

A graphic of a black spider hanging from a web, positioned above a globe. The globe is divided into blue oceans and green continents, with puzzle pieces scattered around it.

Python Web-scraping

Agenda for Today's Session

01 Get a brief understanding of what exactly is Web Scraping



Web-scraping: Usecase

03 Learn about various Python Libraries which are used for web scraping



HTML Basics

05 Scrape data from a website using python libraries beautiful soup and urllib2



Web-scraping: Basics

02

Learn about how web-scraping can be used for different business scenarios

Python Libraries

04

Quick overview of basic HTML Tags

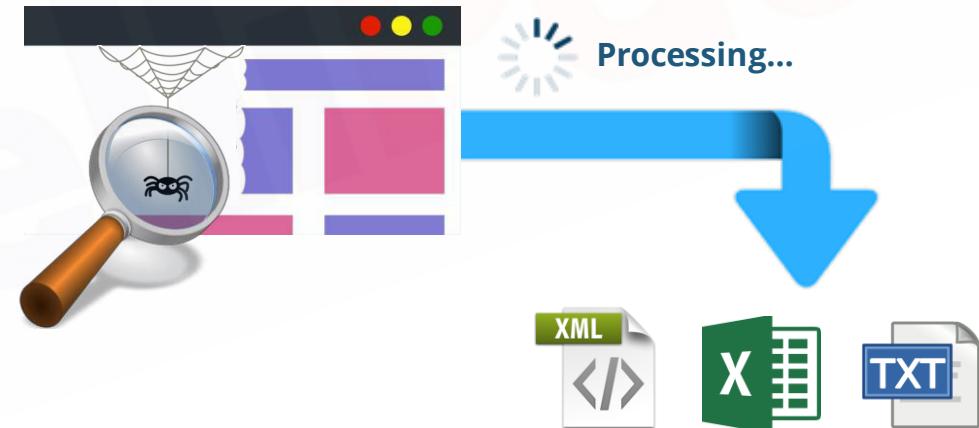
Demo: Scraping a Website

Web-scraping Basics



What is Web-scraping?

Extracting unstructured data from the websites
and transforming it to a structured data...



Web-scraping Basics

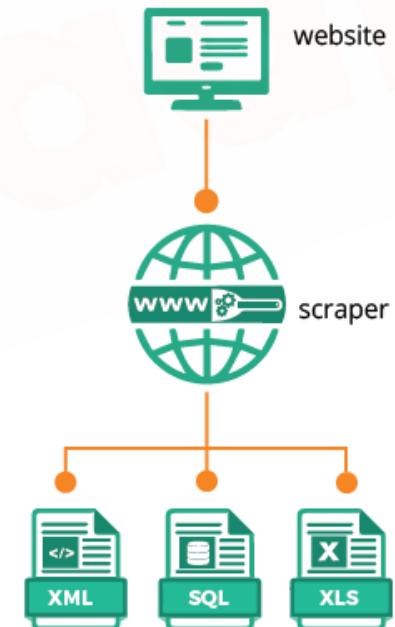


Are they same?

WEB CRAWLER



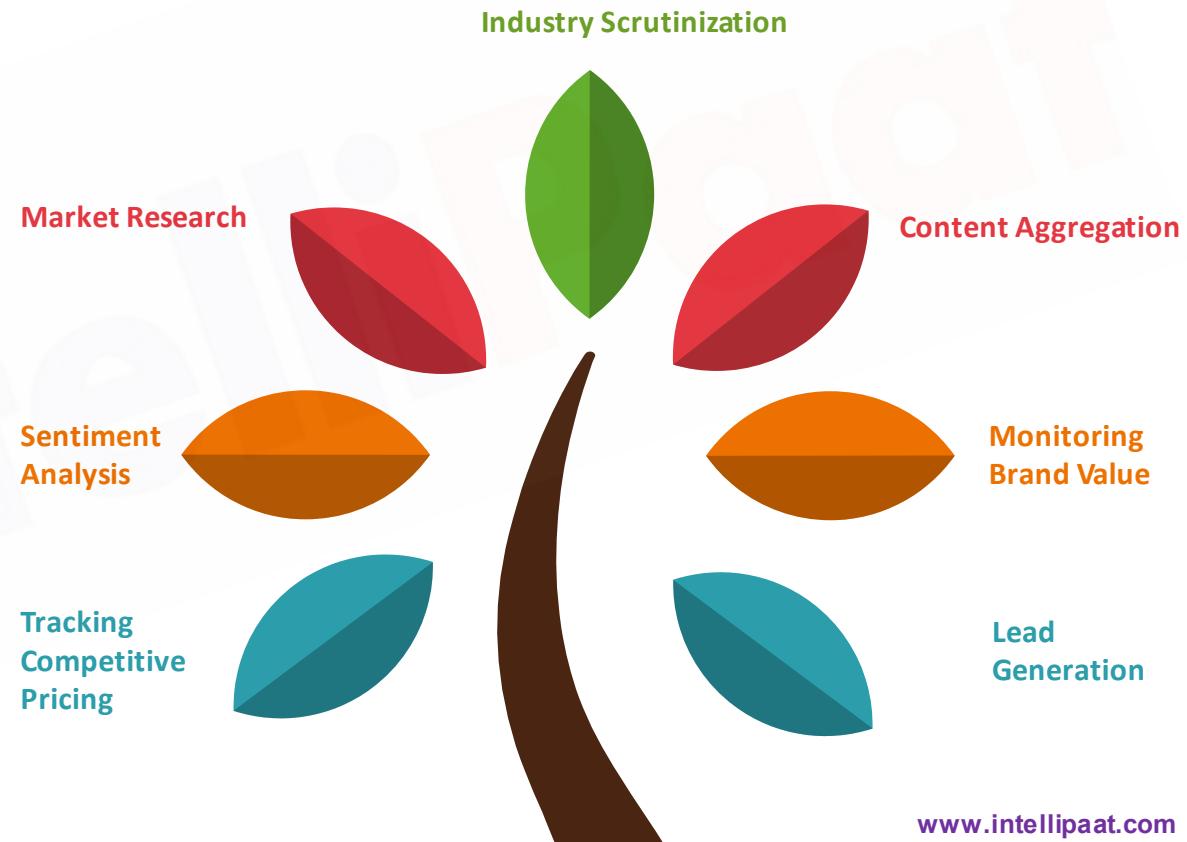
WEB SCRAPING



Need of Web- scraping



Business Application of Web-scraping



Is Web-scraping Legal?



Is Web-scraping / Web-Crawling Legal?

Maybe..



Is Web-scraping Legal?



Is Web-scraping / Web-Crawling Legal?

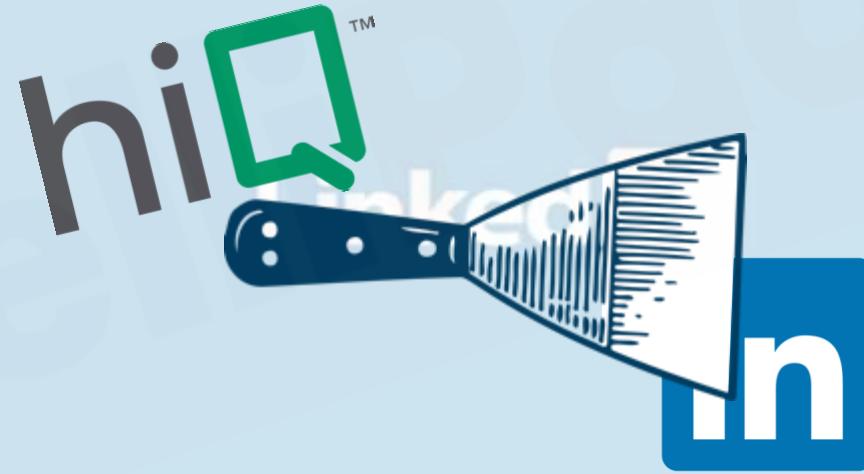
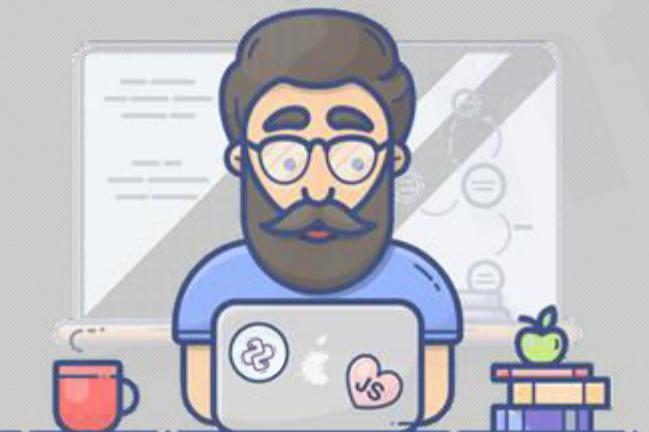
Know The Rules...

- Robots.txt
- Public Content
- Terms of Use
- Authentication Based Websites
- Crawl Delay



Is Web-scraping / Web-Crawling Legal?

Is Web-scraping Legal?



Python Libraries Required



Python Libraries required for Web-Scraping

Urllib2

Used for fetching URLs

`http://`

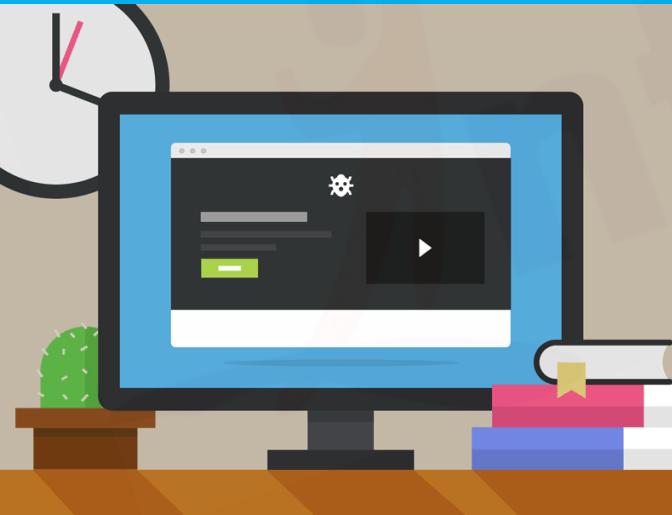


Beautiful Soup

Used for pulling out information from the webpage



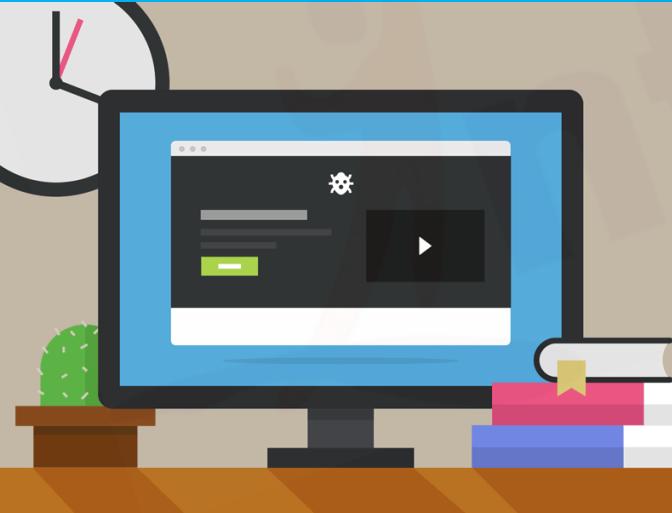
HTML Basics



Get familiar with basic HTML Tags



HTML Basics

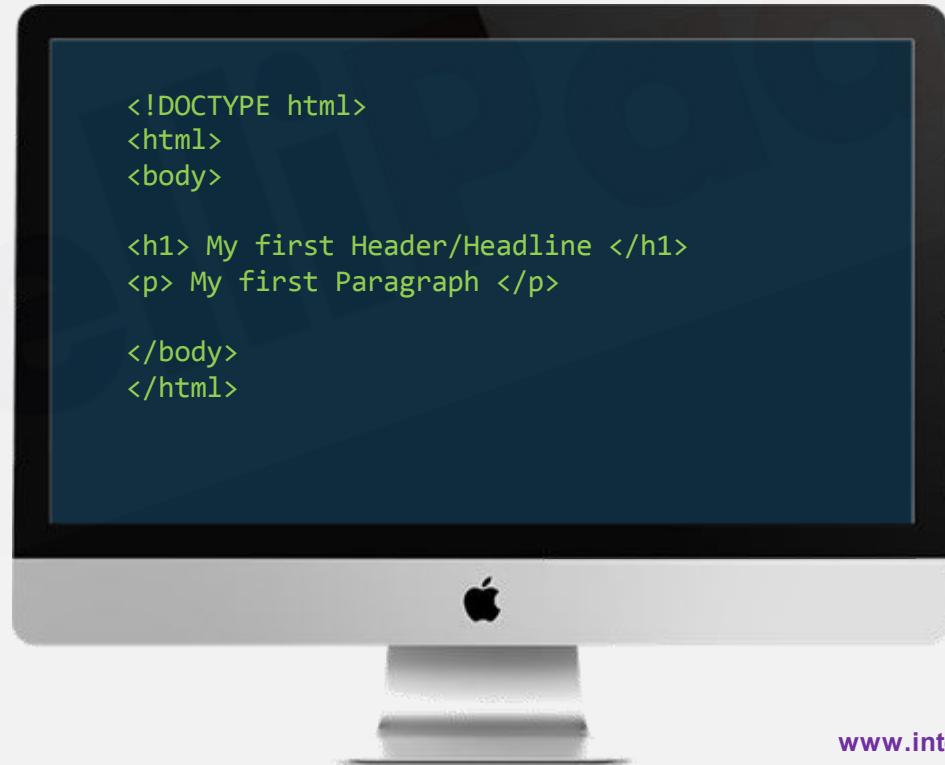


Get familiar with basic HTML Tags

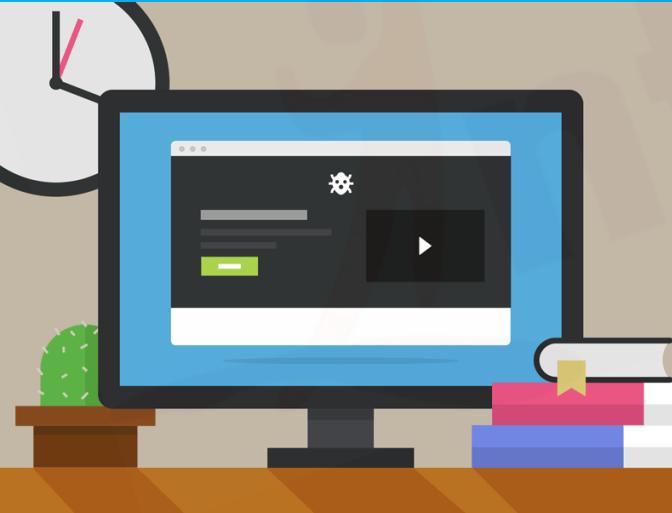
```
<!DOCTYPE html>
<html>
<body>

<h1> My first Header/Headline </h1>
<p> My first Paragraph </p>

</body>
</html>
```



HTML Basics

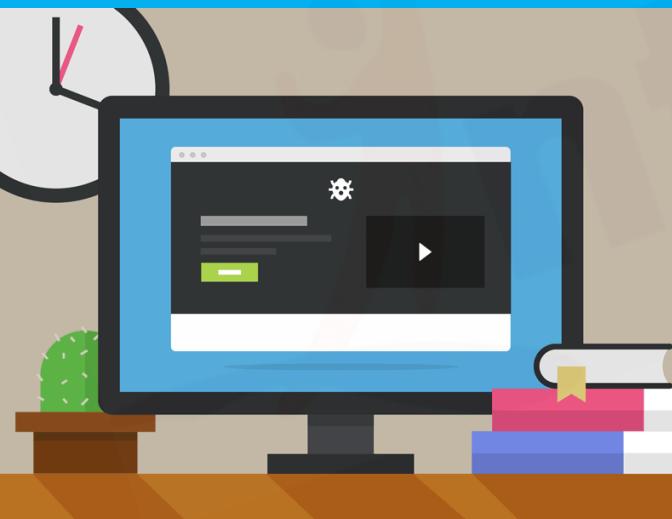


Get familiar with basic HTML Tags

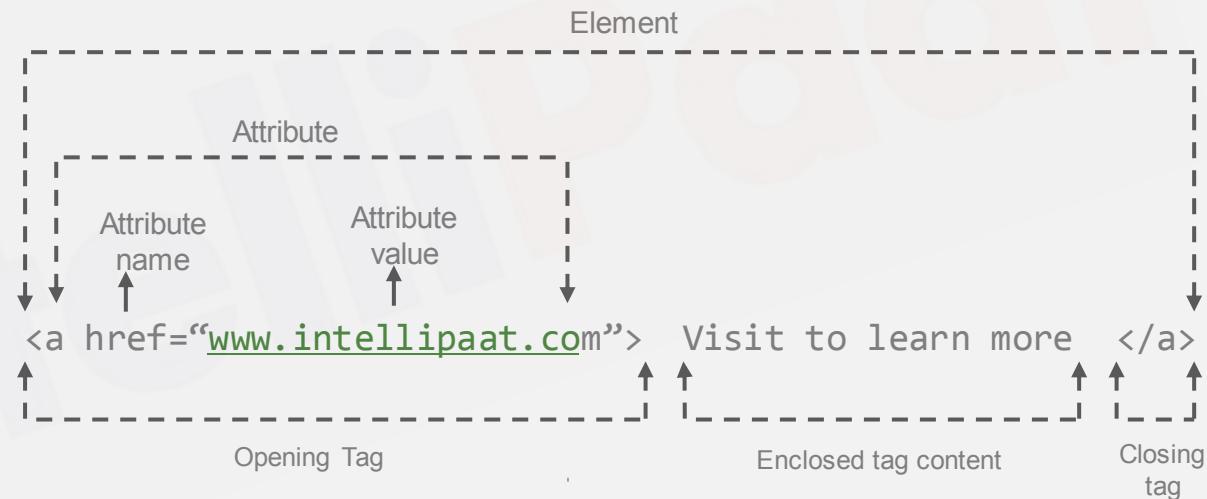
```
<table style="width:100%>
  <tr>
    <td>James</td>
    <td>Smith</td>
    <td>45</td>
  </tr>
  <tr>
    <td>John</td>
    <td>William</td>
    <td>35</td>
  </tr>
</table>
```



HTML Basics



Get familiar with basic HTML Tags



Scraping a Webpage



Thank
You

www.intellipaat.com



 CALL US NOW

India : +91-7847955955

US : 1-800-216-8930 (TOLL FREE)

sales@intellipaat.com