Sathish_K_Rajendiran_6.4_XML-RSS

August 5, 2020

6.4 XML, DOM, and Element Tree Week 6 · Semi-structured Data

```
Name: Sathish Kumar Rajendiran
Task: Class work - XML, DOM and Element Tree
Date: 8/5/2020
```

```python
[2]: #import libraries
     from urllib import request
     from urllib.error import URLError,HTTPError

     import xml.etree.ElementTree as etree
     import io
     import pandas as pd
```

```python
[3]: #define url and import data using request and urlopen
     url = 'http://rss.cnn.com/rss/cnn_topstories.rss.xml'
     #error handling
     try:
         resp = request.urlopen(url).read().decode('utf8')
     except URLError as e:
         if hasattr(e, 'reason'):
             print('We failed to reach a server.')
             print('Reason: ', e.reason)
         elif hasattr(e, 'code'):
             print('The server couldn\'t fulfill the request.')
             print('Error code: ', e.code)
     else:
         print(resp[:500])
```

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" media="screen"
href="/~d/styles/rss2full.xsl"?><?xml-stylesheet type="text/css" media="screen"
href="http://rss.cnn.com/~d/styles/itemcontent.css"?><rss
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:content="http://purl.org/rss/1.0/modules/content/"
xmlns:atom="http://www.w3.org/2005/Atom"
xmlns:media="http://search.yahoo.com/mrss/"
xmlns:feedburner="http://rssnamespace.org/feedburner/ext/1.0"
version="2.0"><channel><ti
```

```python
[4]: #define tree
     xmlfile = io.StringIO(resp)
     tree = etree.parse(xmlfile)
     type(tree)
```

```
[4]: xml.etree.ElementTree.ElementTree
```

```python
[7]: #list the top 10
     alltitles = tree.findall('.//title')
     len(alltitles)

     for title in alltitles[:10]:
         print(title.text)
```

```
CNN.com - RSS Channel - HP Hero
CNN.com - RSS Channel - HP Hero
The coronavirus pandemic has Americans spending more at the supermarket than
they have in years
It's harder to find sale items at the grocery store
Why Fauci thinks the US doesn't have to lockdown
Groceries were hard to find for millions. Now it's getting even worse
Covid-19 patients may develop skin rashes and discoloration, studies find
Second grader tests positive after first day of school
Parents and teachers torn about kids going back to school
Mississippi has the nation's highest test positivity rate
```