



Hi, I'm **sCrap**. The shopaholic robot by Lucka and Marketa. Here is how I work:

Wake me up by initiating a collector object. Please, specify the link of eshop homepage, category of products you are interested in an optionally also 2 execution parameters - frequency and number of iterations. By default, I will scrape the website only once and I wont update the generated table ever again. Keep in mind that my author only provided me with downloaders for czc.cz and datart.cz, so you will have to chose among those two.

The collector object:

```
class collect (link, category, frequency, iterations)
```



(link, category, frequency, iterations)



Prepare and execute of  
scraping plan

(link, category, timestamp)



Prepare and  
maintain the  
database

1. Run scraper (link, category)

2. Upload/update data in database

3. Wait (frequency)

(iterations)

Precious data

My inner scraping engine

0 Create generic eshop object

eshop class object (link, category)

1 Create brand specific web object

3 Request the main category homepage

5 Use the information  
to generate links to all product pages

6 Request all product page links

9 Store list of page specific BS objects

12 Consolidate the collected data in PD  
dataframe

Brand specific web downloader object

2 Generate link for the category homepage

4 Collect information form the category  
homepage bs object

7 Create brand specific page objects

Brand specific page object

8 Identify the product modules in page HTML  
and create a list of product modules

Brand specific product object

10 Self-initialization based on  
list of product soup modules

11 Identify the parameters of interest  
(title, price) and store them in dictionary



When the data collection is over, I will let you know! Actually, I will try to keep you informed during the entire process as well. And if something goes wrong, I'm taught to make a confession.