

Web Scrapping

(پروژه عملی با استفاده از Selenium)

سارا خرم



مباحث جلسات گذشته (جلسه 4 الی 8)

- نصب کتابخانه‌ها و موارد مورد نیاز
- BeautifulSoup Basics و توابع مهم
 - find()
 - find_all() & findAll()
 - select_one()
 - select()
 - get_text() & get attribute values

مباحث جلسات گذشته (جلسه 4 الی 8)

- استخراج دیتا از یک page (page اول)
- استخراج دیتا از چند page (pagination)

سرفصل مطالب



- نصب کتابخانه‌ها و موارد مورد نیاز

- Locators و Selenium Basics

- Class Name

- ID

- Name

- LinkText

- XPath

سرفصل مطالب



- جست و جو کردن
- استخراج دیتا از یک page (page اول)
- استخراج دیتا از چند page (pagination)

XPath : Locators

Path Expressions •

اهمیت بالا •

What is XPath?

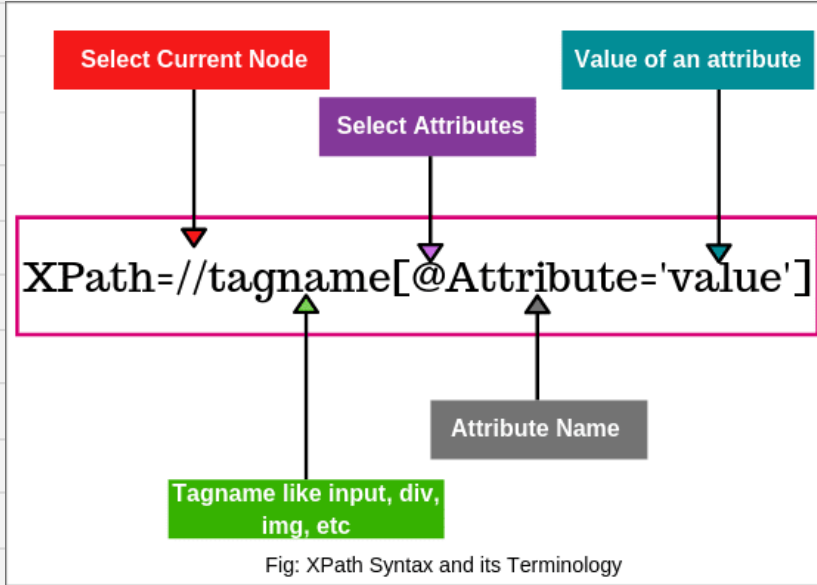


XPath : Locators

مثال :

```
<p class="price"> 200 </p>
```

```
//p[@class="price"]
```



XPath : Locators



- Absolute XPath : از المان root شروع می شود.

```
/html/body/div[2]/div/div[2]/ul/li[1]/div/a
```

- Relative XPath : از المان دلخواهی شروع می شود که root نیست (باید از // استفاده شود).

```
//a[@class='homepage-hero_link']
```


XPath : Locators



- تفاوت / و // :
- / : مستقیماً به child اشاره می‌شود.
- // : امان بعدی لزوماً به child اشاره نمی‌کند (یعنی می‌توان skip کرد).

```
/html/body/div[2]/div/div[2]/ul/li[1]/div/a
```

```
/html/body/div[2]/div/div[2]//li[1]/div/a
```

XPath : Locators



Xpath Expression	Description
<code>nodename</code>	This is the simplest one, it select all nodes with this <i>nodename</i>
<code>/</code>	Selects from the root node (useful for writing absolute path)
<code>//</code>	Selects nodes from the current node that matches
<code>.</code>	Selects the current node
<code>..</code>	Selects the current node's parent
<code>@</code>	Selects attribute
<code>*</code>	Matches any node
<code>@*</code>	Matches any attribute node

XPath : Locators



Xpath Expression	Description
<code>//li[last()]</code>	Selects the last <code>li</code> element
<code>//div[@class='product']</code>	Selects all <code>div</code> elements that have the class attribute with the product value.
<code>//li[3]</code>	Selects the third <code>li</code> element (the index starts at 1)

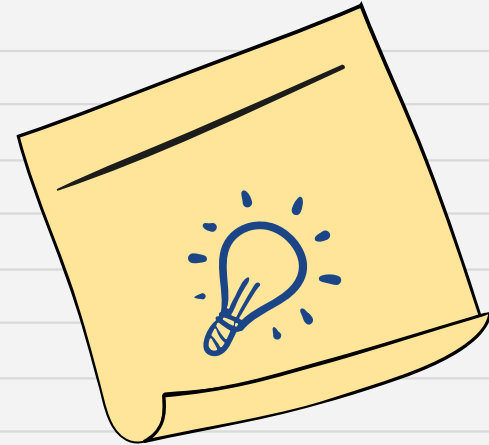
XPath : Locators



- به دست آوردن XPath :
- یادگیری Syntax
- از طریق مرورگر
- با استفاده از افزونه‌ها (extensions)

مصادر

- <https://www.selenium.dev/>
- <https://selenium-python.readthedocs.io/locating-elements.html>
- <https://chromedriver.chromium.org/downloads>
- <https://www.scrapingbee.com/blog/practical-xpath-for-web-scraping/#document-object-model>



Thanks!

آموزش های
رایگان بیشتر

www.data-hub.ir

www.t.me/data_hub_ir

www.linkedin.com/company/data-hub-ir