

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

**Отчет**  
**по лабораторной работе №4**  
**«Исследование протоколов, форматов обмена информацией и языков**  
**разметки документов»**  
**по дисциплине «Информатика»**

Вариант №6

Выполнил: Кокорин Всеволод Вячеславович, группа Р3118  
Преподаватель: Рыбаков Степан Дмитриевич

г. Санкт-Петербург  
2022г.

# Содержание

1	Задание 1	2
2	Задание 2	3
3	Задание 3	4
4	Задание 4	5
5	Задание 5	6
6	Вывод	7
7	Список литературы	8

# 1 Задание 1

Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.

Решение:

[https://github.com/Slonser/ITMO\\_labs/tree/main/cs/labs/lab4/myparser](https://github.com/Slonser/ITMO_labs/tree/main/cs/labs/lab4/myparser)

## 2 Задание 2

- 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.

```
def task2(file="out.json"):
    from dict2xml import dict2xml
    from json import loads
    with open(file) as j:
        return dict2xml(loads(j.read()))
```

Различия:

- Форматирование
- Формат представление массива

### 3 Задание 3

- a) Переписать исходный код, добавив в него использование регулярных выражений.  
b) Сравнить полученные результаты и объяснить их сходство/различие.  
Различия между данным решением и первым отсутствуют

```
def task3(file="out.json"):
    import re
    from dict2xml import dict2xml

    example = {'pair_id': 2460032073, 'subject': 'Дискретная математика (базовый уровень)',
    data = []
    with open(file) as json:
        value = json.read().replace("null", '')
        lessons_count = value.count("pair_id")
        for i in range(lessons_count):
            data.append({})
            for key in example:
                if isinstance(example[key], str):
                    data[-1][key] = eval(list(re.finditer(f'"{key}":\s*("[^"]*)"', value))[i]
                if isinstance(example[key], int):
                    data[-1][key] = eval(list(re.finditer('"' + key + "":\s*([^\,}]+)[\,}]', value))
    return dict2xml(data)
```

## 4 Задание 4

а) Используя свою исходную программу из обязательного задания, программу из дополнительного задания No1 и программу из дополнительного задания No2, сравнить стократное время выполнения парсинга + конвертации в цикле.

б) Проанализировать полученные результаты и объяснить их сходство/различие.

Свое - 0.0011454529099864884

Библиотеки - 0.000519692079978995

Регулярные выражения - 0.0013138183404225855

Библиотечная реализация работает быстрее всего из-за более грамотной реализации лексера, а также не рекурсивного построения дерева xml

Решение через регулярные выражения самое медленное из-за многократного построения деревьев регулярного выражения

## 5 Задание 5

с) 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.

```
def task4(file="out.json"):
    import pandas as pd

    with open(file, encoding='utf-8') as inputfile:
        df = pd.read_json(inputfile)

    df.to_csv('out.csv', encoding='utf-8', index=False)
```

## 6 Вывод

По ходу выполнения данной лабораторной работы, я узнал о форматах представления данных (JSON,XML,CSV), а так же научился парсить данные форматы, как с помощью готовых решений, так и самостоятельно.



## 7 Список литературы

- Лямин А.В., Череповская Е.Н. Объектно-ориентированное программирование. Компьютерный практикум. – СПб: Университет ИТМО, 2017. – 143 с. – Режим доступа: <https://books.ifmo.ru/file/pdf/2256.pdf>.
- Пишем изящный парсер на Питоне - Режим доступа: <https://habr.com/en/post/309242/>