
pymorizon Documentation

Release 0.0.1

Michał Bartlakowski

Nov 23, 2017

Contents:

1	Introduction	1
1.1	Scraping category data	1
1.2	Scraping offer data	1
2	Category methods	3
3	Offer methods	5
4	Utils methods	7
5	Indices and tables	9
	Python Module Index	11

pymorizon supplies two methods that can be used to scrape data from Morizon website

1.1 Scraping category data

This method scrapes available offer urls from Morizon search results with parameters .. autofunction:: morizon.category.get_category

The function above can be used like this:

```
filters = {'[number_of_rooms_from]': 2'}
offers_url = morizon.category.get_category('mieszkania', 'Gdańsk', 'Grunwaldzka', 'do-
↳ wynajecia', None, filters)
```

The code above will put a list of urls containing all apartments found in the given category into the offers_url variable

1.2 Scraping offer data

This method scrapes details of offer .. autofunction:: morizon.offer.get_offer_data

The function above can be used like this:

```
details = morizon.offer.get_offer_data(url)
```

the code above will create dictionary with details of offer from given url

Category methods

`morizon.category.get_category(category='nieruchomosci', city=None, street=None, transaction_type=None, url=None, filters=None)`

Parses available offer urls from given category from every page

Parameters

- **category** (*str*, *None*) – type of property of interest (mieszkania/domy/garaże/działki)
- **city** (*str*, *None*) – city
- **street** (*str*, *None*) – street
- **transaction_type** (*str*, *None*) – type of transaction (sprzedaż/wynajem)
- **url** (*str*, *None*) – User defined url for Morizon page with offers. It overrides other parameters
- **filters** (*dict*) – Dictionary with additional filters.

Returns List of urls of all offers for given parameters

Return type list

`morizon.category.get_offers_from_page(url)`

Parses available offer urls from given category from given page

Parameters **url** (*str*) – Defined url for Morizon page with offers

Returns List of urls of offers from page

Return type list

CHAPTER 3

Offer methods

`morizon.offer.get_offer_data(url)`

Parse data from offer page url

Parameters `url` (*str*) – web page with offer

Returns Dictionary with details of an offer

Return type dict

CHAPTER 4

Utils methods

`morizon.utils.encode_text_to_url(text)`

Change text to lower cases, gets rid of polish characters replacing them with simplified version, replaces spaces with dashes

Parameters `text` (*str*) – raw text

Returns encoded text which can be used in url

Return type `str`

`morizon.utils.get_content_from_source(url)`

Connects with given url

If environmental variable `DEBUG` is `True` it will cache response for url in `/var/temp` directory

Parameters `url` (*str*) – Website url

Returns Response for requested url

`morizon.utils.get_max_page(url)`

Reads total page number on Morizon search page

Parameters `url` (*str*) – web page url

Returns number on sub web pages for search

Return type `int`

CHAPTER 5

Indices and tables

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

m

- `morizon.category`, 3
- `morizon.offer`, 5
- `morizon.utils`, 7

E

`encode_text_to_url()` (in module `morizon.utils`), [7](#)

G

`get_category()` (in module `morizon.category`), [3](#)

`get_content_from_source()` (in module `morizon.utils`), [7](#)

`get_max_page()` (in module `morizon.utils`), [7](#)

`get_offer_data()` (in module `morizon.offer`), [5](#)

`get_offers_from_page()` (in module `morizon.category`), [3](#)

M

`morizon.category` (module), [3](#)

`morizon.offer` (module), [5](#)

`morizon.utils` (module), [7](#)