PFAI_Lab_7

October 15, 2024

0.0.1 Note: To work on colab we have to mount google drive

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

1 1. XML

- Extensible Markup Language
- Designed for storing and transporting data
- It provides a way to structure data using custom tags and elements in a hierarchical format.
- Defines a set of rules for encoding documents in a format that is both human-readable and machine-readable ### Syntax:

```
<tag>Content</tag>
```

Or

```
<parent>
    <child>
      content
    </child>
</parent>
```

Example:

```
    <book category="fiction">
        <title lang="en">Introduction to XML</title>
        <author>John Doe</author>
        <price>19.95</price>
    </book>
    <book category="non-fiction">
        <title lang="en">Python Programming</title>
        <author>Jane Smith</author>
        <price>29.99</price>
    </book>
</bookstore>
```

1.0.1 1.1 Parsing and Reading XML

[3]:
```
import os
import xml.etree.ElementTree as ET
xml_dir = '/content/drive/MyDrive/Colab Notebooks/CTIDataset'
for xml_file in os.listdir(xml_dir):
    if xml_file.endswith('.xml'):
        try:
            # Parse the XML file
            tree = ET.parse(os.path.join(xml_dir, xml_file))
            root = tree.getroot()
            print(root)
        except Exception as e:
          print(f"Error processing {xml_file}: {str(e)}")
```

```
<Element 'CTIMinerDataset' at 0x7ff90886f240>
<Element 'CTIMinerDataset' at 0x7ff90886f060>
<Element 'CTIMinerDataset' at 0x7ff9083d79c0>
<Element 'CTIMinerDataset' at 0x7ff9083f3880>
<Element 'CTIMinerDataset' at 0x7ff9095ddbc0>
<Element 'CTIMinerDataset' at 0x7ff908673880>
<Element 'CTIMinerDataset' at 0x7ff9083d7a60>
<Element 'CTIMinerDataset' at 0x7ff8f41858f0>
<Element 'CTIMinerDataset' at 0x7ff8d77c5850>
<Element 'CTIMinerDataset' at 0x7ff8d759c770>
<Element 'CTIMinerDataset' at 0x7ff8d6d53970>
<Element 'CTIMinerDataset' at 0x7ff8d6d53e20>
<Element 'CTIMinerDataset' at 0x7ff8d56204f0>
<Element 'CTIMinerDataset' at 0x7ff9081d0270>
<Element 'CTIMinerDataset' at 0x7ff8d5c65f80>
<Element 'CTIMinerDataset' at 0x7ff8d5b510d0>
<Element 'CTIMinerDataset' at 0x7ff8d6c3a2f0>
<Element 'CTIMinerDataset' at 0x7ff8d454d670>
<Element 'CTIMinerDataset' at 0x7ff8d6a651c0>
<Element 'CTIMinerDataset' at 0x7ff8d5fcbd80>
```

1.0.2 1.2 Accessing Child Elements

[ ]:
```
for child in root:
  print(child.tag)
```

1.0.3 1.3 Finding Elements

[5]:
```
event_data = root.find('Event')
print(event_data.find('id').text)
```

```
1970
```

[ ]:
```
find_ids = root.findall('Event/id')
for id in find_ids:
    print(id.text)
```

1.0.4 1.4 Write XML file

Example 1:

[9]:
```
# Create the root element
root = ET.Element("AILab")

# Create child elements
child1 = ET.SubElement(root, "Task1")
subchild1=ET.SubElement(child1,"Data Analysis")
child2 = ET.SubElement(root, "Task2")
child1.text = "Data for child 1"
subchild1.text="Mean is 2.4"
child2.text = "Data for child 2"

# Create an ElementTree object
tree = ET.ElementTree(root)

# Write the XML data to a file
tree.write("/content/drive/MyDrive/Colab Notebooks/out_ai.xml")

print("XML data has been written to output.xml.")
```

```
XML data has been written to output.xml.
```

Example 2:

- Adding attributes

[11]:
```
# Create the root element
root = ET.Element("root")

# Create child elements with attributes
child1 = ET.SubElement(root, "child1")
child1.set("attribute", "value1")  # Adding an attribute to child1 element
child1.text = "This is child 1"

child2 = ET.SubElement(root, "child2")
child2.set("attribute", "value2")  # Adding an attribute to child2 element
child2.text = "This is child 2"

# Create an ElementTree object
tree = ET.ElementTree(root)
```

```
# Write to XML file
tree.write("/content/drive/MyDrive/Colab Notebooks/out.xml", encoding="utf-8",␣
 ↪xml_declaration=True)
```

Task 1. Read the same dataset using beautiful soup and show all tags 2. Using XML tree extract the values and categories (data)