Chapter 6 - Data Sourcing via Web

Segment 3 - Data parsing

```
from bs4 import BeautifulSoup

import urllib
import urllib.request
import re

'''from bs4 import BeautifulSoup
import urllib
import urllib.request
import re'''
```

```
'from bs4 import BeautifulSoup\nimport urllib\nimport urllib.request\nimport re'
```

```
with urllib.request.urlopen('https://raw.githubusercontent.com/BigDataGal/Data-Mania-Demos/master/IoT-2018.html') as response:
    html = response.read()

#with urllib.request.urlopen('.html') as response:
    #html= response.read()



soup = BeautifulSoup(html, "lxml")
type(soup)
#BeautifulSoup(html,'lxml')
```

```
bs4.BeautifulSoup
```

Parsing your data

```
print(soup.prettify()[0:100])
#soup.prettify()[0:1000]
```

```
<html>
 <head>
  <title>
   IoT Articles
  </title>
 </head>
 <body>
  <p class="title">
   <b>
```

Getting data from a parse tree

```
text_only = soup.get_text()
print(text_only)
#soup.get_text()
```

```
    IoT Articles

    2018 Trends: Best New IoT Device Ideas for Data Scientists and Engineers
    It’s almost 2018 and IoT is on the cusp of an explosive expansion. In this article, I offer you a listing of new IoT device id


    It’s almost 2018 and IoT is on the cusp of an explosive expansion. In this article, I offer you a listing of new IoT device id
    Looking Back at My Coolest IoT Find in 2017
    Before going into detail about best new IoT device ideas, here’s the backstory. Last month Ericsson Digital invited me to tour

    It wasn’t until I got to the Ericsson Studio that I became extremely impressed by how far IoT has really come. Relying on the

    This car is connected to Ericsson’s Connected Vehicle Cloud, an IoT platform that manages services for the Smart Cars to which

    To understand how it works, imagine you’re pulling your normal 9-to-5 and you know you need to grab some groceries on your way

    To watch some of the amazing IoT device demos I witnessed at Ericsson Studio, make sure to go watch the videos on this page.
    Future Trends for IoT in 2018
    New IoT device ideas won’t do you much good unless you at least know the basic technology trends that are set to impact IoT ov

    Big Data & Data Engineering: Sensors that are embedded within IoT devices spin off machine-generated data like it’s going out
    Machine Learning Data Science: While a lot of IoT devices are still operated according to rules-based decision criteria, the a
    Blockchain-Enabled Security: Above all else, IoT networks must be secure. Blockchain technology is primed to meet the security

    Best New IoT Device Ideas
    This listing of new IoT device ideas has been sub-divided according to the main technology upon which the IoT devices are buil
```

```
Raspberry Pi IoT Ideas
Using Raspberry Pi as open-source hardware, you can build IoT applications that offer any one of the following benefits:

Enable built-in sensing to build a weather station that measures ambient temperature and humidity
Build a system that detects discrepancies in electrical readings to identify electricity theft
Use IoT to build a Servo that is controlled by motion detection readings
Build a smart control switch that operates devices based on external stimuli. Use this for home automation.
Build a music playing application that enables music for each room in your house
Implement biometrics on IoT-connected devices

Arduino IoT Ideas
There are a number of new IoT device ideas that deploy Arduino as a microcontroller. These include:

Integrate Arduino with Android to build a remote-control RGB LED device.
Connect PIR sensors across the IoT to implement a smart building.
Build a temperature and sunlight sensor system to remotely monitor and control the conditions of your garden.
Deploy Arduino and IoT to automate your neighborhood streetlights.
Build a smart irrigation system based on IoT-connected temperature and moisture sensors built-in to your agricultural plants.

[caption id="attachment_3807" align="aligncenter" width="300"] An IoT Chatbot Tree at the Ericsson Studio[/caption]
Wireless (GSM) IoT Ideas
Several new IoT device ideas are developed around the GSM wireless network. Those are:

Monitor soil moisture to automate agricultural irrigation cycles.
Automate and control the conditions of a greenhouse.
Enable bio-metrics to build a smart security system for your home or office building
Build an autonomously operating fitness application that automatically makes recommendations based on motion detection and hea
Build a healthcare monitoring system that tracks, informs, and automatically alerts healthcare providers based on sensor readi

IoT Automation Ideas
```

Searching and retrieving data from a parse tree

Retrieving tags by filtering with name arguments

```
soup.find_all("li")
#soup.find_all('li')
```

```
[<li><strong>Big Data</strong> &amp;amp; Data Engineering: Sensors that are embedded within IoT devices spin off machine-generated
data like it’s going out of style. For IoT to function, the platform must be solidly engineered to handle big data. Be assured,
that requires some serious data engineering.</li>,
```

```
 <li><strong>Machine Learning</strong> Data Science: While a lot of IoT devices are still operated according to rules-based decision criteria, the age of artificial intelligence is upon us. IoT will increasingly depend on machine learning algorithms to control device operations so that devices are able to autonomously respond to a complex set of overlapping stimuli.</li>,
 <li><strong>Blockchain</strong>-Enabled Security: Above all else, IoT networks must be secure. Blockchain technology is primed to meet the security demands that come along with building and expanding the IoT.</li>,
 <li>Enable built-in sensing to build a weather station that measures ambient temperature and humidity</li>,
 <li>Build a system that detects discrepancies in electrical readings to identify electricity theft</li>,
 <li>Use IoT to build a Servo that is controlled by motion detection readings</li>,
 <li>Build a smart control switch that operates devices based on external stimuli. Use this for home automation.</li>,
 <li>Build a music playing application that enables music for each room in your house</li>,
 <li>Implement biometrics on IoT-connected devices</li>,
 <li>Integrate Arduino with Android to build a remote-control RGB LED device.</li>,
 <li>Connect PIR sensors across the IoT to implement a smart building.</li>,
 <li>Build a temperature and sunlight sensor system to remotely monitor and control the conditions of your garden.</li>,
 <li>Deploy Arduino and IoT to automate your neighborhood streetlights.</li>,
 <li>Build a smart irrigation system based on IoT-connected temperature and moisture sensors built-in to your agricultural plants.</li>,
 <li>Monitor soil moisture to automate agricultural irrigation cycles.</li>,
 <li>Automate and control the conditions of a greenhouse.</li>,
 <li>Enable bio-metrics to build a smart security system for your home or office building</li>,
 <li>Build an autonomously operating fitness application that automatically makes recommendations based on motion detection and heart rate sensors that are embedded on wearable fitness trackers.</li>,
 <li>Build a healthcare monitoring system that tracks, informs, and automatically alerts healthcare providers based on sensor readings that describe a patients vital statistics (like temperature, pulse, blood pressure, etc).</li>,
 <li>Build an IoT device that automatically locates and reports the closest nearby parking spot.</li>,
 <li>Build a motion detection system that automatically issues emails or sms messages to alert home owners of a likely home invasion.</li>,
 <li>Use temperature sensors connected across the IoT to automatically alert you if your home windows or doors have been left open.</li>,
 <li>Use bio-metric sensors to build a smart system that automate security for your home or office building</li>]
```

Retrieving tags by filtering with keyword arguments

```
soup.find_all(id="link 7")
#soup.find_all(id='link 7')
```

```
[<a class="preview" href="http://www.skyfilabs.com/iot-online-courses" id="link 7">SkyFi</a>]
```

Retrieving tags by filtering with string arguments

```
soup.find_all('ol')
#soup.find_all('ol')
```

```
[<ol>
 <li><strong>Big Data</strong> &amp;amp; Data Engineering: Sensors that are embedded within IoT devices spin off machine-generated data like it’s going out of style. For IoT to function, the platform must be solidly engineered to handle big data. Be assured, that requires some serious data engineering.</li>
 <li><strong>Machine Learning</strong> Data Science: While a lot of IoT devices are still operated according to rules-based decision criteria, the age of artificial intelligence is upon us. IoT will increasingly depend on machine learning algorithms to control device operations so that devices are able to autonomously respond to a complex set of overlapping stimuli.</li>
 <li><strong>Blockchain</strong>-Enabled Security: Above all else, IoT networks must be secure. Blockchain technology is primed to meet the security demands that come along with building and expanding the IoT.</li>
 </ol>, <ol>
 <li>Enable built-in sensing to build a weather station that measures ambient temperature and humidity</li>
 <li>Build a system that detects discrepancies in electrical readings to identify electricity theft</li>
 <li>Use IoT to build a Servo that is controlled by motion detection readings</li>
 <li>Build a smart control switch that operates devices based on external stimuli. Use this for home automation.</li>
 <li>Build a music playing application that enables music for each room in your house</li>
 <li>Implement biometrics on IoT-connected devices</li>
 </ol>, <ol>
 <li>Integrate Arduino with Android to build a remote-control RGB LED device.</li>
 <li>Connect PIR sensors across the IoT to implement a smart building.</li>
 <li>Build a temperature and sunlight sensor system to remotely monitor and control the conditions of your garden.</li>
 <li>Deploy Arduino and IoT to automate your neighborhood streetlights.</li>
 <li>Build a smart irrigation system based on IoT-connected temperature and moisture sensors built-in to your agricultural plants.</li>
 </ol>, <ol>
 <li>Monitor soil moisture to automate agricultural irrigation cycles.</li>
 <li>Automate and control the conditions of a greenhouse.</li>
 <li>Enable bio-metrics to build a smart security system for your home or office building</li>
 <li>Build an autonomously operating fitness application that automatically makes recommendations based on motion detection and heart rate sensors that are embedded on wearable fitness trackers.</li>
 <li>Build a healthcare monitoring system that tracks, informs, and automatically alerts healthcare providers based on sensor readings that describe a patients vital statistics (like temperature, pulse, blood pressure, etc).</li>
 </ol>, <ol>
 <li>Build an IoT device that automatically locates and reports the closest nearby parking spot.</li>
 <li>Build a motion detection system that automatically issues emails or sms messages to alert home owners of a likely home invasion.</li>
 <li>Use temperature sensors connected across the IoT to automatically alert you if your home windows or doors have been left open.</li>
 <li>Use bio-metric sensors to build a smart system that automate security for your home or office building</li>
 </ol>]
```

Retrieving tags by filtering with list objects

```
soup.find_all(['ol', 'b'])
```

```
#soup.find_all(['ol','b'])

    [<b>2018 Trends: Best New IoT Device Ideas for Data Scientists and Engineers</b>,
     <ol>
     <li><strong>Big Data</strong> &amp;amp; Data Engineering: Sensors that are embedded within IoT devices spin off machine-generated
    data like it’s going out of style. For IoT to function, the platform must be solidly engineered to handle big data. Be assured,
    that requires some serious data engineering.</li>
     <li><strong>Machine Learning</strong> Data Science: While a lot of IoT devices are still operated according to rules-based
    decision criteria, the age of artificial intelligence is upon us. IoT will increasingly depend on machine learning algorithms to
    control device operations so that devices are able to autonomously respond to a complex set of overlapping stimuli.</li>
     <li><strong>Blockchain</strong>-Enabled Security: Above all else, IoT networks must be secure. Blockchain technology is primed
    to meet the security demands that come along with building and expanding the IoT.</li>
     </ol>,
     <ol>
     <li>Enable built-in sensing to build a weather station that measures ambient temperature and humidity</li>
     <li>Build a system that detects discrepancies in electrical readings to identify electricity theft</li>
     <li>Use IoT to build a Servo that is controlled by motion detection readings</li>
     <li>Build a smart control switch that operates devices based on external stimuli. Use this for home automation.</li>
     <li>Build a music playing application that enables music for each room in your house</li>
     <li>Implement biometrics on IoT-connected devices</li>
     </ol>,
     <ol>
     <li>Integrate Arduino with Android to build a remote-control RGB LED device.</li>
     <li>Connect PIR sensors across the IoT to implement a smart building.</li>
     <li>Build a temperature and sunlight sensor system to remotely monitor and control the conditions of your garden.</li>
     <li>Deploy Arduino and IoT to automate your neighborhood streetlights.</li>
     <li>Build a smart irrigation system based on IoT-connected temperature and moisture sensors built-in to your agricultural
    plants.</li>
     </ol>,
     <ol>
     <li>Monitor soil moisture to automate agricultural irrigation cycles.</li>
     <li>Automate and control the conditions of a greenhouse.</li>
     <li>Enable bio-metrics to build a smart security system for your home or office building</li>
     <li>Build an autonomously operating fitness application that automatically makes recommendations based on motion detection and
    heart rate sensors that are embedded on wearable fitness trackers.</li>
     <li>Build a healthcare monitoring system that tracks, informs, and automatically alerts healthcare providers based on sensor
    readings that describe a patients vital statistics (like temperature, pulse, blood pressure, etc).</li>
     </ol>,
     <ol>
     <li>Build an IoT device that automatically locates and reports the closest nearby parking spot.</li>
     <li>Build a motion detection system that automatically issues emails or sms messages to alert home owners of a likely home
    invasion.</li>
     <li>Use temperature sensors connected across the IoT to automatically alert you if your home windows or doors have been left
    open.</li>
     <li>Use bio-metric sensors to build a smart system that automate security for your home or office building</li>
     </ol>]
```

Retrieving tags by filtering with regular expressions

```
t = re.compile("t")
for tag in soup.find_all(t):
    print(tag.name)
''' t = re.compile('t')
for tag in soup.find_all(t):
    print(tag.name)'''
```

```
 html
 title
 strong
 strong
 strong
 strong
 strong
 strong
```

Retrieving tags by filtering with a Boolean value

```
for tag in soup.find_all(True):
    print(tag.name)
```

```
 html
 head
 title
 body
 p
 b
 p
 br
 br
 h1
 span
 strong
 a
 a
 a
 img
 a
 span
```

```
strong
a
h1
ol
li
strong
li
strong
li
strong
h1
a
a
a
h2
ol
li
li
li
li
li
li
h2
ol
li
li
li
li
li
a
img
h2
ol
li
li
li
li
li
h2
ol
```

Retrieving weblinks by filtering with string objects

```
for link in soup.find_all('a'):
    print(link.get('href'))
```

```
''' for link in soup.find_all('a'):
print(link.get('href'))'''
```

```
http://bit.ly/LPlNDJj
http://www.data-mania.com/blog/m2m-vs-iot/
bit.ly/LPlNDJj
http://mat.se/
http://bit.ly/LPlNDJj
https://click.linksynergy.com/deeplink?id=*JDLXjeE*wk&mid=39197&murl=https%3A%2F%2Fwww.udemy.com%2Ftopic%2Finternet-of-things%2F%
http://www.skyfilabs.com/iot-online-courses
https://www.coursera.org/specializations/iot
bit.ly/LPlNDJj
http://bit.ly/LPlNDJj
```

Retrieving strings by filtering with regular expressions

```
soup.find_all(string=re.compile("data"))
'''soup.find_all(string=recompile('data'))'''
print(soup.find_all(string=re.compile("data")))
```

```
[' & Data Engineering: Sensors that are embedded within IoT devices spin off machine-generated data like it’s going out of style.
```

Colab paid products - Cancel contracts here