
WG Gesucht Crawler CLI Documentation

Release 0+untagged.11.g1adf23b.dirty

Grant Williams

Jul 07, 2018

Contents

1	WG Gesucht Crawler	3
1.1	Installation	3
1.2	Use	3
1.3	Features	4
2	Installation	5
3	Usage	7
3.1	Paramaters	7
4	Contributing	9
4.1	Types of Contributions	9
4.2	Get Started!	10
4.3	Pull Request Guidelines	10
5	Credits	13
5.1	Maintainer	13
5.2	Contributors	13
6	History	15
7	Indices and tables	17

Contents:

CHAPTER 1

WG Gesucht Crawler

Python web crawler / scraper for WG-Gesucht. Crawls the WG-Gesucht site for new apartment listings and send a message to the poster, based off your saved filters and saved text template.

1.1 Installation

```
$ pip install wg-gesucht-crawler-cli
```

Or, if you have virtualenvwrapper installed:

```
$ mkvirtualenv wg-gesucht-crawler-cli  
$ pip install wg-gesucht-crawler-cli
```

1.2 Use

Can be run directly from the command line with:

```
$ wg-gesucht-crawler-cli --help
```

Or if you want to use it in your own project:

```
from wg_gesucht.crawler import WgGesuchtCrawler
```

Just make sure to save at least one search filter as well as a template text on your wg-gesucht account.

- Free software: MIT license
- Documentation: <https://wg-gesucht-crawler-cli.readthedocs.org>.

1.3 Features

- Searches <https://wg-gesucht.de> for new WG ads based off your saved filters
- Sends your saved template message and applies to all matching listings
- Reruns every ~5 minutes
- Run on a RPi or free EC2 micro instance 24/7 to always be one of the first to apply for new listings

Getting Caught with reCAPTCHA

I've made the crawler sleep for 5-8 seconds between each request to try and avoid their reCAPTCHA, but if the crawler does get caught, you can sign into your wg-gesucht account manually through the browser and solve the reCAPTCHA, then start the crawler again. If it continues to happen, you can also increase the sleep time in the `get_page()` function in `wg_gesucht.py`

CHAPTER 2

Installation

At the command line:

```
$ pip install wg-gesucht-crawler-cli
```

Or, if you have virtualenvwrapper installed:

```
$ mkvirtualenv wg-gesucht-crawler-cli  
$ pip install wg-gesucht-crawler-cli
```


To use WG Gesucht Crawler in a project:

```
from wg_gesucht.crawler import WgGesuchtCrawler

crawler = WgGesuchtCrawler(login_info, ad_links_folder, offline_ad_folder, logs_
    ↪ folder)
crawler.sign_in()
crawler.search()
```

3.1 Paramaters

3.1.1 *login_info*

dict: containing wg-gesucht login details

keys: 'email', 'password', 'phone'(*optional*)

3.1.2 *ad_links_folder*

path to folder where a 'csv' file will be kept with previously applied for ads

3.1.3 *offline_ad_folder*

path to folder where offline ads will be saved

3.1.4 *logs_folder*

path to folder where the log files will be kept

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given. You can contribute in many ways:

4.1 Types of Contributions

4.1.1 Report Bugs

Report bugs at <https://github.com/grantwilliams/wg-gesucht-crawler-cli/issues>.

If you are reporting a bug, please include:

- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

4.1.2 Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with “bug” is open to whoever wants to implement it.

4.1.3 Implement Features

Look through the GitHub issues for features. Anything tagged with “feature” is open to whoever wants to implement it.

4.1.4 Write Documentation

WG Gesucht Crawler CLI could always use more documentation, whether as part of the official WG Gesucht Crawler CLI docs, in docstrings, or even on the web in blog posts, articles, and such.

4.1.5 Submit Feedback

The best way to send feedback is to file an issue at <https://github.com/grantwilliams/wg-gesucht-crawler-cli/issues>.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)

4.2 Get Started!

Ready to contribute? Here's how to set up *WG-Gesucht-Crawler-CLI* for local development.

1. Fork the *wg-gesucht-crawler-cli* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/wg-gesucht-crawler-cli.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv wg-gesucht-crawler-cli
$ cd wg-gesucht-crawler-cli/
$ python setup.py develop
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

Now you can make your changes locally.

5. When you're done making changes, check that your changes pass flake8 and the tests, including testing other Python versions with tox:

```
$ flake8 wg_gesucht tests
$ python setup.py test
$ tox
```

To get flake8 and tox, just pip install them into your virtualenv.

6. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -m "Your detailed description of your changes."
$ git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request through the GitHub website.

4.3 Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests.

2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring, and add the feature to the list in README.rst.
3. The pull request should work for Python 3.3, 3.4, 3.5, 3.6 and for PyPy. Check https://travis-ci.org/grantwilliams/wg-gesucht-crawler-cli/pull_requests and make sure that the tests pass for all supported Python versions.

5.1 Maintainer

- Grant Williams <grant.williams2986@gmail.com>

5.2 Contributors

None yet. Why not be the first? See: CONTRIBUTING.rst

CHAPTER 6

History

Pre-release

CHAPTER 7

Indices and tables

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