

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
ИТМО»
Факультет Программной Инженерии и Компьютерной Техники

Лабораторная работа №4 по дисциплине «Информатика»
Исследование протоколов, форматов обмена информацией и языков разметки
документов
Вариант 28

Выполнил:
Студент группы Р3112
Думцев В.С.
Преподаватель:
Авксентьева Е.Ю. ,к.п.н., доцент факультета ПИиКТ

Санкт-Петербург 2023

Оглавление

Задание	3
Обязательное задание	4
Дополнительное задание 1	5
Дополнительное задание 2	5
Дополнительное задание 4 и сравнение программ	6
Вывод	6
Код программы.....	7
Список источников	7

Задание

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).
- а) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.
 - б) Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать.
 - с) 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).
- а) Переписать исходный код, добавив в него использование регулярных выражений.
 - б) 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).
- 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.

б) 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 № 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.

2

15. Написать отчёт о проделанной работе.

16. Подготовиться к устным вопросам на защите.

Обязательное задание

Содержимое итогового файла, после работы программы для обязательного задания, приведено на Рисунке 1.1

```

<friday>
  <linearAlgebra>
    <lecture>
      <start>8:20</start>
      <end>9:50</end>
      <teacher>"Lapin Ivan Aleksandrovich"</teacher>
      <location>Kronv</location>
      <room>1405</room>
    </lecture>
    <practice>
      <start>10:00</start>
      <end>11:30</end>
      <teacher>"Shavrin Andrey Andreevich"</teacher>
      <location>Kronv</location>
      <room>2336</room>
    </practice>
  </linearAlgebra>
</friday>

```

Рисунок 1.1

Программа работает быстро, использует рекурсивную функцию для замены строк.

Дополнительное задание 1

Итоговый файл после применения программы, представлен на Рисунке 1.2

```

<root><friday><linearAlgebra><lecture><start>8-20</start><end>9-50</end><teacher>Lapin Ivan Aleksandrovich</teacher><location>Kronv</location><room>1405</room></lecture><practice><start>10-00</start><end>11-30</end><teacher>Shavrin Andrey Andreevich</teacher><location>Kronv</location><room>2336</room></practice></linearAlgebra></friday></root>

```

Рисунок 1.2

При создании программы использовались библиотеки `elementpath` и `ruyaml`(см. список источников).

Происходит конвертация из `yaml` в структуру данных на `python`. Эта программа способна преобразовывать более сложные структуры, чем исходный вариант программы, но работает медленнее.

Дополнительное задание 2

Результат на Рисунке 1.3

```

<friday>linearAlgebra:</friday>
  <lecture>start: 8-20</lecture>
  <end>9-50</end>
  <teacher>"Lapin Ivan Aleksandrovich"</teacher>
  <location>Kronv</location>
  <room>1405</room>
  <practice>start: 10-00</practice>
  <end>11-30</end>
  <teacher>"Shavrin Andrey Andreevich"</teacher>
  <location>Kronv</location>
  <room>2336</room>

```

Рисунок 1.3

Как видите, некоторые структуры не обработались, а некоторые обработались неправильно. Дело в том, что при конвертации из YAML необходимо учитывать отступы и уровень вложенности. А регулярные выражения в данном случае не помогают, а только мешают.

Дополнительное задание 4 и сравнение программ

Таблица 1 – Сравнение программ

	Стократное время выполнения (округленное)	Использование сторонних библиотек	Обрабатывает сложные структуры	Добавляет отступы в результатирующем формате
main.py	0.0175714	Нет	Нет	Да
add1.py	0.3041744	Да	Да	Нет
add2.py	0.0483751	Из стандартной библиотеки	Нет	Добавляет, но не всегда правильно

Вывод

В ходе выполнения работы, я научился работать с форматами yaml, xml и json. Также, я улучшил своё понимание регулярных выражений и навыки владения Python.

Код программы

Все программы из заданий находятся в репозитории https://github.com/G0ld3nDr4g0n3301/ITMO_Informatics_labs/tree/main/lab4 .

Список источников

Электронные ресурсы:

1. ElementPath documentation [Электронный ресурс]. – 2018. – URL: <https://elementpath.readthedocs.io/en/latest/introduction.html> (дата обращения 06.11.2023).
2. PyYAML documentation [Электронный ресурс]. – 2023. – URL: <https://pyyaml.org/wiki/PyYAMLDocumentation> (дата обращения 06.11.2023).