

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 4

Выполнил:

Горин Семён Дмитриевич

Группа Р3108

Проверила:

Бострикова Дарья Константиновна

Санкт-Петербург 2024

Содержание

<i>Задание.....</i>	<i>3</i>
<i>Основное задание л/р</i>	<i>4</i>
<i>Дополнительное задание №1</i>	<i>4</i>
<i>Дополнительное задание №2</i>	<i>4</i>
<i>Дополнительное задание №3</i>	<i>5</i>
<i>Дополнительное задание №4</i>	<i>5</i>
<i>Дополнительное задание №5</i>	<i>5</i>
<i>Вывод:.....</i>	<i>5</i>
<i>Источники:.....</i>	<i>6</i>

Задание

Обязательное задание (позволяет набрать до 45 процентов от максимального числа баллов BaPC 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 BaPC 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 BaPC 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 BaPC за данную лабораторную).

- a) 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.
- c) 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 BaPC 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.

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.

Проверить, что все пункты задания выполнены и выполнены верно.

Написать отчёт о проделанной работе.

Подготовиться к устным вопросам на защите.

Основное задание л/р

Составленное для парсинга расписание в формате YAML можно посмотреть по [ссылке](#).

Листинг программы – парсера из YAML в XML доступен по [ссылке](#).

Результат работы программы также доступен на [Github](#).

Дополнительное задание №1

Ссылка на листинг программы на [Github](#).

Результат выполнения программы на [Github](#).

Значительных различий в выводе программ нет, отличается лишь формат отступов перед тегами. В моем коде они сделаны при помощи пробелов, тогда как библиотека использует символы табуляции для создания отступов. Также можно заметить колоссальное уменьшение количества кода необходимого для преобразования файла.

Дополнительное задание №2

Ссылка на листинг программы на [Github](#).

Результат выполнения программы на [Github](#).

Различий в выводе исходной и измененной программ нет, так как регулярные выражения были использованы лишь в блоке считывания данных из YAML в словарь, в то время когда функция записи из словаря в файл XML осталась нетронутой.

Дополнительное задание №3

Листинг программы – парсера из YAML в XML доступен по [ссылке](#).

Результат работы программы также доступен на [Github](#).

Различия в работе программы отсутствуют, так как код изначально писался с применением знаний о формальных грамматиках, для полноценного парсинга из YAML в XML.

Дополнительное задание №4

Код программы для измерения времени работы доступен для просмотра на [Github](#).

Результат выполнения программы:

```
MainTaskAndAdditionalTask3: 0.5120919999899343  
AdditionalTask1: 0.5600823999848217  
AdditionalTask2: 0.49209389998577535
```

Можем заметить, что дольше всего выполнялся код дополнительного задания №1. Это происходит из-за того, что библиотека обрабатывает значения данных с сохранением типизации, в то время как я записываю данные как строки. Также заметим, что код, использующий регулярные выражения работает быстрее нежели код без них, так как при большом объеме данных идентификация терминалов происходит быстрее.

Дополнительное задание №5

Листинг программы доступен на [Github](#).

[Ссылка](#) на результат выполнения программы.

YAML позволяет создавать объекты большой вложенности (т.н. маппинг). Представление подобных объектов в табличных форматах (таких как CSV или TSV) крайне неудобно. Из предложенных форматов остается лишь PROTOBUF и WML, и так как WML очень похож на хорошо знакомый мне XML выбор пал именно на него. WML (Wireless Markup Language) — это язык разметки, созданный специально для разработки веб-страниц и приложений для мобильных устройств, использующих протокол WAP (Wireless Application Protocol). Впервые разработанный в 1998 году, WML сыграл важную роль в период становления мобильного интернета, когда устройства имели ограниченные возможности по обработке графики и памяти. Сегодня WML практически не используется, так как его заменили более современные технологии, такие как HTML5 и CSS3, которые поддерживаются современными мобильными браузерами и обеспечивают гораздо больший функционал и гибкость.

Вывод:

В ходе выполнения лабораторной работы я узнал о языках разметки XML и YAML, научился с ними работать и переводить один тип в другой с помощью средств языка Python. Узнал о формальных грамматиках и БНФ.

Источники:

1. Балакшин П.В., Соснин В.В., Калинин И.В., Малышева Т.А., Раков С.В., Руценко 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 / recenzyent: polyakov v.i..htm>
2. 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>