Date
14/02/22

Information Retrieval - 8

Aim - Write a program to parse XML text

Description

1) An xml Parse - The XML Dom (Document object Model) defines the properties and methods for accessing and editing XML. However, before an xml document can be accessed, it must be loaded into an XML DOM object. All modern browsers have a built in xml parser than that can convert text into an xml DOM object.

2) DOM - The Dom defines a standard for accessing and manipulating documents. The W3C documents object Model (DOM) is a platform and language. Natural interface that allows programs and scripts to dynamically access & update the content, structure and style of document.

3) xml-dom-minidom - It is a minimal implementation of all the document object model interface, with an API similar to that in other languages. It is intended to the simpler than the full DOM and also significant smaller. Uses who are not already proficient with the DOM should consider using the XML tree. Efficient tree module for their xml processing instead.

```
<collection shelf="car">

<car title="Turismo R">
    <typec>Supercar</typec>
    <year>2014</year>
    <brand>Grotti</brand>
    <tier>3</tier>
    <description>based on ferrai laferrai fxx</description>
</car>

<car title="Autarch">
    <typec>Supercar</typec>
    <year>2019</year>
    <brand>Overflod</brand>
    <tier>2</tier>
    <description>based on SCG 003</description>
</car>

<car title="Vagner">
    <typec>Supercar</typec>
    <year>2018</year>
    <brand>Dewbauchee</brand>
    <tier>1</tier>
    <description>based on aston martin valkyrie</description>
</car>


</collection>
```

Information Retreival Pratical-8.py - E:/fffiiles/college pracs and projects/IR/Information Retreival Pratical-8.py (3.9.6)

File   Edit   Format   Run   Options   Window   Help

```python
from xml.dom.minidom import parse
import xml.dom.minidom

# Open XML document using minidom parser
DOMTree = xml.dom.minidom.parse("cars.xml")

collection = DOMTree.documentElement
if collection.hasAttribute("shelf"):
    print("Root element :",collection.getAttribute("shelf"))
print("Neeraj Appari T073")

# Get all the car in the collection
cars = collection.getElementsByTagName("car")

# Print detail of each movie.
for car in cars:
    print("*****CARS*****")
    if car.hasAttribute("title"):
        print("Title: %s" % car.getAttribute("title"))

    typec = car.getElementsByTagName('typec')[0]
    print("Type: %s" % typec.childNodes[0].data)
    brand = car.getElementsByTagName('brand')[0]
    print("brand: %s" % brand.childNodes[0].data)
    tier = car.getElementsByTagName('tier')[0]
    print("tier: %s" % tier.childNodes[0].data)
    description = car.getElementsByTagName('description')[0]
    print("Description: %s" % description.childNodes[0].data)
```

```
Python 3.9.6 (tags/v3.9.6:db3ff76, Jun 28 2021, 15:26:21) [MSC v.1929 64 bit (AM
D64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: E:/fffiiles/college pracs and projects/IR/Information Retreival Prati
cal-8.py
Root element : car
Neeraj Appari T073
****CARS****
Title: Turismo R
Type: Supercar
brand: Grotti
tier: 3
Description: based on ferrai laferrai fxx
****CARS****
Title: Autarch
Type: Supercar
brand: Overflod
tier: 2
Description: based on SCG 003
****CARS****
Title: Vagner
Type: Supercar
brand: Dewbauchee
tier: 1
Description: based on aston martin valkyrie
>>>
```