



Data Journalism

Strategie di Web Scraping

InfoUma 2021-22 *Andrea Marchetti*

Scuola in Chiaro

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

Verificare se esistono dei dati open già pronti

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

La ricerca restituisce l'elenco delle scuole che rispondono ai criteri impostati (con la denominazione, l'indirizzo, il codice scuola e altri dati principali) e, sulla mappa, la posizione geografica delle scuole selezionate.

I marker rappresentati in mappa con tre differenti colori indicano se le scuole che stai visualizzando sono **Istituti Statali (blu)**, **Paritari (verde)** o se si tratta di **Centri di Formazione Professionale (rosso)**.

Le **scuole statali** vengono supportate dallo Stato ed assicurano il diritto allo studio promuovendo un'istruzione laica e gratuita.

Le **scuole paritarie** invece non vengono gestite dallo Stato e hanno la totale libertà da un punto di vista dell'orientamento culturale e didattico.

Le **scuole paritarie** svolgono un servizio pubblico e sono inserite nel sistema nazionale di istruzione. Per gli alunni, la regolare frequenza della scuola paritaria costituisce assolvimento dell'obbligo di istruzione. ... l'abilitazione a rilasciare titoli di studio aventi lo stesso valore legale delle scuole statali

Comprendere il dominio

circolo didattico

istituto comprensivo

plesso scolastico

...

Ricerca per posizione

The image shows a web interface for 'Scuola in Chiaro' with a search focus on location. At the top, the logo 'Scuola in Chiaro' is centered, featuring a location pin icon. Below it are three circular icons: a magnifying glass, a location pin, and a list icon. The main search area is titled 'Ricerca per posizione' and includes a text input field with the placeholder 'Inserisci un indirizzo o attiva la localizzazione e'. Below the input field are two dropdown menus: 'Raggio *' with a 'Seleziona' dropdown arrow and 'Tipo di istruzione *' with a 'Seleziona' dropdown arrow. There are two radio buttons: 'Statale' (unselected) and 'Paritaria' (unselected). Below these are two buttons: 'Visualizza in mappa' and 'Visualizza elenco'. To the right is a map of Italy and surrounding regions, showing major cities like Milano, Genova, Roma, and Napoli. The map includes a legend with three categories: 'Statale' (blue location pin), 'Paritaria' (green location pin), and 'Centri Formazione Professionale' (red location pin). The map also shows labels for 'Croazia', 'Bosnia ed Erzegovina', 'Montenegro', and 'Albania'. At the bottom of the map, there is a Google logo and copyright information: 'Scorciatoie da tastiera | Dati mappa ©2022 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional, Mapa GISrael | Termini e condizioni d'uso'.

Ricerca avanzata

Selezionando **la regione** posso coprire con certezza tutto il territorio nazionale



Scuola in Chiaro

Ricerca avanzata

Regione	Provincia
<input type="text" value="Seleziona"/>	<input type="text"/>
Comune	Tipo di istruzione
<input type="text"/>	<input type="text" value="Seleziona"/>
Statale	Paritaria
<input type="checkbox"/>	<input type="checkbox"/>
Denominazione	
<input type="text"/>	
Codice meccanografico	
<input type="text"/>	

Cerca

Controllare le URL generate

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

codiceRegione=VA_VALLE+D'AOSTA&

codiceProvincia=&

codiceOrdine=&

radioBiennioTriennio=Biennio&

denominazione=&

codMecc=&

tipoRicerca=AVANZATA&

gidf=1

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 input fields: "Regione" (a dropdown menu), "Provincia" (a dropdown menu), "Comune" (a dropdown menu), "Tipo di istruzione" (a dropdown menu), "Statale" (a checkbox), "Paritaria" (a checkbox), "Denominazione" (a text input), and "Codice meccanografico" (a text input). A green "Cerca" button is at the bottom.

The developer tools window shows the HTML structure of the form. The selected element is a `<select>` element with the following attributes and content:

```
<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">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>
```

The developer tools also show the CSS styles for the selected element, including a width of 100% and a height of 40px.

Uso di f:string per generare url

```
for codiceRegione in ["AB_ABRUZZO", "BA_BASILICATA", ...]:  
    url = f"https://../risultati?codiceRegione={codiceRegione}"  
    driver.get(url)
```