Министерство науки и высшего образования Российской Федерации 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 №2

Группа: P3112

Выполнил: Балин А. А.

Проверил: к. т. н., преподаватель Белозубов А. В.

# Введение

В этой лабораторной работе я хочу научиться переводить данные из одного формата в другой разными способами.

# Порядок выполнения работы

1. Определить номер варианта как остаток деления на 36 порядкового номера в списке группы в ISU. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь.
2. 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.
3. Написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый (нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов).
4. Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов. Переписать исходный код, применив найденные библиотеки. Регулярные выражения также нельзя использовать. Сравнить полученные результаты и объяснить их сходство/различие.
5. Переписать исходный код, добавив в него использование регулярных выражений. Сравнить полученные результаты и объяснить их сходство/различие.
6. Используя свою исходную программу из обязательного задания, программу из дополнительного задания №1 и программу из дополнительного задания №2, сравнить стократное время выполнения парсинга + конвертации в цикле. Проанализировать полученные результаты и объяснить их сходство/различие.
7. 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. Проанализировать полученные результаты, объяснить особенности использования формата.
8. Проверить, что все пункты задания выполнены и выполнены верно.

# Выполнение заданий

№1

Мой порядковый номер в списке: 2, остаток от деления 2 на 36: 2, следовательно, номер моего варианта: 2

№2

Для моего варианта исходный формат: JSON, результирующий формат: YAML, день недели: понедельник

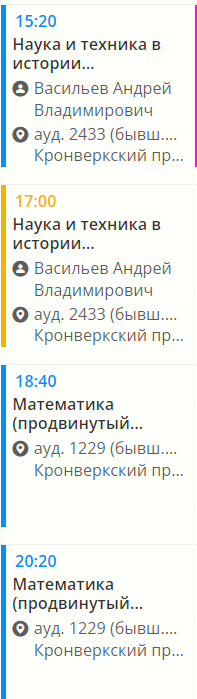


Рисунок 1. Моё расписание на понедельник

Сформируем объект JSON на основе расписания (Рисунок 1):

{

    "Первая пара": {

        "Название": "Наука и техника в истории цивилизации",

        "Преподаватель": "Васильев Андрей Владимирович",

        "Аудитория": "2433",

        "Время начала": "15:20",

        "Время окончания": "16:50",

        "Корпус": "Кронверкский проспект, 49",

        "Тип": "Лекция"

    },

    "Вторая пара": {

        "Название": "Наука и техника в истории цивилизации",

        "Преподаватель": "Васильев Андрей Владимирович",

        "Аудитория": "2433",

        "Время начала": "17:00",

        "Время окончания": "18:30",

        "Корпус": "Кронверкский проспект, 49",

        "Тип": "Практика"

    },

    "Третья пара": {

        "Название": "Математика (продвинутый уровень)",

        "Преподаватель": "Абжандадзе Зураб Леванович",

        "Аудитория": "1229",

        "Время начала": "18:40",

        "Время окончания": "20:10",

        "Корпус": "Кронверкский проспект, 49",

        "Тип": "Лекция"

    },

    "Четвертая пара": {

        "Название": "Математика (продвинутый уровень)",

        "Преподаватель": "Абжандадзе Зураб Леванович",

        "Аудитория": "1229",

        "Время начала": "20:20",

        "Время окончания": "21:50",

        "Корпус": "Кронверкский проспект, 49",

        "Тип": "Лекция"

    }

}

Этот объект будет лежать в файле Monday.json

№3

Код программы:

f = open('./Monday.json', 'r', encoding='utf-8')

F = f.readlines()

f.close()

s = ''

for S in F:

    s += S.strip()

result = ''

tab = -1

i = 0

name = True

while i < len(s):

    if s[i] == '{':

        tab += 1

        name = True

        i += 1

        if tab != 0:

            result += '\n'

    elif s[i] == '\"':

        if name:

            t = ''

            i += 1

            while s[i] != '\"':

                t += s[i]

                i += 1

            i += 1

            result += "  "\*tab + t

            name = False

        else:

            t = ''

            i += 1

            while s[i] != '\"':

                t += s[i]

                i += 1

            i += 1

            q = True

            for x in t:

                if x.isalpha():

                    q = False

                    break

            if q:

                result += "\'"+t+"\'"

            else:

                result += t

            name = True

    elif s[i] == ':':

        result += ': '

        i += 1

    elif s[i] == ',':

        result += '\n'

        i += 1

    elif s[i] == '}':

        tab -= 1

        i += 1

    else:

        i += 1

f = open("./Monday.yaml", 'w', encoding='utf-8')

f.write(result)

f.close()

И результат программы (содержимое файла Monday.yaml):

Первая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время начала: '15:20'

  Время окончания: '16:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Вторая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время начала: '17:00'

  Время окончания: '18:30'

  Корпус: Кронверкский проспект, 49

  Тип: Практика

Третья пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время начала: '18:40'

  Время окончания: '20:10'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Четвертая пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время начала: '20:20'

  Время окончания: '21:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Немного изменим входной файл, объединив время начала и конца в новый объект «время», например:

"Время": {

      "Начало": "15:20",

      "Конец": "16:50"

    }

Вывод программы:

Первая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время:

    Начало: '15:20'

    Конец: '16:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Вторая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время:

    Начало: '17:00'

    Конец: '18:30'

  Корпус: Кронверкский проспект, 49

  Тип: Практика

Третья пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время:

    Начало: '18:40'

    Конец: '20:10'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Четвертая пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время:

    Начало: '20:20'

    Конец: '21:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

№4

В своей программе для этого задания я использовал библиотеки json и pyyaml. Код программы:

import json, yaml

js = json.loads(open('./Monday.json',encoding='utf8').read())

s = yaml.safe\_dump(js, default\_flow\_style=False, allow\_unicode=True, sort\_keys=False)

open('./Monday\_lib.yaml', 'w', encoding='utf8').write(s)

Вывод программы (содержимое файла Monday\_lib.yaml):

Первая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время:

    Начало: '15:20'

    Конец: '16:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Вторая пара:

  Название: Наука и техника в истории цивилизации

  Преподаватель: Васильев Андрей Владимирович

  Аудитория: '2433'

  Время:

    Начало: '17:00'

    Конец: '18:30'

  Корпус: Кронверкский проспект, 49

  Тип: Практика

Третья пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время:

    Начало: '18:40'

    Конец: '20:10'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Четвертая пара:

  Название: Математика (продвинутый уровень)

  Преподаватель: Абжандадзе Зураб Леванович

  Аудитория: '1229'

  Время:

    Начало: '20:20'

    Конец: '21:50'

  Корпус: Кронверкский проспект, 49

  Тип: Лекция

Как можно заметить, внешне результаты №3 и №4 неотличимы, однако на самом деле при использовании библиотек в конце файла получается пустая строка, чего нет при использовании моего алгоритма. Разница заключается лишь в условиях переноса строки: я это делаю после запятой, библиотека ставит перенос после символа «}». Сходство объясняется тем, что при конвертации в любом случае нужно соблюдать разметку YAML.

№5