

Data Collection

Web Scraping Esercitazione

Andrea Marchetti

Metodi di PlayWright

Localizzazione elementi

```
# Per elementi con ruoli ARIA.  
page.get_by_role("button", name="Invia")  
# Per elementi associati a etichette.  
page.get_by_label("provincia")  
# Per elementi contenenti un testo specifico.  
page.get_by_text("Testo da cercare")  
# locator(css): standard, semplice e leggibile.  
page.locator("css=selettore CSS")  
# locator(xpath): potente, ma meno leggibile.  
page.locator("xpath=selettore XPath")
```

Estrazione di contenuto

```
# Estrazione del contenuto visibile di un elemento.  
titolo = page.locator("h1").inner_text()  
  
# Estrazione di tutto il contenuto di un elemento.  
paragrafo = page.locator("div.content").text_content()  
  
# Estrazione di più elementi.  
paragrafi = page.locator("p").all_inner_texts()  
  
# Estrazione di attributi.  
url = page.locator("a").get_attribute("href")  
immagine = page.locator("img").get_attribute("src")
```

Interazione con Form elements

```
# Inserisce "Testo di esempio".
page.locator("input#text-input").fill("Testo di esempio")

# Seleziona l'opzione con value "opzione2".
page.locator("select#select").select_option("opzione2")

# Seleziona il radio button con id "radio2".
page.locator("input#radio2").check()

# Clicca sul pulsante di submit per inviare il form.
page.locator("button[type='submit']").click()
```

Ambiente di lavoro

—>esercizi (directory contenente i testi degli esercizi)

—>HTML (directory dove inserire i file .html)

scraperTemplate.py

test_playwright.py

web_server.py (eseguire su un terminale a parte: è il server web)