



Data Journalism

Strategie di Web Scraping

InfoUma 2023 *Andrea Marchetti*

Sintesi

Esaminiamo diverse strategie di Web Scraping tramite Selenium tramite 3 esempi

- calcio femminile [femminile.football.it/ricerca.php](https://www.femminile.football.it/ricerca.php)
- carceri www.poliziapenitenziaria.it
- scuole cercalatuascuola.istruzione.it
-

Calcio femminile

<https://femminile.football.it/ricerca.php>

Femminile.Football.it

<https://femminile.football.it/ricerca.php>

Simulo un click sulla lettera Q che è un link

Giocatori

Cerca un Giocatore

Seleziona l'iniziale del cognome:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Ricerca libera: inserisci le lettere iniziali o il cognome intero

ents Console Sources Network Performance Memory Application Security Audits

```
▶ <td class="cellatopleft" height="18" nowrap>...</td>
▶ <td class="cellatopleft" height="18" nowrap>...</td>
▼ <td class="cellatopleft" height="18" nowrap>
  ▼ <div align="center"> == $0
    ▼ <b>
      <a href="?Ricerca=giocatori&iniziale=Q" title="18 risultanti">Q</a>
    </b>
  </div>
  (t)
```

Si apre da Chrome "Strumenti per sviluppatori" e si "ispeziona" gli elementi del DOM

```
}
div[Attrib
utes Style]
{
  text-
  alian
```

Questo è un link con il testo=Q

```
lettera = driver.find_element_by_link_text("Q").click()
```

Femminile.Football.it

Giocatori

Cerca un Giocatore

Seleziona l'iniziale del cognome:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Ricerca libera: inserisci le lettere iniziali o il cognome intero

Risultati ottenuti:

Clicca su un nome per accedere alla relativa scheda:

- QAFOKU MARTINA (4) [nato a il 28/11/2000]
- QUADRELLI SARA (57) [nato a Cattolica il 28/08/1997]
- QUAGGIOTTO ELENA (57) [nato a il 30/04/1992]
- QUAGLIA FRANCESCA (57) [nato a Padova il 13/04/1996]
- QUAGLIOTTO JESSICA (57) [nato a il 16/03/1993]
- QUARANTA MARTINA (57) [nato a il 19/05/1995]
- QUARESIMA MARA (57) [nato a il 15/04/2002]
- QUARTICELLI VALENTINA (57) [nato a Roma il 20/02/1990]
- QUARTO FEDERICA (57) [nato a il 13/04/1992]
- QUARTULLO FEDERICA (57) [nato a il 20/06/1991]
- QUATTROCCHI FRANCESCA (57) [nato a Palermo il 03/08/1988]
- QUAZZICO FRANCESCA (57) [nato a Taranto il 07/05/2001]
- QUERCIA MARZIA (57) [nato a il 16/09/1980]
- QUERZOLA LAURA (57) [nato a il 20/08/1991]
- QUIN JENNIFER (57) [nato a il]
- QUINTO MICHELINA (57) [nato a il 30/03/1988]
- QUISTINI VALENTINA (57) [nato a il 30/11/1985]
- QUITADAMO SONIA (57) [nato a il 10/10/1985]

Elements Console Sources >> 4 2

```
><table width="100%" border="0"
cellspacing="0" cellpadding="2"
bgcolor="#F3f5f8" align="center">...
</table>
...
▼<table width="100%" border="0"
cellspacing="0" cellpadding="0"
height="20" bgcolor="#FFFFFF"> == $0
▼<tbody>
▼<tr>
▼<td height="20" nowrap
bgcolor="#F3f5f8" align="left">
"
"
<a href="/
schedagiocatore.php?
id_giocatore=8316">QAFOKU
MARTINA (4) [nato a il
28/11/2000 ]</a>
</td>
</tr>
```

tbody

Styles Ev

XPATH della tabella contenente i dati
`/html/body/table/tbody/tr/td/table/tbody/tr/td[2]`
`/table/tbody/tr/td[2]/table[2]`

Save a list of dictionary to file

```
import json
with open('playersDB.json', 'w') as jsonFile:
    json.dump(playersDB, jsonFile)
```



```
import csv
keys = playersDB[0].keys() # Recupero i nomi dei campi dal primo dictionary
with open('playersDB.csv', 'w', newline='') as csvFile:
    dict_writer = csv.DictWriter(csvFile, keys) # imposto i nomi dei campi
    dict_writer.writeheader()                 # scriva la prima riga
    dict_writer.writerows(playersDB)          # scrivo i dati
```

Scheda player

https://femminile.football.it/schedagiocatore.php?id_giocatore=2082

Seleziona Scheda : 2018-2019

MARA ZABBEO

Scheda anagrafica	
Data di nascita	07/08/1981
luogo di nascita	Camposanpiero (Padova) - Italia
nazione	Italia
altezza	
peso forma	
ruolo	Difensore
sito ufficiale	
Nazionalita	 Italiana



Presenze				
	Presenze	Minuti (dal 0 in poi)	Gol	Squalifiche
Totali	97	9977	1	1
Serie A	0		0	0
Serie A2	43	4646	1	1
Serie B	54	5331	0	0
Regionali	0	0	0	0

Squadre					
Squadra	Campioni x squadra disputati	Serie A	Serie A2	Serie B	Regionali
Totali	8	0	0	5	0
Vicenza	1	0	0	1	0
Mestre	1	0	0	1	0
Schio	2	0	1	1	0
Padova	4	0	2	2	0

Cambiando id ottengo i dati di una calciatrice.

`getPlayerData(id)`

`/html/body/table/tbody/tr/td/table/tbody/tr/td[2]/table[3]/tbody/tr[1]/td[2]/table//tr`

`/html/body/table/tbody/tr/td/table/tbody/tr/td[2]/table[5]/tbody/tr[1]/td[2]/table//tr`

Espressioni regolari in Python

Python string methods

Python fornisce delle funzioni per manipolare le stringhe le più usate sono:

- `split('-')` # prende una stringa e restituisce una lista di stringhe
- `strip()` # elimina gli spazi bianchi marginali
- `replace()` # sostituisce una sottostringa con un'altra

Le altre possono essere trovate su [w3schools](https://www.w3schools.com/python/python_string_methods.asp)

Python Espressioni Regolari

```
import re # importare la libreria
```

match

restituisce un'istanza di MatchObject se zero o più caratteri all'inizio della stringa corrispondono al pattern. È comunque possibile specificare una posizione da cui iniziare la ricerca diversa dalla testa della stringa, e anche un limite superiore). Restituisce None se la stringa non corrisponde al pattern.

search

scandisce la stringa cercando una corrispondenza con il pattern e restituisce l'istanza MatchObject corrispondente. Restituisce None se in nessuna posizione la stringa corrisponde al pattern. Anche qui è possibile specificare un punto d'inizio e di termine per la ricerca come per match.

split

in breve divide la stringa secondo le occorrenze del pattern e restituisce le porzioni ottenute come lista.

escape

prende in ingresso una stringa e la restituisce con i caratteri alfanumerici protetti da barre oblique inverse.

sub

restituisce la stringa ottenuta sostituendo alle occorrenze non sovrapposte del pattern l'entità rimpiazzo, che può essere una stringa o una funzione, eventualmente per un numero di volte prefissato.

Sintassi Espressioni Regolari

Character	Description
[]	A set of characters
\	Signals a special sequence (can also be used to escape special characters)
.	Any character (except newline character)
^	Starts with
\$	Ends with
*	Zero or more occurrences
+	One or more occurrences
{ }	Exactly the specified number of occurrences
	Either or
()	Capture and group

Regular Expressions

[documentazione](#)

test on regular expression tester [online](#)

Carceri italiane

<https://www.poliziapenitenziaria.it/>

<https://www.poliziapenitenziaria.it/wp-content/uploads/2019/05/elenco-carceri-italiane.pdf>

Carceri Italiane

POLIZIA PENITENZIARIA

CARCERI LAZIO

Provveditorato Piemonte- Lazio – Abruzzo – Molise

`[[list style="check"]`

`[li]Scuola di Formazione e Aggiornamento del Corpo di Polizia e del Personale di`

Indirizzo: Via di Brava, 99 – 00163 Roma (RM)

Telefono: 06661741

Email: sfp.roma@giustizia.it

`[/li]`

`[li]Casa Reclusione Civitavecchia “Giuseppe Passerini”`

Indirizzo: Via Tarquinia, 20 – 00053 Civitavecchia (RM)

Telefono: 076650821

Email: cr.civitavecchia@giustizia.it

`[/li]`

`[li]Casa Circondariale Cassino`

Indirizzo: Via Sferracavalli, 3 – 03043 Cassino (FR)

Telefono: 077621019

Email: cc.cassino@giustizia.it

`[/li]`

Posizionarsi
sull'elemento di
interesse e con il
tasto destro del
mouse seleziona
**Ispeziona
Elemento**
per attivare
“Risorse per
sviluppatori”

```
<p> == $0
  "[list style="check"]"
  <br>
  " [li]Scuola di Formazione e Aggiornamento del Corpo di
  Polizia e del Personale dell'Amministrazione Penitenziaria
  Roma "Giovanni Falcone""
</p>
<div style="margin-left: 35px;">
  <strong>Indirizzo</strong>
  ":Via di Brava, 99 – 00163 Roma (RM)"
  <br>
  <strong>Telefono</strong>
  ": "
  <a href="tel:+3906661741">06661741</a>
  <br>
  <strong>Email</strong>
  ": "
  <a href="mailto:sfp.roma@giustizia.it">sfp.roma@giustizia.it
  </a>
</div>
```

poliziapenitenziaria.it/carceri-italiane/

CARceri ITALIANE

Falcone"

Indirizzo: Via di Brava, 99 – 00163 Roma (RM)
 Telefono: 06661741
 Email: sfp.roma@giustizia.it

✓ Casa Reclusione Civitavecchia "Giuseppe Passerini"

Indirizzo: Via Tarquinia, 20 – 00053 Civitavecchia (RM)
 Telefono: 076650821
 Email: cr.civitavecchia@giustizia.it

✓ Casa Circondariale Cassino

Indirizzo: Via Sferracavalli, 3 – 03043 Cassino (FR)
 Telefono: 077621019
 Email: cc.cassino@giustizia.it

✓ Istituto di istruzione Sulmona

Indirizzo: Via Fonte D'Amore, 1 – 67039 Sulmona (AQ)
 Telefono: 0864251280
 Email: sfp.sulmona@giustizia.it

✓ Casa Lavoro Vasto

Indirizzo: Via Torre Sinello, 23/A – 66054 Vasto (CH)
 Telefono: 0873310901
 Email: cl.vasto@giustizia.it

Console Elements Sources Network Performance Memory

```

    <blockquote class="modern-quote full">...</blockquote>
    <p>...</p>
    <h2 id="sovraffollamento">...</h2>
    <p>...</p>
    <h2 id="carezza-personale">...</h2>
    <p>...</p>
    <h2 id="suicidi">...</h2>
    <p>...</p>
    <h2 id="elenco-carceri">...</h2>
    <h3 class="regioni">...</h3>
    <h4>Provveditorato Piemonte- Lazio - Abruzzo - Molise</h4>
    ...
    <div class="sc-list sc-list-check fa-ul"> == $0
  
```

- ▼ <ul class="sc-list sc-list-check fa-ul"> == \$0
 - ▶ ... Add attribute
- ▶ ... Edit attribute
- ▶ ... Edit as HTML
- ▶ ... Duplicate element
- ▶ ... Delete element
- ▶ ... Copy
 - Cut element
 - Copy element
 - Paste element
 - Copy outerHTML
 - Copy selector
 - Copy JS path
 - Copy styles
 - Copy XPath
 - Copy full XPath
- ▶ ... Hide element
- ▶ ... Force state
- ▶ ... Break on
- ▶ ... Expand recursively
- ▶ ... Collapse children
- ▶ ... Capture node screenshot
- ▶ ... Scroll into view
- ▶ ... Focus
- ▶ ... Store as global variable

... page.status-publish.has-post-thumbnail div.post-content div.sc-list.sc-list-check.fa-ul

Console What's New Search

Highlights from the Chrome 89 update

Posizionarsi sull'elemento lato source di interesse. A sx il testo viene evidenziato

Premere il tasto dx per selezionare copy xpath

```
▼ <article id="post-59285" class="post-59285 page
  ▶ <header class="post-header">...</header>
  ▶ <div class="post-content">...</div> == $0
  </article>
```

```
//*[@id="post-59285"]/div/div[3]
```

A partire dalla radice (//) prendi un elemento generico (*) che ha l'attributo id (@id) uguale a post-59285

```
▼<article id="post-59285" class="post-59285 page type-page status-publish has-post-thumbnail">
  ▶<header class="post-header">...</header>
  ▼<div class="post-content">
    1 ▶<div id="ez-toc-container" class="ez-toc-v2_0_16 counter-hierarchy counter-decimal">...</div>
      ▶<p> </p>
    2 ▶<div class="sc-list sc-list-check fa-ul">...</div>
      ▶<p>...</p>
      ▶<blockquote class="modern-quote full">...</blockquote>
      ▶<p>...</p>
      ▶<h2 id="situazione">...</h2>
      ▶<p>...</p>
      ▶<blockquote class="modern-quote full">...</blockquote>
      ▶<p>...</p>
      ▶<h2 id="sovraffollamento">...</h2>
      ▶<p>...</p>
      ▶<h2 id="carenza-personale">...</h2>
      ▶<p>...</p>
      ▶<h2 id="suicidi">...</h2>
      ▶<p>...</p>
      ▶<h2 id="elenco-carceri">...</h2>
      ▶<h3 class="regioni">...</h3>
      <h4>Provveditorato Piemonte- Lazio - Abruzzo - Molise</h4>
    3 ▶<div class="sc-list sc-list-check fa-ul">...</div> == $0
      ▶<h3 class="regioni">...</h3>
      <h4>Provveditorato Piemonte- Valle D'Aosta - Liguria</h4>
```

/** [@id="post-59285"] /div/div[3]

Scuola in Chiaro

<https://cercalatuascuola.istruzione.it/>

Verificare se esistono dei dati open

tramite ricerca su google

- <https://dati.istruzione.it/opendata/opendata/>
- <http://www.datiopen.it/it/catalogo-opendata/miur-scuola-chiaro>
- <https://gist.github.com/steko/3895362>
- ...

Cosa c'è nel sito?



Il progetto Scuola in Chiaro

Il progetto "Scuola in chiaro" nasce nel 2011 per rispondere all'esigenza di mettere a disposizione della collettività tutte le informazioni disponibili relative alle scuole italiane di ogni ordine e grado, in una forma organica e strutturata.

Negli anni successivi l'applicazione è stata oggetto di continui interventi migliorativi. Sono state aggiunte, infatti, via via nuove funzionalità ed implementate quelle originarie; sono state introdotte nuove sezioni (come quella interamente rivolta all'offerta formativa della scuola o quella dedicata all'edilizia scolastica); è stato modificato il motore di ricerca che, con riferimento alla scuola primaria e secondaria di primo grado, ha consentito di impostare un filtro sulla tipologia del "tempo scuola" e, con riferimento alla scuola secondaria di secondo grado ha permesso di selezionare l'indirizzo di studio specifico. Nel corso del tempo, inoltre, sono state apportate modifiche "strutturali" per rendere l'applicazione adattabile alle esigenze dell'utente e alle caratteristiche uniche del dispositivo da lui utilizzato. Infine, è stato migliorato l'aspetto grafico dell'applicazione, nel rispetto degli standard di usabilità e dei criteri di accessibilità.

L'applicazione **Scuola in Chiaro** permette di:

- cercare una scuola o un centro di formazione professionale regionale sul territorio nazionale;
- conoscere tutte le informazioni disponibili sugli istituti scolastici di ogni ordine e grado e sui centri di formazione professionale ricercati;
- mettere a confronto l'offerta formativa delle scuole e dei centri di formazione selezionati;
- accedere direttamente ad alcuni servizi legati alla ricerca di scuole come, per esempio, le "Iscrizione on-line".

La base informativa che alimenta l'applicazione è costituita da dati già presenti nel sistema informativo del MI (continuamente aggiornati) e dalle informazioni inserite da ciascuna istituzione scolastica attraverso le funzioni presenti sul portale SIDI.

Si accede a **Scuola in Chiaro** direttamente dalla homepage del sito del MI <https://www.miur.gov.it/>.

Comprendere organizzazione sito



Guida alla navigazione

Nella parte alta della pagina, hai a disposizione un menu generale a scomparsa ☰ da cui puoi:

- tornare alla home page;
- aprire la mappa del sito;
- consultare la guida di navigazione;
- visualizzare l'elenco delle scuole preferite;
- accedere alla pagina dei contatti dell'Ufficio Relazioni con il Pubblico (URP).

Il breadcrumb, nella parte alta della pagina, indica dove ti trovi in ogni momento della navigazione.

Per trovare facilmente le informazioni desiderate puoi utilizzare tre differenti modalità di ricerca:

- "Ricerca per posizione geografica",
- "Ricerca rapida",
- "Ricerca avanzata".



Con la "**Ricerca rapida**", inserendo le parole chiave che intendi cercare (come, ad esempio, il nome della scuola o il suo codice, il comune di interesse o il tipo di percorso di studio che vuoi conoscere), ottieni l'elenco delle scuole corrispondente alla somma delle parole "descrittive" che hai inserito.



Con la "**Ricerca per posizione geografica**", individui tutte le scuole presenti nell'area geografica che hai indicato o compilando i campi presenti nella scheda o cliccando sulla mappa geografica; puoi anche specificare il raggio entro cui effettuare la ricerca. Seleziona il tipo d'istruzione. Poi procedi con la ricerca. L'indicatore sulla mappa evidenzia il punto esatto in cui è collocata la scuola o le scuole che rispondono ai criteri che hai indicato.



Per personalizzare i risultati di ricerca e trovare esattamente ciò che desideri utilizza la "**Ricerca avanzata**". Questa modalità ti permette di specificare la regione, la provincia, il comune, il tipo d'istruzione e altre caratteristiche che vuoi che siano presenti nella scuola che cerchi. Puoi anche inserire il nome della scuola o in alternativa il suo codice identificativo.

Comprendere il dominio

L'istruzione in Italia è regolata dal ministero dell'istruzione e dal ministero dell'università e della ricerca con modalità diverse a seconda della forma giuridica

- **scuole pubbliche** (gestite dallo Stato)
- **scuole paritarie** (gestite da enti privati ma godono di finanziamenti pubblici)
- **scuole private** (gestite da enti privati e non ricevono finanziamenti pubblici)

Nel sito scuola in chiaro non sono presenti le scuole private



Statale



Paritaria



Centri Formazione Professionale

Comprendere il dominio

In Italia, l'organizzazione delle scuole è suddivisa

Scuola dell'infanzia (per bambini dai 3 ai 6 anni, non obbligatoria). ASILO

Scuola primaria (per bambini dai 6 agli 11 anni, obbligatoria). ELEMENTARI

Scuola secondaria di primo grado (per ragazzi dagli 11 ai 14 anni, obbligatoria). MEDIE

Scuola secondaria di secondo grado (obbligatoria per ragazzi dai 14 ai 16 anni, non obbligatoria per ragazzi dai 16 ai 19 anni). LICEO, ...

Comprendere il dominio

circolo didattico (è la totalità delle scuole di infanzia e primaria appartenenti a uno stesso territorio)

istituto comprensivo (costituiti da più scuole nella stessa sede con un solo preside ed un solo consiglio di istituto.)

plesso scolastico (è la scuola specifica, la sede fisica e concreta nel quale gli studenti si recano)

Organizzazione del sito

Consente di ricercare le “Scuole” in tre modalità

- per parola chiave o rapida
- per posizione
- per livello amministrativo (regione, provincia, comune)

Ricerca per posizione

The image shows a web interface for 'Scuola in Chiaro' with a search section titled 'Ricerca per posizione'. At the top, there are three icons: a magnifying glass, a location pin, and a list icon. Below these is a search bar with the placeholder text 'Inserisci un indirizzo o attiva la localizzazione e'. To the right of the search bar are two dropdown menus: 'Raggio *' with a 'Seleziona' dropdown arrow and 'Tipo di istruzione *' with a 'Seleziona' dropdown arrow. Below the dropdowns are two radio buttons: 'Statale' (unselected) and 'Paritaria' (unselected). At the bottom of the search section are two buttons: 'Visualizza in mappa' and 'Visualizza elenco'. To the right of the search section is a map of Italy and surrounding regions, including Monaco, Croazia, Bosnia ed Erzegovina, Montenegro, and Albania. The map shows major cities like Milano, Genova, Bologna, Firenze, Roma, Napoli, and Palermo. At the top of the map, there are three location pin icons with labels: 'Statale' (blue), 'Paritaria' (green), and 'Centri Formazione Professionale' (red). The map also has 'Mappa' and 'Satellite' tabs, a 'Seft' logo, and a person icon in the bottom right corner. At the bottom of the map, there is a Google logo and a small text line: 'Scorciatoie da tastiera | Dati mappa ©2022 GeoBasis-DE/BKG (©2009), Google, Inqt, Geogr. Nacional, Mapa GISrael | Termini e condizioni d'uso'.

Ricerca avanzata

Selezionando tutte le **regioni** posso coprire con certezza tutto il territorio nazionale

Per le Regioni più popolate potrei passare alle province



The screenshot shows the 'Scuola in Chiaro' advanced search interface. At the top, there is a logo with a location pin and the text 'Scuola in Chiaro'. Below the logo are three circular icons: a magnifying glass, a location pin, and a list icon. The main search form is titled 'Ricerca avanzata' and contains the following fields:

- Regione:** A dropdown menu with the text 'Seleziona' and a downward arrow.
- Provincia:** A dropdown menu with a downward arrow.
- Comune:** A dropdown menu with a downward arrow.
- Tipo di istruzione:** A dropdown menu with the text 'Seleziona' and a downward arrow.
- Statale:** A checkbox.
- Paritaria:** A checkbox.
- Denominazione:** A text input field.
- Codice meccanografico:** A text input field.

At the bottom of the form is a green button labeled 'Cerca'.

URL generate da una ric. avanzata

<https://cercalatuascuola.istruzione.it/cercalatuascuola/ricerca/resultati?>

codiceRegione=VA_VALLE+D'AOSTA&

codiceProvincia=&

codiceComune=&

codiceOrdine=&

checkStatale=S&

checkNonStatale=S&

radioBiennioTriennio=Biennio&

denominazione=&

codMecc=&

tipoRicerca=AVANZATA&

gidf=1

} Livelli amministrativi, almeno uno deve essere presente

} Obbligatorio

Query String

Ispezionare elemento select

The image displays a web browser interface on the left and a developer tools window on the right. The browser shows a search form titled "Ricerca avanzata" with several dropdown menus and checkboxes. The developer tools window shows the HTML structure of the page, with the `<select>` element for the region selection highlighted. The `<select>` element has the following attributes: `id="codiceRegione"`, `name="codiceRegione"`, `class="required"`, and `title="Selezionare una regione"`. The `onchange` attribute is set to `javascript:caricaProvincia(this.value);`. The `<option>` elements are listed with their values and labels, including "AB_ABRUZZO", "BA_BASILICATA", "CL_CALABRIA", "CA_CAMPANIA", "EM_EMILIA-ROMAGNA", "FR_FRIULI VENEZIA GIULIA", "LA_LAZIO", "LI_LIGURIA", "LO_LOMBARDIA", "MA_MARCHE", "MO_MOLISE", "PI_Piemonte", "PU_PUGLIA", "SA_SARDEGNA", "SI_SICILIA", "TO_TOSCANA", "TR_TRENTINO-ALTO ADIGE", "UM_UMBRIA", "VA_VALLE D'AOSTA", and "VE_VENETO".

```
HTML Structure:  
...  
<div class="sc-field sc-field-1-2">  
  <label>Regione </label>  
  <div class="sc-select"> == $0  
    <select id="codiceRegione" name="codiceRegione" class="required" title="Selezionare una regione"  
      onchange="javascript:caricaProvincia(this.value);">  
      <option value selected="selected">Seleziona</option>  
      <option value="AB_ABRUZZO">ABRUZZO</option>  
      <option value="BA_BASILICATA">BASILICATA</option>  
      <option value="CL_CALABRIA">CALABRIA</option>  
      <option value="CA_CAMPANIA">CAMPANIA</option>  
      <option value="EM_EMILIA-ROMAGNA">EMILIA-ROMAGNA</option>  
      <option value="FR_FRIULI VENEZIA GIULIA">FRIULI VENEZIA GIULIA</option>  
      <option value="LA_LAZIO">LAZIO</option>  
      <option value="LI_LIGURIA">LIGURIA</option>  
      <option value="LO_LOMBARDIA">LOMBARDIA</option>  
      <option value="MA_MARCHE">MARCHE</option>  
      <option value="MO_MOLISE">MOLISE</option>  
      <option value="PI_PIEMONTE">PIEMONTE</option>  
      <option value="PU_PUGLIA">PUGLIA</option>  
      <option value="SA_SARDEGNA">SARDEGNA</option>  
      <option value="SI_SICILIA">SICILIA</option>  
      <option value="TO_TOSCANA">TOSCANA</option>  
      <option value="TR_TRENTINO-ALTO ADIGE">TRENTINO-ALTO ADIGE</option>  
      <option value="UM_UMBRIA">UMBRIA</option>  
      <option value="VA_VALLE D'AOSTA">VALLE D'AOSTA</option>  
      <option value="VE_VENETO">VENETO</option>  
    </select>  
  </div>  
  <div class="sc-field sc-field-1-2" id="provincia_container">...</div>  
</div>  
...  
<div class="sc-fields">...</div>  
<div class="sc-fields" id="tipoScuola_containerAvanzata">...</div>  
<div class="sc-fields" id="tipologiaStataleNuovoOrdinamento_containerAvanzata" style="display: none;">...</div>  
<div class="sc-fields" id="nuovi_indirizzi_professionale_containerAvanzata" style="display: none;">...</div>
```

Form Fields:
- Regione: Seleziona
- Provincia: [Dropdown]
- Comune: [Dropdown]
- Tipo di istruzione: Seleziona
- Statale:
- Paritaria:
- Denominazione: [Text Field]
- Codice meccanografico: [Text Field]
- Cerca: [Button]

Uso di f:string per generare url

```
# Array dei codici delle regioni usato dal sito
```

```
codiceRegioni = ["AB_ABRUZZO", "BA_BASILICATA", "CL_CALABRIA", "CA_CAMPANIA",  
                "EM_EMILIA-ROMAGNA", "FR_FRIULI VENEZIA GIULIA", "LA_LAZIO",  
                "LI_LIGURIA", "LO_LOMBARDIA", "MA_MARCHE", "MO_MOLISE",  
                "PI_Piemonte", "PU_PUGLIA", "SA_SARDEGNA", "SI_SICILIA",  
                "TO_TOSCANA", "TR_TRENTINO-ALTO ADIGE", "UM_UMBRIA", "VA_VALLE D'AOSTA", "VE_VENETO"]
```

```
# Scansione delle regioni
```

```
fixedPart = "https://cercalatuascuola.istruzione.it/cercalatuascuola/ricerca/risultati?tipoRicerca=AVANZATA&"  
totale = 0  
for codiceRegione in codiceRegioni:  
    url = fixedPart + f"codiceRegione={codiceRegione}"  
    driver.get(url)
```

&codiceComune=G628_PIETRASANTA

https://cercalatuascuola.istruzione.it/cercalatuascuola/ricerca/risultati?tipoRicerca=AVANZATA&codiceComune=G628_PIETRASANTA



Scuola in Chiaro



```
/html/body/section/div/div/h1
```

23 risultati

Nuova ricerca

Scuole selezionate 0

Istituto Principale	Plesso / Scuola	Indirizzo	Telefono	E-mail	Mappa	Seleziona
PIETRASANTA 2 LUIC850009	"ALESSIO RICCI" LUUE85001B	VIA PONTENUOVO, 55045 PIETRASANTA (LU)	0584791088	LUIC850009@istruzione.it		<input type="checkbox"/>
PIETRASANTA 1 LUIC83400B	"ANNA PILI" LUAA83404B	VIA DELLA CHIESA 13, 55045 PIETRASANTA (LU)	0584771059	LUIC83400B@istruzione.it		<input type="checkbox"/>
PIETRASANTA 1 LUIC83400B	"BARSANTI" PIETRASANTA LUMM83401C	P.ZZA MATTEOTTI, 37, 55045 PIETRASANTA (LU)	0584793619	LUIC83400B@istruzione.it		<input type="checkbox"/>

```
rows =  
driver.find_elements(By.XPATH  
, "//*[@id='tabellaRisultati']  
/tbody/tr")
```

PIETRASANTA 1 LUIC83400B	"P.G. BARSOTTINI" LUEE83402E	VIA DELLA SVOLTA, 55045 PIETRASANTA (LU)	0584735379	LUIC83400B@istruzione.it		<input type="checkbox"/>
PIETRASANTA 2 LUIC850009	"PETER PAN" LUAA850016	VIA COMUNALE, 183, 55040 PIETRASANTA (LU)	0584772147	LUIC850009@istruzione.it		<input type="checkbox"/>
PIETRASANTA 1 LUIC83400B	"SIRIO GIANNINI" LUAA83403A	VIA CROCE VERDE, 55045 PIETRASANTA (LU)	0584757123	LUIC83400B@istruzione.it		<input type="checkbox"/>

Statale Paritaria Centri Formazione Professionale

Pagina 1/3

Scuole selezionate 0

```
driver.find_element(By.CLASS_NAME, "sc-pager-next").click()
```

Pagina 1/3

```

pager= driver.find_element(By.CLASS_NAME,"sc-pager")
pager= pager.find_element(By.TAG_NAME,"span")

```