НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО

Факультет ПИиКТ

Информатика

Лабораторная работа № 3

«Регулярные выражения и языки разметки документов. Python»

Выполнил студент:

Двоеглазова Наталья Николаевна

Группа №P3123

Преподаватель:

Болдырева Елена Александровна

г. Санкт-Петербург

2023

Содержание

[Основные этапы вычислений: 3](#_Toc149524218)

[Задание №1 3](#_Toc149524219)

[Задание №2 5](#_Toc149524220)

[Задание №3 7](#_Toc149524221)

[Вывод: 11](#_Toc149524222)

[Список литературы: 11](#_Toc149524223)

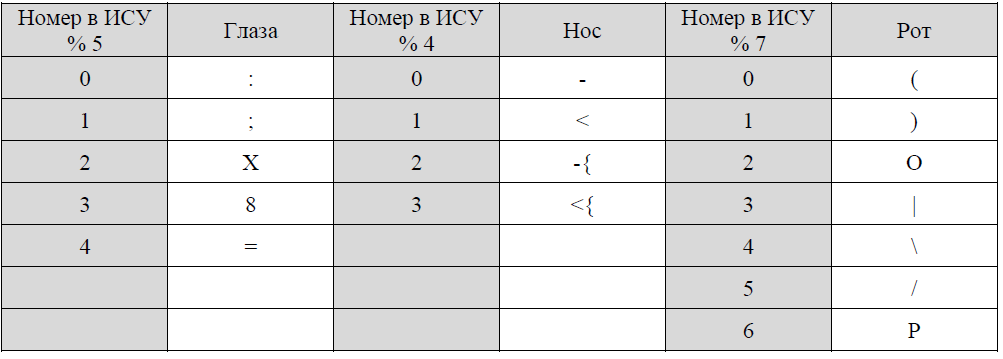
# Основные этапы вычислений:

# Задание №1

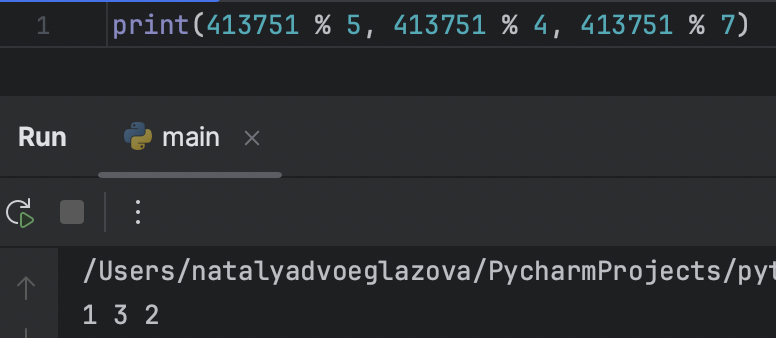
1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.

2) Для своей программы придумайте минимум 5 тестов. Каждый тест является отдельное сущностью, передаваемой регулярному выражению для обработки. Для каждого теста необходимо самостоятельно (без использования регулярных выражений) найти правильные ответ. После чего сравнить ответ, выданные программой, и полученные самостоятельно.

3) Программа должна считать количество смайликов определённого вида (вид смайлика описан в таблице вариантов) в предложенном тексте. Все смайлики имеют такую структуру:



Номер ИСУ: 413751



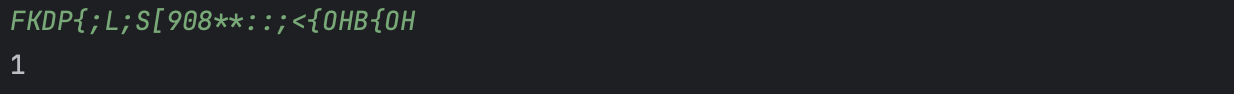
= > Мой смайлик ;<{O

Код:

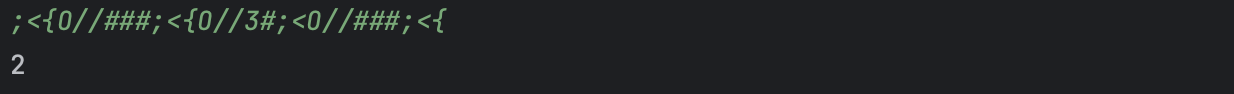
Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Тест №1:



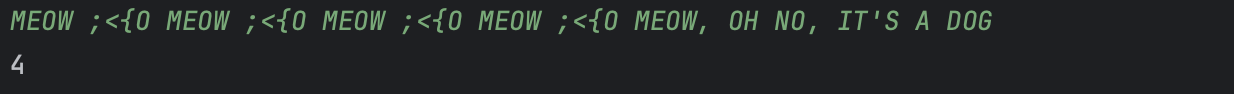
Тест №2:



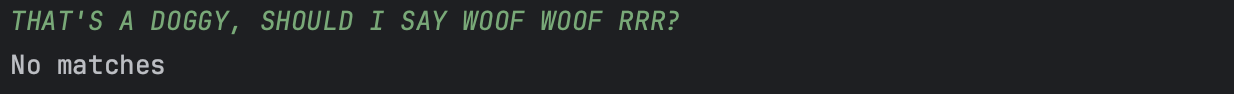
Тест №3:



Тест №4:



Тест №5:



С ручным подсчетом совпало ☺

# Задание №2

413751 % 6 = 3 вариант

Дан текст. Необходимо найти в нём любой̆ фрагмент, где сначала идёт слово «ПИиКТ», затем не более 4 слов, и после этого идёт слово «ИТМО». Для простоты будем считать словом любую последовательность букв, цифр и знаков «\_» (то есть символов \w).

Пример:

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

1) Реализуйте программный продукт на языке Python, используя регулярные выражения по варианту, представленному в таблице.

2) Для своей программы придумайте минимум 5 тестов.

3) Протестируйте свою программу на этих тестах.

Поскольку я с факультета «ПИиКТ», то буду использовать его вместо «ВТ» ☺

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

Если мы хотим рассмотреть все возможные символы в качестве слов, то воспользуемся этим регулярным выражением (релевантный пример №3):

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

# Задание №3

1. Определить номер варианта как остаток деления номера в ИСУ на 36. В случае, если в данный день недели нет занятий, то увеличить номер варианта на восемь.
2. Изучить форму Бэкуса-Наура.
3. Изучить особенности протоколов и форматов обмена информацией между системами: JSON, YAML, XML.
4. Понять устройство страницы с расписанием для своей группы: https://itmo.ru/schedule/0/P3123/schedule.htm
5. Исходя из структуры расписания конкретного дня, сформировать файл с расписанием в формате, указанном в задании в качестве исходного.
6. **Обязательное задание**: написать программу на языке Python 3.x, которая бы осуществляла парсинг и конвертацию исходного файла в новый.
7. Нельзя использовать готовые библиотеки, в том числе регулярные выражения в Python и библиотеки для загрузки XML-файлов.

Вариант: 413751%36 = 3 (понедельник, 5 пар ☹….)

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

8. Дополнительное задание №1

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.

9. Дополнительное задание №2

a) Переписать исходный код, добавив в него использование регулярных выражений.

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

Исходный файл:

Изображение выглядит как текст, снимок экрана, Шрифт, документ

Автоматически созданное описание

Программа:

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Вывод:

Изображение выглядит как текст, программное обеспечение, Мультимедийное программное обеспечение, компьютер

Автоматически созданное описание

Дополнительное задание №1:

Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание

# Вывод:

В ходе работы мы изучили регулярные выражения, способы их использования и на практике отработали различные задачи. Также были изучены системы JSOM, YAML, XML. Была написана программа для парсинга и конвертации исходного JSON файла в YAML. Более того научилась устанавливать и подключать библиотеки к Python.

# Список литературы:

1. Лямин А.В., Череповская Е.Н. Объектно-ориентированное программирование. Компьютерный практикум. – СПб: Университет ИТМО, 2017. – 143 с.

Вопросы:

1. В чём разница между Markup и Markdown?

**Markup**-это просто способ обеспечения функциональности выше обычного текста. Например: форматирование, ссылки, изображения и т.д. **Markup**-это общий термин для форматирования содержимого, например HTML, но **markdown**-это библиотека, которая генерирует HTML **markup**.

2) В чём заключается особенность PROTOBUF по сравнению с другими форматами?

Фишка протобaфа в том, что, при сериализации целых чисел, по-умолчанию, используется **формат** переменной длины (varint), который занимает меньше места для небольших положительных чисел. Протобаф добавляет в бинарный поток номер поля и его тип, что увеличивает итоговый размер.

это **формат** сериализации данных, разработанный компанией Google. Он эффективно и компактно хранит структурированные данные в двоичной форме, что позволяет быстрее передавать их по сети.

3) Чем формат CSV отличается от формата TSV?

**CSV** и **TSV** имеют схожий **формат**, но использование разделителей символов делает их сильно **отличающимися** друг от друга. В **CSV**-**файлах** объекты данных разделены запятыми, а в **TSV**-**файлах**используются табуляции в качестве разделителя.

4) Чем обусловлено постоянное появление новых форматов представления данных?

5) Каким образом в формате XML представляются символы ‘>’ и ‘<’?

Для ограничения тегов в разметке **XML** используются угловые скобки: тег начинается со **знака** "меньше" (<) и завершается **знаком** "больше" (>). Но необходимо помнить, что в отличие от HTML вся разметка **XML** чувствительна к регистру **символов**, это касается как имен тегов, так и значений атрибутов.

< &lt;

> &gt;

6) Что такое сериализация данных?

Сериализация — это **процесс преобразования объекта в поток байтов для сохранения или передачи в память, базу данных или файл**. Эта операция предназначена для того, чтобы сохранить состояния объекта для последующего воссоздания при необходимости. Обратный процесс называется десериализацией.

7) Каким образом в YAML обозначаются комментарии?

Comments in YAML are denoted by the "#" symbol and are ignored by the interpreter when the code is executed.

8) Пояснить, как в языке разметки Markdown создать заголовки разных

уровней, оформить код, вывести полужирный, курсивный и зачеркнутый текст?

9) Какие форматы обмена данных используются в современных популярных мессенджерах (Viber, WhatsApp, Telegram и т.д.)?

Тг – **MTProto,** CSV, TSV и XML (RSS/ATOM) - facebook

10) Как расшифровывается аббревиатура SVG?

**Scalable** **Vector** **Graphics** (**SVG**) is an XML-based vector image format for defining two-dimensional graphics, having support for interactivity and animation. Масштабируемая векторная графика

11) Привести пример использования в языке HTML тега, который создаёт гиперссылку на url.

ег <a> превращает любой объект в ссылку.

<a href="https://www.yandex.ru/">Яндекс</a>

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

12) Какие две структуры может представлять собой в закодированном виде JSON-текст?

Объекты и массивы

1. Набор пар *ключ:значение*. В различных языках это реализовано как объект, запись, структура, словарь, хэш-таблица, список с ключом или ассоциативный массив. Ключом может быть только строка, значением — любая форма.
2. Упорядоченный набор значений. Во многих языках это реализовано как массив, вектор, список или последовательность.