнительных проверок.

Дополнительно задание №5

Исходный код программы: <https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/extratask5.py>

Результат работы программы: <https://github.com/sophiaabd/ITMO/blob/main/1%20%D0%98%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0/lab4/schedule.csv>

CSV - это текстовый формат, где данные организованы в виде таблиц, строки данных разделяются переносом строки, а столбцы - запятыми или другим разделителями, например, точками с запятой. В моей программе данные из JSON преобразуются в строки, и каждый элемент из массива lessons становится отдельной строкой в CSV файле с указанными полями. Также в CSV метки полей указываются только в заголовке, а все последующие строки содержат только значения. Для обработки нужно лишь разделить строку по заданному разделителю, и этот формат не поддерживает вложенные структуры данных. В JSON обрабатываются вложенные структуры, и каждая запись сопровождается метками полей, что увеличивает объем данных. Следовательно, CSV файлы занимают меньше места в памяти, и этот формат обрабатывается быстрее JSON.

Вывод:

Во время выполнения лабораторной работы я узнала о таких языках разметки, как JSON и YAMl, научилась работать с ними, производить перевод из одного формата в другой с помощью библиотек и регулярных выражений в Python. Также узнала о формальных грамматиках и форме Бэкуса-Наура.

Список литературы:

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. - экз. - Режим доступа:

<https://books.ifmo.ru/book/2248/informatika:_laboratornye_raboty_i_testy:_uchebno-metodicheskoe_posobie_/_recenzent:_polyakov_v.i..htm>

1. Грошев А.С. Г89 Информатика: Учебник для вузов / А.С. Грошев. – Архангельск, Арханг. гос. техн. ун-т, 2010. -470с. -Режим доступа <https://narfu.ru/university/library/books/0690.pdf>