



XPath

Definizione

XPath = percorso XML.

È una sintassi o un linguaggio per trovare qualsiasi elemento nella pagina Web, utilizzando l'espressione del percorso XML.

XPath viene utilizzato per trovare la posizione di qualsiasi elemento su una pagina Web utilizzando la struttura DOM HTML.



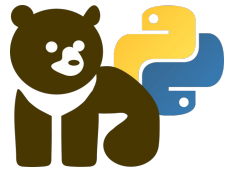
Sintassi

XPath contiene il percorso dell'elemento situato nella pagina web.

La sintassi standard per la creazione di XPath è la seguente:

XPath=//tagname[@attribute='value']

- // : seleziona tag corrente
- Tagname: Tagname del nodo
- @: seleziona attributo
- Attribute: Nome attributo del nodo
- Value: Valore dell'attributo



Tipi di XPath - Assoluto

E' il modo diretto per trovare l'elemento.

Ma lo svantaggio dell'XPath assoluto è che se vengono apportate modifiche al percorso dell'elemento, l'XPath non funziona più.

La caratteristica chiave di XPath è che inizia con la singola barra (/) , il che significa che puoi selezionare l'elemento dal nodo radice.

Esempio:

XPath assoluto: /html/body/div[1]/div/div[2]/header/div/div[2]/a/img



Tipi di XPath - Relativo

Il percorso inizia dal centro della struttura DOM HTML.

Inizia con la doppia barra (//), il che significa che può **cercare** l'elemento in qualsiasi punto della pagina web.

Vantaggio: puoi iniziare dal centro della struttura DOM HTML senza la necessità di scrivere un lungo XPath.

Esempio:

XPath relativo: `//*[@id="block-perfecto-main-menu"]/ul/li[6]/a`



Tools

Possiamo verificare un'espressione Xpath nello **strumento per sviluppatori di Chrome** o con **Firefox Firebug**.

Possiamo aprire lo strumento per sviluppatori in Chrome premendo F12, quindi la scheda **Console** deve essere selezionata.

Possiamo convalidare xpath con l'espressione `$x("<xpath value>")`.
Facendo clic su Invio dopo aver immesso l'espressione, verrà visualizzato un array di elementi corrispondenti. Passando il mouse sopra il risultato restituito, l'elemento effettivo viene evidenziato nella pagina. Se non ci sono elementi corrispondenti, verrà restituito un array vuoto.


Più indicazioni:

<https://yizeng.me/2014/03/23/evaluate-and-validate-xpath-css-selectors-in-chrome-developer-tools/>



Tools

Possiamo testare le nostre espressioni XPATH sulle pagine ad esempio tramite il sito:
<https://codebeautify.org/Xpath-Tester>

 **Code Beautify**

JSON Formatter | Hex Color Codes | HMAC Generator | My Ip | Search

XML Input

Option 1: Copy-paste your XML string here [Sample](#)

```
<!DOCTYPE html><html prefix="" lang="it"><link rel="preconnect" href="https://s1.immobiliare.it"/><link rel="preconnect" href="https://pwm.im-cdn.it"/><link rel="preconnect" href="https://crt.im-cdn.it"/><head><meta name="apple-mobile-web-app-status-bar-style" content="black-translucent"/><meta name="msapplication-TileColor" content="#2b6481"/><meta name="theme-color" content="#2b6481"/><meta name="apple-mobile-web-app-title" content="Immobiliare.it"/><meta name="application-name" content="Immobiliare.it"/><link rel="apple-touch-icon" href="https://s1.immobiliare.it/_next/static/media/apple-touch-icon.6eaaec1a.png"/><link rel="icon" type="image/png" sizes="32x32" href="https://s1.immobiliare.it/_next/static/media/favicon-32x32.c18ee3c9.png"/><link rel="icon" type="image/png" sizes="16x16" href="https://s1.immobiliare.it/_next/static/media/favicon-16x16.b6a5d8cf.png"/><link rel="mask-icon" href="https://s1.immobiliare.it/_next/static/media/safari-pinned-tab.ee8311f7ab52c53c.svg" color="#2b6481"/><link rel="shortcut icon" href="https://s1.immobiliare.it/_next/static/media/favicon.96dc93cb.ico"/><link rel="preload" href="https://s1.immobiliare.it/_next/static/media/inter-latin-400.493934f7829ce6ee.woff2" as="font" type="font/woff2" crossorigin=""></style>
```

Option 2: Or Load your XML file

Load Url

Browse

XPath Expression

TEST XPATH

RESET

Result : Generated XPath

```
1 href="https://www.immobiliare.it/annunci/98708906/"
2 href="https://www.immobiliare.it/annunci/93999130/"
3 href="https://www.immobiliare.it/annunci/102587674/"
4 href="https://www.immobiliare.it/annunci/93999128/"
5 href="https://www.immobiliare.it/annunci/77755068/"
6 href="https://www.immobiliare.it/annunci/102968710/"
7 href="https://www.immobiliare.it/annunci/98667808/"
8 href="https://www.immobiliare.it/annunci/101538113/"
9 href="https://www.immobiliare.it/annunci/98677260/"
10 href="https://www.immobiliare.it/annunci/103086888/"
11 href="https://www.immobiliare.it/annunci/104519631/"
12 href="https://www.immobiliare.it/annunci/104767525/"
13 href="https://www.immobiliare.it/annunci/103676024/"
14 href="https://www.immobiliare.it/annunci/56292390/"
15 href="https://www.immobiliare.it/annunci/104710567/"
16 href="https://www.immobiliare.it/annunci/104730391/"
17 href="https://www.immobiliare.it/annunci/104730389/"
18 href="https://www.immobiliare.it/annunci/86535732/"
19 href="https://www.immobiliare.it/annunci/93208666/"
20 href="https://www.immobiliare.it/annunci/94194408/"
21 href="https://www.immobiliare.it/annunci/80083987/"
22 href="https://www.immobiliare.it/annunci/83943086/"
23 href="https://www.immobiliare.it/annunci/94113722/"
24 href="https://www.immobiliare.it/annunci/90174199/"
25 href="https://www.immobiliare.it/annunci/93731518/"
26
```



Espressioni XPath

- Particolarmente utili nello scraping per localizzare e automatizzare
 - XPath base
 - **Contains()**
 - **OR & AND**
 - Funzione **Starts-With()**
 - **Text()**
 - Uso di indici



XPath base

L'espressione XPath seleziona nodi o elenchi di nodi sulla base di attributi come ID, nome, nome classe, ecc. dal documento XML come illustrato di seguito.

```
XPath=//*[@nome='nome utente']
```




Contains()

Viene utilizzato quando il valore di qualsiasi attributo cambia in modo dinamico, ad esempio le informazioni di registrazione.

La funzione **contains()** ha la capacità di trovare l'elemento con testo parziale come mostrato nell'esempio seguente.

In questo esempio, abbiamo cercato di identificare l'elemento utilizzando solo il valore di testo parziale dell'attributo. Nell'espressione XPath sottostante, viene utilizzato il valore parziale 'Name'.

Il valore completo di 'id' è 'FirstName' ma utilizza solo il valore parziale 'Name.'

```
XPath=//*[contains(@id,'Name')]
```



OR & AND

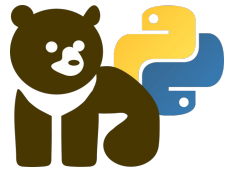
Nell'espressione OR vengono utilizzate due condizioni: qualsiasi condizione deve essere vera per trovare l'elemento.

Con AND entrambe devono essere vere.

Esempi:

```
XPath=//*[@id='Nome' or @nome='Cognome']
```

```
XPath=//*[@id='FirstName' and @name='FirstName']
```



StartsWith()


Trova l'elemento il cui valore dell'attributo cambia durante l'aggiornamento o qualsiasi operazione sulla pagina Web.

Si può ovviamente usare anche per trovare elemento statici.

Nell'espressione seguente sono presenti due elementi con una classe che inizia con "form". Quindi, XPath trova quegli elementi la cui "classe" inizia con "form".

```
XPath=//input[starts-with(@class,'form')]
```



 **Perfecto**

Sign in with Perfecto

Username or email

Password

☐ Remember me [Forgot Password?](#)

SIGN IN

Elements Console Sources Network >> 1

```
<div id="kc-form-wrapper" class="
  <form id="kc-form-login" class="form-horizontal" onsubmit="
    login.disabled = true; return true;" action="https://
    auth.perfectomobile.com/auth/realms/mobilecloud-perfectomobile-com/
    login-actions/authenticate?session_code=01ZVgBY25r3QgC-
    3U63xXhhnIg2lAEvb-zycAPVIYEE&execution=d9fd03f3-4d8a-452f-977c-
    6f03e48aefe8&client_id=reportium&tab_id=r9ig5Zk73Hc" method="post">
    <div class="form-group">
      <div class="col-xs-12 col-sm-12 col-md-12 col-lg-12">
        <label for="username" class="control-label">Username or
        email</label>
      </div>
      <div class="col-xs-12 col-sm-12 col-md-12 col-lg-12">
        <input tabindex="1" id="username" class="form-control"
        name="username" value type="text" autofocus autocomplete=
        "off">
      </div>
    </div>
    <div class="form-group">
      <div class="col-xs-12 col-sm-12 col-md-12 col-lg-12">_</div>
      <div class="col-xs-12 col-sm-12 col-md-12 col-lg-12">
        <input tabindex="2" id="password" class="form-control"
        name="password" type="password" autocomplete="off"> == $0
      </div>
    </div>
    <div class="form-group">_</div>
  </form>
</div>
</div>
</div>
<div id="support-message">_</div>
</div>
<div class="footer">_</div>
<iframe src="about:blank" frameborder="0" style="border: 1px solid #ccc; visibility:
hidden; z-index: 2147483647; position: fixed; width: 100%; left: 0px; bottom:
0px; height: 200px;" id="E1c9bua1IT">_</div>
... #kc-form #kc-form-wrapper #kc-form-login div div input#password,form-control
```

Two (2) Nodes Matched

1 of 2

Cancel

`//input[starts-with(@class,'form')]`




Text()

Troviamo l'elemento con la corrispondenza esatta del testo come mostrato di seguito.

In questo caso troviamo l'elemento con il testo "User name or email".

```
XPath=//*[text()='User name or email']
```



Sign in with Perfecto

Username or email

Password

☐ Remember me [Forgot Password?](#)

SIGN IN

One (1) Node Matched

<label for="username" class="control-label">Username or email</label> == \$0

//*[@text()='Username or email']

1 of 1

Cancel

```
<html xmlns="http://www.w3.org/1999/xhtml" class="login-pf" debug="true">
  <div id="FirebugChannel" style="display: none;"></div>
  <script src="chrome-extension://ehemiojjcpldeipjhjkepfdaohajpbdo/csp-test.js"
    id="FirebugLiteCSPTest" firebugignore="true" extension="Chrome" extension-id=
    "ehemiojjcpldeipjhjkepfdaohajpbdo" csp-mode="true"></script>
  <head></head>
  <script src="chrome-extension://ehemiojjcpldeipjhjkepfdaohajpbdo/firebug-
    lite.js" id="FirebugLite" firebugignore="true" extension="Chrome" extension-
    id="ehemiojjcpldeipjhjkepfdaohajpbdo"></script>
  <script src="chrome-extension://ehemiojjcpldeipjhjkepfdaohajpbdo/
    googleChrome.js"></script>
  <body class="data-gr-c-s-loaded=true">
    <div id="kc-container" class="container-fluid">
      <div id="kc-container-wrapper" class="">
        <div id="kc-header" class="login-pf-page-header">
          <div class="header-image-wrapper">
            <div class="kc-header-image"></div>
          </div>
          <div id="kc-header-wrapper" class=""> Sign in with Perfecto
          </div>
        </div>
        <div id="kc-content" class="col-sm-8 col-sm-offset-2 col-md-6 col-md-
          offset-3 col-lg-6 col-lg-offset-3">
          <div id="kc-content-wrapper" class="row">
            <div id="kc-form" class="col-sm-10 col-sm-offset-1 col-md-8 col-md-
              offset-2 col-lg-8 col-lg-offset-2">
                <div id="kc-form-wrapper" class="">
                  <form id="kc-form-login" class="form-horizontal" onsubmit=
                    "login.disabled = true; return true;" action="https://
                    auth.perfectomobile.com/auth/realms/mobilecloud-perfectomobile-com/
                    login-actions/authenticate?session_code=olZVgBY25r3QgC-
                    3e1d1hnlq2lAEvb-zycAPVIYEE&execution=d9fd03f3-4d8a-452f-977c-
                    6f03e4...&client_id=reportium&tab_id=r9ig5Zk73Hc" method="post">
                      <div class="form-group">
                        <div class="col-xs-12 col-sm-12 col-md-12 col-lg-12">
                          <label for="username" class="control-label">Username or
                          email</label> == $0
                        </div>
                      </div>
                    </form>
                  </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </body>
</html>
```



Uso di indici

Questo approccio viene utilizzato quando si desidera specificare un determinato nome di tag in termini di indice / ordine di posizione che si desidera individuare.

```
XPath=(//*[@attribute='valore'])[indice]
```

Nell'espressione seguente, un DOM ha più tag di input per ogni diverso valore di campo e desideri inserire del testo nel secondo campo, ad esempio LastName.

```
XPath=(//input[@type="text"])[2]
```


The screenshot shows the Selenium IDE interface with the 'Elements' tab selected. The DOM tree on the right displays the HTML structure of a form. A red box highlights the following HTML element:

```
<input id="LastName" name="LastName" maxlength="255" type="text" class="mktoField mktoTextField mktoHasWidth mktoRequired">
```

A speech bubble points to this element with the text: "One (1) Node matched".

At the bottom, the breadcrumb trail shows the path: `div > div > div > input#LastName.mktoField.mktoTextField.mktoHasWidth.mktoRequired`. The selected XPath expression is `(//input[@type="text"])[2]`, which is highlighted with a red box. The status bar indicates "1 of 1" matches.