

Zadanie domowe do zajęć z siódmego dnia kursu 'Python for Web'

Zakres:

Zadanie bazują na materiale omówionym w trakcie zajęć 5-7.

Rozwiązania:

Rozwiązanie umieść w swoim prywatnym repozytorium.

Przykładowa nazwa pliku z rozwiązaniem:

- zadanie.py

Zadanie

W katalogu 'Day07/small_big_data' umieściłem dwa pliki:

- save_allegro_offer_details.py
- offers.zip

Pierwszy plik zawiera szkielet aplikacji do odczytu i przetwarzania plików ofert – ten plik należy rozbudować w trakcie rozwiązywania zadania.

Drugi plik zawiera oferty aukcji; zawartość tego archiwum (pliki html) umieść w katalogu 'Day07/small_big_data/offers'.

Zadanie polega na uzupełnieniu szkieletu o:

- obsługę sytuacji, gdy brakuje etykiety. Technika wyciągania danych – dowolna, w której najlepiej się czujesz (regex, find, BeautifulSoup, XPATH lub CSS Selectors),
- ustaleniu zakresu danych, które Ciebie interesują (zakres może być węższy niż zaproponowany w szkielecie),
- wyciągnięciu danych z plików ofert i umieszczeniu ich w pamięci komputera (słownik lub obiekt typu Oferta; jeśli zdecydujesz się na klasę Oferta trzeba ją zdefiniować)
- utworzeniu pliku, który będzie przechowywał wyciągnięte dane (CSV lub JSON – który preferujesz)
- zapisaniu kolekcji ofert do pliku

Po zakończeniu przetwarzania porównaj wielkość pliku wynikowego (CSV lub JSON) z sumaryczną wielkością plików html.