Национальный исследовательский университет компьютерных технологий, механики и оптики

Факультет ПИиКТ

Лабораторная работа №3

«Исследование языков разметки документов»

Вариант №11

Работу выполнил: Кулаков Никита

Группа: P3130

Преподаватель: Балакшин П.В.

Город: Санкт-Петербург

2020 год

Задание №1.

*Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного. (исходный формат в Варианте №11 – перевести файл в XML)*

<?xml version="1.0"?>

<rasp\_day\_mobile>

<lesson>

<time>08:20-09:50</time>

<parity>нечетная неделя</parity>

<auditory>101 АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</subject>

<parity\_2>нечетная неделя</parity\_2>

<teacher>Перминов Илья Валентинович</teacher>

<format>Очно - дистанционный</format>

</lesson>

<lesson>

<time>08:20-09:50</time>

<parity>четная неделя</parity>

<auditory>303/2 (УСЛ) АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ИНФОРМАТИКА(ЛАБ)</subject>

<parity\_2>четная неделя</parity\_2>

<teacher>Калинин Игорь Владимирович</teacher>

<format>Очно - дистанционный</format>

</lesson>

<lesson>

<time>10:00-11:30</time>

<parity>четная неделя</parity>

<auditory>303/2 (УСЛ) АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ИНФОРМАТИКА(ЛАБ)</subject>

<parity\_2>четная неделя</parity\_2>

<teacher>Калинин Игорь Владимирович</teacher>

<format>Очно - дистанционный</format>

</lesson>

<lesson>

<time>10:00-11:30</time>

<parity>нечетная неделя</parity>

<auditory>101 АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)</subject>

<parity\_2>нечетная неделя</parity\_2>

<teacher>Перминов Илья Валентинович</teacher>

<format>Очно - дистанционный</format>

</lesson>

<lesson>

<time>11:40-13:10</time>

<parity>четная неделя</parity>

<auditory>101 АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ПРОГРАММИРОВАНИЕ(ЛАБ)</subject>

<parity\_2>четная неделя</parity\_2>

<teacher>Николаев Владимир Вячеславович</teacher>

<format>Очно - дистанционный</format>

</lesson>

<lesson>

<time>13:30-15:00</time>

<parity>четная неделя</parity>

<auditory>101 АУД.</auditory>

<corps>Кронверкский пр., д. 49, лит. А</corps>

<subject>ПРОГРАММИРОВАНИЕ(ЛАБ)</subject>

<parity\_2>четная неделя</parity\_2>

<teacher>Николаев Владимир Вячеславович</teacher>

<format>Очно - дистанционный</format>

</lesson>

</rasp\_day\_mobile>

Задание №2.

*Написать код на языке Python 3.x, который бы осуществлял парсинг и конвертацию исходного файла в другой. (У меня по заданию это JSON файл)*

def changer(lines):

r\_count = line.count('<')

l\_count = line.count('>')

sl\_count = line.count('/')

if r\_count == 1 and l\_count == 1:

if sl\_count == 0:

return '\t\t{'

return '\t\t},'

lines = lines.replace('<', '"', 1).replace('>', '": "', 1)

line2 = '\t'

for tmp in lines:

if tmp != '<':

line2 += tmp

else:

return line2 + '",'

file = open('daytime.xml’, encoding='utf-8')

file.readline()

line = file.readline()

print(‘{‘, line.replace('<', '"').replace('>', '": {'))

line = file.readline()

print(line.replace('<', '"').replace('>', '": ['))

k = 2 # костыль в лице проверки на кратность, чтобы убрать ненужные запятые и вывести конец программы (можно сначала подсчитать кол-во уроков и кол-во параметров, чтобы не писать костыли, но я решил не реализовывать)

while line != '' and k != 61:

line = changer(line)

if k != 0 and k % 10 == 0:

line = line[:-1]

print(line)

line = file.readline()

k += 1

print('\t\t}\n\t]\n}}')

Результат:

"rasp\_day\_mobile": {

"lesson": [

{

"time": "08:20-09:50",

"parity": "нечетная неделя",

"auditory": "101 АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)",

"parity\_2": "нечетная неделя",

"teacher": "Перминов Илья Валентинович",

"format": "Очно - дистанционный"

},

{

"time": "08:20-09:50",

"parity": "четная неделя",

"auditory": "303/2 (УСЛ) АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ИНФОРМАТИКА(ЛАБ)",

"parity\_2": "четная неделя",

"teacher": "Калинин Игорь Владимирович",

"format": "Очно - дистанционный"

},

{

"time": "10:00-11:30",

"parity": "четная неделя",

"auditory": "303/2 (УСЛ) АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ИНФОРМАТИКА(ЛАБ)",

"parity\_2": "четная неделя",

"teacher": "Калинин Игорь Владимирович",

"format": "Очно - дистанционный"

},

{

"time": "10:00-11:30",

"parity": "нечетная неделя",

"auditory": "101 АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)",

"parity\_2": "нечетная неделя",

"teacher": "Перминов Илья Валентинович",

"format": "Очно - дистанционный"

},

{

"time": "11:40-13:10",

"parity": "четная неделя",

"auditory": "101 АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",

"parity\_2": "четная неделя",

"teacher": "Николаев Владимир Вячеславович",

"format": "Очно - дистанционный"

},

{

"time": "13:30-15:00",

"parity": "четная неделя",

"auditory": "101 АУД.",

"corps": "Кронверкский пр., д. 49, лит. А",

"subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)",

"parity\_2": "четная неделя",

"teacher": "Николаев Владимир Вячеславович",

"format": "Очно - дистанционный"

}

]

}

Задание №3

*a) Найти готовые библиотеки, осуществляющие аналогичный парсинг и конвертацию файлов.*

*b) Сравнить полученные результаты и объяснить их сходство/различие.*

* библиотека xmltodict (она переводит из xml в словарь, а уже потом нужно будет осуществлять перевод из словаря в json), minidom, однако я эту библиотеку не использовал

Программа с использованием библиотеки xmltodict:

import xmltodict, json

file = open('daytime.xml', encoding='utf-8')

now = file.read()

wow = xmltodict.parse(now)

kek = json.dumps(wow, ensure\_ascii=False)

print(kek)

Результат выводится в одну строку в виде файла xml:

{"rasp\_day\_mobile": {"lesson": [{"time": "08:20-09:50", "parity": "нечетная неделя", "auditory": "101 АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)", "parity\_2": "нечетная неделя", "teacher": "Перминов Илья Валентинович", "format": "Очно - дистанционный"}, {"time": "08:20-09:50", "parity": "четная неделя", "auditory": "303/2 (УСЛ) АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ИНФОРМАТИКА(ЛАБ)", "parity\_2": "четная неделя", "teacher": "Калинин Игорь Владимирович", "format": "Очно - дистанционный"}, {"time": "10:00-11:30", "parity": "четная неделя", "auditory": "303/2 (УСЛ) АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ИНФОРМАТИКА(ЛАБ)", "parity\_2": "четная неделя", "teacher": "Калинин Игорь Владимирович", "format": "Очно - дистанционный"}, {"time": "10:00-11:30", "parity": "нечетная неделя", "auditory": "101 АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ОСНОВЫ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ(ЛАБ)", "parity\_2": "нечетная неделя", "teacher": "Перминов Илья Валентинович", "format": "Очно - дистанционный"}, {"time": "11:40-13:10", "parity": "четная неделя", "auditory": "101 АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)", "parity\_2": "четная неделя", "teacher": "Николаев Владимир Вячеславович", "format": "Очно - дистанционный"}, {"time": "13:30-15:00", "parity": "четная неделя", "auditory": "101 АУД.", "corps": "Кронверкский пр., д. 49, лит. А", "subject": "ПРОГРАММИРОВАНИЕ(ЛАБ)", "parity\_2": "четная неделя", "teacher": "Николаев Владимир Вячеславович", "format": "Очно - дистанционный"}]}}

Различие между моим выводом и выводом программы, использующей библиотеки xmltodict и json:

Моя программа выводит файл в том формате, в котором изначально был xml файл, парсинг и кодирование в json заключается в обычной замене символов, т.к у них почти одинаковый синтаксис. Однако программа с библиотеками сначала осуществляет парсинг, переводя файл из xml в словарь, а потом словарь преобразует в json файл, именно поэтому вторая программа выводит файл в строчку, т.к сам словарь записывается в строку.

Задание №4:

*a) Используя свою программу и найденные готовые библиотеки, сравнить десятикратное время выполнения парсинга + конвертации в цикле.*

*b) Проанализировать полученные результаты и объяснить их сходство/различие.*

Для выполнения данного задания я использовал библиотеку time и ее методы. Примерная структура указана ниже:

import time

start\_time = time.time()

main() # тело функции

print(time.time() - start\_time, ‘seconds’)

Для проведения десятикратного прохода по каждой программе возьмем все тело главной функции в цикл for (for I in range(10)). В результате получим: # в таком случае я считаю десятикратное открытие файла

1) Для моей программы среднее значение: 0.018 секунд

2) Для программы с использованием библиотек: 0.015 секунд

# В случае возникновения вопросов могу показать значения time через отладку

Вывод: в ходе проделанной работы я постарался разобраться с библиотеками xmltodict, json, time, очень много времени потратил на debug кода ~~особенно с библиотекой time~~. Кроме того, научился отыскивать html код сайта в браузере да и получил представление о таких языках как html, xml, json. Вдобавок, если мы учтем задания для подготовки, то я узнал многое из этого, пока искал ответы на данные вопросы.