

## Scraper:

### Software required:

- Mysql (database)
- Myphp (server)
- Scrapy(python-framework)
- Httpd: Local host(While integration testing)

### Usage:

As of now There are 4 crawlers.

There are 4 spiders (amazon,shopclues,flipkart,shopclues)

To use it in your database change the password and the username for the database in the spider and the pipelines.py file located in every scraper .

The database commands are located in the db.py file.

The database file contains the table\_3 and the final price tables generated after the final database modification.

The urls have to be updated in the links table and the urls which aren't present for a specific item should be given a other url which is present in the zip file named unavailable url. Include the file link instead of the null row because the start\_urls need to have a url to execute.

After scraping the data from the websites run the db.py file to generate the table\_3 and price tables...

The php files showing the tables are included in the zip.

The complete url must be included so that it will match the url in the final tables of amazon,flipkart,snapdeal and shopclues.

### Commands for execution:

To Run the crawler:

```
scrapy crawl _name_(amazon,flip,snapdeal,shopclues){example: scrapy crawl amazon}
```

After running the crawler

Run:

```
python db.py //to generate the databae files
```

To see scrapy documentation visit:

<https://docs.scrapy.org/en/latest/#>