

MARTA ARAMU



# WEB SCRAPER

Automatizzare compiti ripetitivi  
per risparmiare tempo

MARTA ARAMU

# COSA È IL WEB SCRAPING

Il web scraping serve a estrarre dati dalle pagine web per poi raccoglierli in database o tabelle locali per analizzarli. Si tratta di un sistema in grado di estrapolare una grande varietà di informazioni: dati di contatto, indirizzi di posta elettronica, numeri di telefono, così come singoli termini di ricerca o URL.



# IL PROGETTO

In questo progetto andremo a creare un web scraper incaricato di trovare tutte le offerte sul sito Subito.it relative ad una specifica macchina fotografica, per poi salvarle in un foglio di testo con i rispettivi URL.





# STRUMENTI UTILIZZATI



Beautiful Soup è una libreria Python per l'analisi di documenti HTML e XML, inclusi quelli con markup non valido. Crea un albero di analisi per i documenti che può essere utilizzato per estrarre dati da HTML, utile per il web scraping.



Requests

Python Requests è una libreria per effettuare richieste HTTP. Fornisce un'interfaccia facile da usare che rende molto semplice lavorare con HTTP, in quanto semplifica il processo di invio e ricezione di dati dai siti Web fornendo un'interfaccia uniforme per entrambi i metodi GET e POST.



# RISULTATI



```
n.py      ≡ risultati_salvati.txt ×  
raper > ≡ risultati_salvati.txt  
https://www.subito.it/fotografia/panasonic-dc-tz90-4k-messina-543016148.htm  
https://www.subito.it/fotografia/panasonic-lumix-dc-tz90-perugia-528797051.htm  
https://www.subito.it/fotografia/fotocamera-panasonic-lumix-dc-tz90-leica-lens-venezia-540583185.htm  
https://www.subito.it/fotografia/panasonic-dc-tz90-vicenza-537849590.htm  
https://www.subito.it/fotografia/fotocamera-digitale-compatta-lumix-dc-tz90-palermo-541531006.htm  
https://www.subito.it/fotografia/panasonic-lumix-tz10-varese-429915684.htm  
https://www.subito.it/fotografia/panasonic-lumix-dc-tz90-verbano-cusio-ossola-504730320.htm
```

Avviando il programma, ecco cosa dobbiamo aspettarci: la creazione di un file .txt contenente tutti gli url che rimandano agli annunci pubblicati dagli utenti del sito.

```
[Running] python -u "c:\Users\marta\Desktop\webScraper\main.py"  
Ci sono nuovi risultati. Apertura in corso...
```



Inoltre verremo avvisati con un messaggio in console nel caso vengano rilevati nuovi annunci rispetto all'ultima volta che il programma è stato avviato.

MARTA ARAMU



GRAZIE PER  
L'ATTENZIONE!

