
Python Email Crawler Documentation

Release 1.0

Junda Ong

Nov 20, 2017

Contents

1	Requirements	3
2	Usage	5
3	Indices and tables	7

This python script search certain keywords on Google, crawls the webpages from the results, and return all emails found.

For each result from Google, the crawler will crawl that page for an email. If it could not find an email, it will crawl the linked pages (up to 2nd level).

This is useful when the result returns the homepage of a website, and the email is usually in the Contact Us page.

CHAPTER 1

Requirements

- sqlalchemy
- urllib2

CHAPTER 2

Usage

Start the search with a keyword. We use “iphone developers” as an example.

```
$ ./email_crawler.py "iphone developers"
```

The search and crawling process will take quite a while, as it retrieve up to 500 search results (from Google), and crawl up to 2 level deep. It should crawl around 10,000 webpages :)

After the process finished, run this command to get the list of emails

```
$ ./email_crawler.py --emails
```

The emails will be saved in ./data/emails.csv

Contents:

CHAPTER 3

Indices and tables

- `genindex`
- `modindex`
- `search`