Requirments of Dim Load Scrapper

2020年4月7日 16:37

I am looking for someone to help me download an online database(directory)-dimload.com using python and python libraries only.

Dimload.com is a directoy or money lenders in Hong Kong.

Objective:
The idea is to download the database into .csv, .json, and .xml data that can be manipulated by Microsoft access or imported into other databases. I need the data of both these in order to promote my services to them.

Programming language used:

1. Python-3

Potential Libraries:

1. Lxml
2. Json
3. Selenium
4. Beautiful Soup

Output:

1. Must Have CSV
2. Optional JSON or XML Output

Where to get the data?

2020年4月7日　16:45

The data should be scarpped based on the csv sheet column D.

For example:
For Line 5 of the file, "Sagarmatha Finance Company Limited" should be used so the email should be "http://www.dimloan.com/moneylender/Sagarmatha-Finance-Company-Limited"

Special Character and space handling
Space should be replaced by "-"
Special Character should be omitted.

ml_licensees

What data to get?

2020年4月7日 16:38

The data is in the dimloan.com website.

What kind of data we should get?

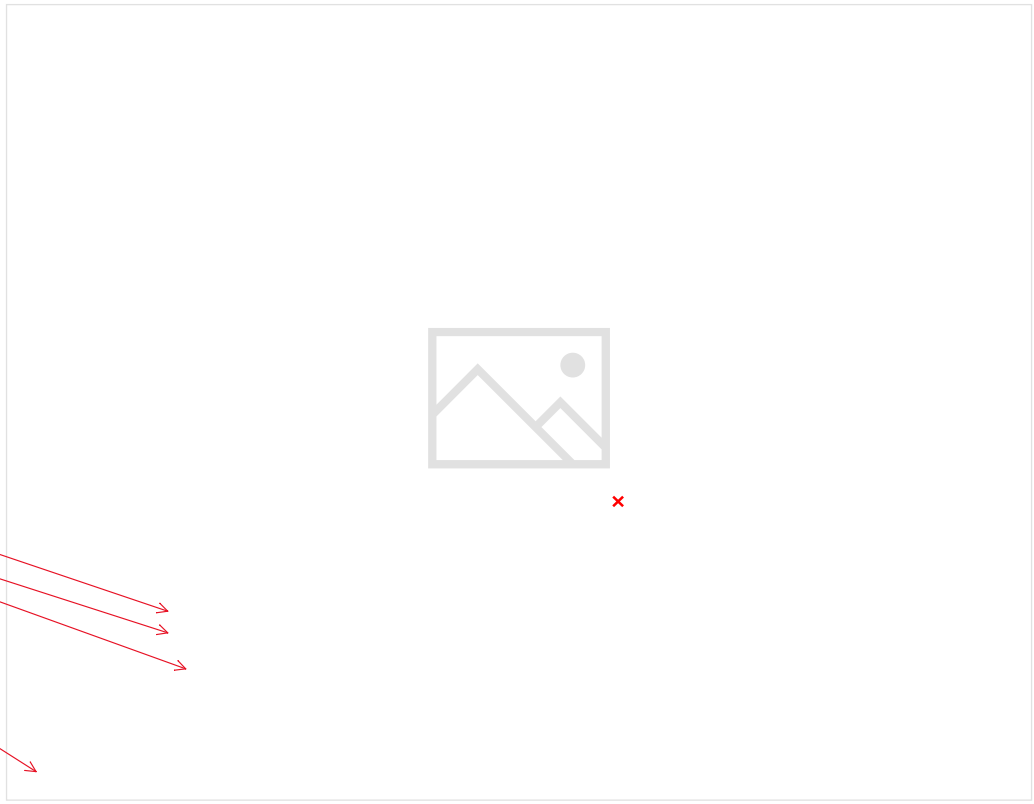
Name
Chinese Name:
English Name:

Licence Details
Money Lender Licence Status:
Money Lender Licence Valid Until:
Money Lender Licence No:

Contact Method
Website
Phone Number
Email

Registered Address
Registered Address

URL
URL



In rare occassions, there are more than one entries in address or phone number. For example https://www.dimloan.com/moneylender/Promise-Hong-Kong-Co-Limited

In that case, add another row as shown in the sample file.

Output Formating

2020年4月3日 16:20

Hints:

- Please remember to output the data as .csv, .xml and .json.
- The encode should be UTF-8.
- The output should include

- Website (This is what you need to scrape, and put into MLR No.)
- Phone Number (This is what you need to scrape)
- Email (This is what you need to scrape)
- Registered Address (This is what you need to scrape)

MLR_Email

MLR_Addr…

MLR_Web…

MLR_Phone