# Requirements of the MSO Licensee Register List Scapper

Saturday, February 29, 2020 4:46 PM

I am looking for someone to help me download an online database(directory)-The Licensee Register List of Money Services Operators using python and python libraries only.

The Customs of Hong Kong is responsible for licensing Money Service Operators (A.K.A Money Changers) 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

#### **Libraries:**

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

## **Output:**

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

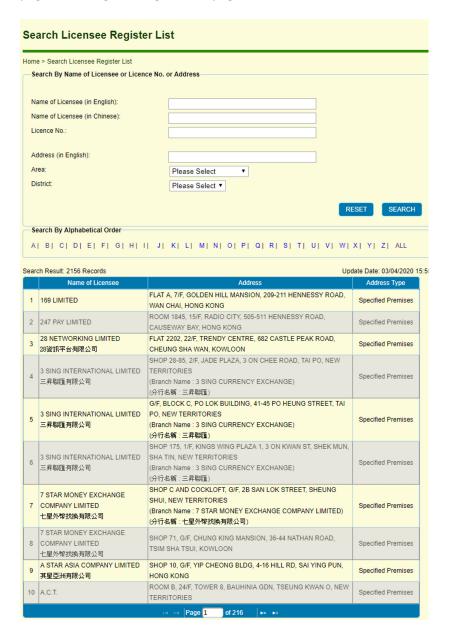
# Where is the data?

Saturday, February 29, 2020

4.51 PM

You can search for the The Licensee Register List of Money Services Operators here <a href="https://eservices.customs.gov.hk/MSOS/wsrh/001s1w">https://eservices.customs.gov.hk/MSOS/wsrh/001s1w</a>

Once you reach this page you can click "ALL" to find all the licencees, you can also click the pagination to go through all the pages.



# What data should be scrapped?

2020年4月3日

16:00



The output data should have a name MSO\_YYYYMMDD where YYYY is the year the scrapper scrapped results for, MM the month, DD the day.

#### In the file there are 8 Columns

MSOID	English Name of	English Trading	Chinese Name of	Chinese Trading	English	English Branch	Chinese Branch	Address Type
	Licensee	As	Licensee	As	Address	Name	Name	

I need you to fill in the data of the website into 8 columns,

- MSOID: A unique Identifier that identifies a unique record, a value we will use as a key in the database, I would prefer it to start with 0.
- English Name of Licensee: Is the english name of the licencesee, you will find it in the "Name of Licensee" column in your data.
- English Trading as: Is the trading name of the licencesee, some licencees will have it, some wont. For those that do, you will find it in the "Name of Licensee" column in your data. I want the data after "TRADING AS" to be included in this column.
- Chinese Name of Licensee: Is the Chinese name of the licencesee, you will find it in the "Name of Licensee" column in your data.
- Chinese Trading as: Is the trading name of the licencesee, some licencees will have it, some wont. For those that do, you will find it in the "Name of Licensee" column in your data. I want the data after "營業名稱" to be included in this column.
- English Address: Obviously this is the address, which you can find in the "Address" column in the data.
- English Branch Name: For some licencee, they will have a English branch name, look for the branch name in the address column.
- Chinese Branch Name: For some licensee, they will have a Chinese branch name, look for the branch name in the address column. I want the data after "分行名稱".
- Address Type: I want the Address type of the data to be included in this column.



# **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.



MSO\_2020...