

04/29/2020

DEVELOPER GUIDE

[HONEYDEW CONSULTING]

PREPARED BY: GROUP 7

VAIBHAV BHASIN

BIJAYALAXMI PANDA

MANIKANTA CHINTAKUNTA

RAJU KOUSHIK GORANTLA

TABLE OF CONTENTS

To Change Frontend	3
Q. How to add new websites to a search list?	3
Q. How to add new certifications keywords to a search list?	3
To Change Backend	4
Q. How to see more responses using google's custom search?	4
Q. How to add new fields from linkedin in terms of scrapping?	4
Q. Why sometimes I receive blank certifications columns in the database?	4
To Change Config file	5
Q. Where can I find config files to add user ids and passwords?	5
To Change Database	5
Q. How to edit in a database directly?	5
Q. How to add a new table to Database?	5

TO CHANGE FRONTEND

Q. HOW TO ADD NEW WEBSITES TO A SEARCH LIST?

All the HTML pages are generally kept in a “template” folder in Django. This folder in our project can be found in the crawler folder. The file will be home.html to add new websites. New options can be added to the select tag.

```
<div id="bloc1">
  <select name="apply_website" id="website">
    <option value="linkedin"> LinkedIn</option>
    <option value="monster"> Monster</option>
    <option value="indeed"> Indeed</option>
  </select>
</div>
```

Q. HOW TO ADD NEW CERTIFICATIONS KEYWORDS TO A SEARCH LIST?

All the HTML pages are generally kept in a “template” folder in Django. This folder in our project can be found in the crawler folder. The file will be home.html to add new websites. New options can be added to the select tag.

```
<div id="bloc2">
  <select name="submit_new" id="submit_new">
    <option value="epic beaker anatomic pathology">Epic Beaker Anatomic Pathology</option>
    <option value="epic beaker clinical pathology">Epic Beaker Clinical Pathology</option>
    <option value="epic radiant"> Epic Radiant</option>
    <option value="epic cupid"> Epic Cupid</option>
  </select>
</div>
```

This value attribute in option is important, as this is used by us to filter data from databases.

TO CHANGE BACKEND

Q. HOW TO SEE MORE RESPONSES USING GOOGLE'S CUSTOM SEARCH?

Currently we have designed it in such a way that it will scrape only 10 results from google. In future we can change it to any number, the code change will be in views.py::crawler() in Crawler app.

```
# making a request to a google custom search engine
custom_search_engine_url = "https://www.googleapis.com/customsearch/v1"

linkedin_url_list = set()

for i in range(0, 1):
    PARAMS = {'key': 'AIzaSyByUxDR0Y0701Y0ETLSJZn6bfFNWIjtQBM', 'cx': '009462381166450434430:ecyvn9zudgu',
              'q': query, 'start': i * 10}

    # sending get request and saving the response as response object
    r = requests.get(url=custom_search_engine_url, params=PARAMS)
```

Here i*10 suggests that 10 entries. This value can be changed to extract more profiles.

Q. HOW TO ADD NEW FIELDS FROM LINKEDIN IN TERMS OF SCRAPPING?

Linkedin does not like to be scrapped. That's why they change their HTML tags, attributes in every specific time period. So to add new attributes we suggest to read the HTML page carefully and then find the correct HTML tag to extract like we have done in views::get_individual_result_from_linkedin().

Q. WHY SOMETIMES I RECEIVE BLANK CERTIFICATIONS COLUMNS IN THE DATABASE?

First reason can be that the user does not have any entry in the certifications section in linkedin.

Second reason could be minimizing the chrome window during scrapping. Linkedin page takes time to load and expand all Javascript code to HTML. We have achieved this in code by scrolling. Once the window is minimized selenium can not do scrolling which implies no loading of HTML tags in the linkedin page.

So the solution will be not to minimize the chrome window during scrapping.

To CHANGE CONFIG FILE

Q. WHERE CAN I FIND CONFIG FILES TO ADD USER IDS AND PASSWORDS?

The configuration file name is “credentials.ini” located in the root folder itself. You can add and edit the details in this file.

Provide your user id and password to get started.

To CHANGE DATABASE

Q. HOW TO EDIT IN A DATABASE DIRECTLY?

Db.sqlite3 is the database file created by the Django framework. In case you are creating any new model(Table) in database following command will be helpful:

```
Python manage.py migrate
```

“DB Browser for sqlite3” is also a free software available to view these sqlite3 files. It also provides add/delete/update functionality for each table, row and column.

Q. HOW TO ADD A NEW TABLE TO DATABASE?

To do this you need to create a model in models.py file found in Crawler App. It will be similar in which we have created a Profile and RFP table. Then 3 commands will be useful;

```
python manage.py makemigrations crawler
```

```
python manage.py migrate
```

```
python manage.py runserver
```