

Read.me

Beautiful soup library is used to web scrape the amazon webpages.

The provided excel file is read by using the function `read_excel` from pandas library. Two different lists are maintained for both ansi and country codes and the values of these lists are imported from the excel sheet. We loop through these lists and form an amazon link for the combinations of country codes and ansi values.

For every pair of country code and ansi values we send a GET request to the corresponding amazon page and check if the response code is 404 or not. If 404 then we skip and move to next pair of country code and ansi value. Otherwise, we scrape the webpage and take product title from element having id "productTitle", price from element having attributes class = "a-offscreen" and we retrieve image link from the block having id = "img-canvas". Different functions are created for performing each of these tasks. A dictionary is formed with these values and appended into a list. These values are also stored on an online hosted database, elephantSQL.

After performing the above tasks for every pair of ansi and country codes, we convert the list containing every dictionary to a json object using `json.dumps` and then write this json object to a json file.

Github link: https://github.com/Anas-Ahmad-Siddiqui/amazon_webscraping.git

Snapshot of the values stored in database

2022-04-23 13:43:23 +0000	Musik für Kinder: Im Fünftenraum. Vol. 1. Singstimme Blockflöte und Schlagzeug. Sing- und Spielpartitur. (Orff- Schulwerk)	https://images-eu.ssl-images-amazon.com/images/I/419-3yqvVML_SX198_BO1,204,203,200_QL40_ML2_.jpg	€22.00
2022-04-23 13:43:26 +0000	Schulwerk vol. 1 (ted.) (orff/keetman)	https://images-eu.ssl-images-amazon.com/images/I/419-3yqvVML_SX218_BO1,204,203,200_QL40_ML2_.jpg	0% (0%)
2022-04-23 13:43:32 +0000	5 antike Oden: nach Gedichten von Sappho. op. 57. mittlere Frauenstimme Viola und Klavier. mittel. Stimmensatz. (Edition Schott)	https://images-eu.ssl-images-amazon.com/images/I/41uGG66YO-L_SX198_BO1,204,203,200_QL40_ML2_.jpg	32,00 €
2022-04-23 13:43:34 +0000	Fünf antike oden op. 57 -ensemble de partitions	https://images-eu.ssl-images-amazon.com/images/I/41uGG66YO-L_SX218_BO1,204,203,200_QL40_ML2_.jpg	41,45 €
2022-04-23 13:44:08 +0000	Etuden 1 & 2 orgue	https://images-na.ssl-images-amazon.com/images/I/31+SH2b7VAL_SX258_BO1,204,203,200_.jpg	31,19 €
2022-04-23 13:44:11 +0000	Etuden 1 & 2 orgue: Orgel.	https://images-na.ssl-images-amazon.com/images/I/31+SH2b7VAL_SX258_BO1,204,203,200_.jpg	31,19 €
2022-04-23 13:45:00 +0000	Arien(2) (grand macabre) 2e akte	https://images-eu.ssl-images-amazon.com/images/I/41usGis9EJL_SX198_BO1,204,203,200_QL40_ML2_.jpg	Not Available
2022-04-23 13:45:25 +0000	Fantasie c piano	https://images-eu.ssl-images-amazon.com/images/I/41qQkChj52L_SX198_BO1,204,203,200_QL40_ML2_.jpg	9,03 €
2022-04-23 13:45:29 +0000	Adelaide: op. 46. mittlere Singstimme und Klavier. (Edition Schott Einzelausgabe)	https://images-eu.ssl-images-amazon.com/images/I/41zo5m7iCzL_SX198_BO1,204,203,200_QL40_ML2_.jpg	€6.50