tld Documentation

Release 0.8

Artur Barseghyan <artur.barseghyan@gmail.com>

Jun 13, 2018

Contents

Extract the top level domain (TLD) from the URL given. List of TLD names is taken from Mozilla.

Optionally raises exceptions on non-existing TLDs or silently fails (if `fail_silently` argument is set to True). Knows about active and inactive TLDs. If only active TLDs shall be matched against, `active_only` argument shall be set to True (default - False).

CHAPTER 1

Prerequisites

- Python 2.6, 2.7, 3.4, 3.5, 3.6 and PyPy

CHAPTER 2

Documentation

Documentation is available on Read the Docs.

CHAPTER 3

Installation

Latest stable version on PyPI:

```
pip install tld
```

Or latest stable version from GitHub:

```
pip install https://github.com/barseghyanartur/tld/archive/stable.tar.gz
```

Or latest stable version from BitBucket:

```
pip install https://bitbucket.org/barseghyanartur/tld/get/stable.tar.gz
```

CHAPTER 4

Usage examples

Get the TLD name **as string** from the URL given:

```
from tld import get_tld, get_fld

get_tld("http://www.google.co.uk")
# 'co.uk'

get_fld("http://www.google.co.uk")
# 'google.co.uk'

get_tld("http://www.google.idontexist", fail_silently=True)
# None

get_fld("http://www.google.idontexist", fail_silently=True)
# None
```

If you wish, you could get the TLD as **an object**:

```
from tld import get_tld

res = get_tld("http://some.subdomain.google.co.uk", as_object=True)

res
# 'co.uk'

res.subdomain
# 'some.subdomain'

res.domain
# 'google'

res.tld
# 'co.uk'
```

(continues on next page)

(continued from previous page)

```
res.fld
# 'google.co.uk'
```

Get TLD name, **ignoring the missing protocol**:

```
from tld import get_tld, get_fld

get_tld("www.google.co.uk", fix_protocol=True)
# 'co.uk'

get_fld("www.google.co.uk", fix_protocol=True)
# 'google.co.uk'
```

CHAPTER 5

Update the list of TLD names

To update/sync the tld names with the most recent version run the following from your terminal:

```
update-tld-names
```

Or simply do:

```
from tld.utils import update_tld_names

update_tld_names()
```

CHAPTER 6

Troubleshooting

If somehow domain names listed here are not recognised, make sure you have the most recent version of TLD names in your virtual environment:

```
update-tld-names
```

CHAPTER 7

Testing

Simply type:

```
./runtests.py
```

Or use tox:

```
tox
```

Or use tox to check specific env:

```
tox -e py36
```

CHAPTER 8

Writing documentation

Keep the following hierarchy.

```
=====
title
=====

header
======

sub-header
----------

sub-sub-header
~~~~~~~~~~~~~~

sub-sub-sub-header
^^^^^^^^^^^^^^^^^^

sub-sub-sub-sub-header
++++++++++++++++++++++

sub-sub-sub-sub-sub-header
**************************
```

CHAPTER 9

License

MPL 1.1/GPL 2.0/LGPL 2.1

CHAPTER 10

Support

For any issues contact me at the e-mail given in the *Author* section.

CHAPTER 11

Author

Artur Barseghyan <artur.barseghyan@gmail.com>

CHAPTER 12

Docs

Contents:

12.1 Release history and notes

Sequence based identifiers are used for versioning (schema follows below):

```
major.minor[.revision]
```

- It's always safe to upgrade within the same minor version (for example, from 0.3 to 0.3.4).
- Minor version changes might be backwards incompatible. Read the release notes carefully before upgrading (for example, when upgrading from 0.3.4 to 0.4).
- All backwards incompatible changes are mentioned in this document.

12.1.1 0.8

2018-06-13

Note: This release contains backward incompatible changes. You should update your code.

Old *get_tld* functionality remain is moved to *get_fld* (first-level domain definition). The *as_object* argument (False by default) has been deprecated for *get_fld*.

```
res = get_tld("http://www.google.co.uk", as_object=True)
```

New behaviour

```
In: res.domain
Out: 'google'
```

(continues on next page)

(continued from previous page)

```
In: res.extension
Out: 'co.uk'

In: res.subdomain
Out: 'www'

In: res.suffix
Out: 'co.uk'

In: res.tld
Out: 'google.co.uk'
```

Old behaviour

```
In: res.fld
Out: 'google.co.uk'

In: res.tld
Out: 'co.uk'

In: res.domain
Out: 'google'

In: res.subdomain
Out: 'www'
```

When used without `as_object` it returns `co.uk`.

Recap

If you have been happily using old version of *get_tld* function without *as_object* argument set to True, you might want to replace *get_tld* import with *get_fld* import:

```
# Old
from tld import get_tld
get_tld('http://google.co.uk')

# New
from tld import get_fld
get_fld('http://google.co.uk')
```

- Move to a Trie to match TLDs. This brings a speed up of 15-20%.
- It's now possible to search in public, private or all suffixes (old behaviour). Use *search_public* and *search_private* arguments accordingly. By default (to support old behavior), both are set to True.
- Correct TLD definitions.
- Domains like ******.xn–fiqs8s* are now recognized as well.
- Due to usage of *urlsplit* instead of *urlparse*, the initial list of TLDs is assembled quicker (a speed-up of 15-20%).
- Docs/ directory is included in source distribution tarball.
- More tests.

12.1.2 0.7.10

2018-04-07

- The `fix_protocol` argument respects protocol relative URLs.
- Change year in the license.
- Improved docstrings.
- TLD source updated to the latest version.

12.1.3 0.7.9

2017-05-02

- Added base path override for local .dat file.
- *python setup.py test* can used to execute the tests

12.1.4 0.7.8

2017-02-19

- Fix relative import in non-package for update-tls-names script. #15
- `get_tld` got a new argument `fix_protocol`, which fixes the missing protocol, having prepended "https" if missing or incorrect.

12.1.5 0.7.7

2017-02-09

- Tested against Python 3.5, 3.6 and PyPy.
- pep8 fixes.
- removed deprecated *tld.update* module. Use `update-tld-names` command instead.

12.1.6 0.7.6

2016-01-23

- Minor fixes.

12.1.7 0.7.5

2015-11-22

- Minor fixes.
- Updated tld names file to the latest version.

12.1.8 0.7.4

2015-09-24

- Exposed TLD initialization as `get_tld_names`.

12.1.9 0.7.3

2015-07-18

- Support for wheel packages.
- Fixed failure on some unicode domains.
- TLD source updated to the latest version.
- Documentation updated.

12.1.10 0.7.2

2014-09-28

- Minor fixes.

12.1.11 0.7.1

2014-09-23

- Force lower case of the URL for correct search.

12.1.12 0.7

2014-08-14

- Making it possible to obtain object instead of just extracting the TLD by setting the `as_object` argument of `get_tld` function to True.

12.1.13 0.6.4

2014-05-21

- Softened dependencies and lowered the `six` package version requirement to 1.4.0.
- Documentation improvements.

12.1.14 0.6.3

2013-12-05

- Speed up search

12.1.15 0.6.2

2013-12-03

- Fix for URLs with a port not handled correctly.
- Adding licenses.

12.1.16 0.6.1

2013-09-15

- Minor fixes.
- Credits added.

12.1.17 0.6

2013-09-12

- Fixes for Python 3 (Windows encoding).

12.1.18 0.5

2013-09-13

- Python 3 support added.

12.1.19 0.4

2013-08-03

- Tiny code improvements.
- Tests added.

12.2 tld package

12.2.1 Subpackages

12.2.1.1 tld.commands package

12.2.1.1.1 Submodules

12.2.1.1.2 tld.commands.update_tld_names module

tld.commands.update_tld_names.**main**()
Updates TLD names.

Example python src/tld/commands/update_tld_names.py

12.2.1.1.3 Module contents

12.2.2 Submodules

12.2.3 tld.conf module

12.2.4 tld.defaults module

12.2.5 tld.exceptions module

exception tld.exceptions.**TldBadUrl**(*url*)
Bases: exceptions.ValueError

TldBadUrl.

Supposed to be thrown when bad URL is given.

exception tld.exceptions.**TldDomainNotFound**(*domain_name*)
Bases: exceptions.ValueError

TldDomainNotFound.

Supposed to be thrown when domain name is not found (didn't match) the local TLD policy.

exception tld.exceptions.**TldImproperlyConfigured**(*msg=None*)
Bases: exceptions.Exception

TldImproperlyConfigured.

Supposed to be thrown when code is improperly configured. Typical use-case is when user tries to use *get_tld* function with both *search_public* and *search_private* set to False.

exception tld.exceptions.**TldIOError**(*msg=None*)
Bases: exceptions.IOError

TldIOError.

Supposed to be thrown when problems with reading/writing occur.

12.2.6 tld.helpers module

tld.helpers.**project_dir**(*base*)
Project dir.

tld.helpers.**PROJECT_DIR**(*base*)
Project dir.

12.2.7 tld.test module

12.2.8 tld.utils module

tld.utils.**get_fld**(*url*, *active_only=False*, *fail_silently=False*, *fix_protocol=False*, *search_public=True*, *search_private=True*, ***kwargs*)
Extract the first level domain.

Extract the top level domain based on the mozilla's effective TLD names dat file. Returns a string. May throw `TldBadUrl` or `TldDomainNotFound` exceptions if there's bad URL provided or no TLD match found respectively.

Parameters

- **url** (*str*) – URL to get top level domain from.
- **active_only** (*bool*) – If set to True, only active patterns are matched.
- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **fix_protocol** (*bool*) – If set to True, missing or wrong protocol is ignored (https is appended instead).
- **search_public** (*bool*) – If set to True, search in public domains.
- **search_private** (*bool*) – If set to True, search in private domains.

Returns String with top level domain (if `as_object` argument is set to False) or a `tld.utils.Result` object (if `as_object` argument is set to True); returns None on failure.

Return type str

`tld.utils.`**get_tld**(*url*, *active_only=False*, *fail_silently=False*, *as_object=False*, *fix_protocol=False*, *search_public=True*, *search_private=True*)

Extract the top level domain.

Extract the top level domain based on the mozilla's effective TLD names dat file. Returns a string. May throw `TldBadUrl` or `TldDomainNotFound` exceptions if there's bad URL provided or no TLD match found respectively.

Parameters

- **url** (*str*) – URL to get top level domain from.
- **active_only** (*bool*) – If set to True, only active patterns are matched.
- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **as_object** (*bool*) – If set to True, `tld.utils.Result` object is returned, `domain`, `suffix` and `tld` properties.
- **fix_protocol** (*bool*) – If set to True, missing or wrong protocol is ignored (https is appended instead).
- **search_public** (*bool*) – If set to True, search in public domains.
- **search_private** (*bool*) – If set to True, search in private domains.

Returns String with top level domain (if `as_object` argument is set to False) or a `tld.utils.Result` object (if `as_object` argument is set to True); returns None on failure.

Return type str

`tld.utils.`**get_tld_names**(*fail_silently=False*, *retry_count=0*)

Build the `tlds` list if empty. Recursive.

Parameters

- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **retry_count** (*int*) – If greater than 1, we raise an exception in order to avoid infinite loops.

Returns List of TLD names

Type iterable

tld.utils.**process_url**(*url*, *active_only=False*, *fail_silently=False*, *fix_protocol=False*, *search_public=True*, *search_private=True*)

Process URL.

Parameters

- **url** –
- **active_only** –
- **fail_silently** –
- **fix_protocol** –
- **search_public** –
- **search_private** –

Returns

class tld.utils.**Result**(*tld*, *domain*, *subdomain*)

Bases: object

Container.

domain

extension

Alias of tld.

Return str

fld

First level domain.

Returns

subdomain

suffix

Alias of tld.

Return str

tld

tld.utils.**update_tld_names**(*fail_silently=False*)

Update the local copy of TLDs file.

Parameters **fail_silently** (*bool*) – If set to True, no exceptions is raised on failure but boolean False returned.

Returns True on success, False on failure.

Return type bool

12.2.9 Module contents

tld.**get_fld**(*url*, *active_only=False*, *fail_silently=False*, *fix_protocol=False*, *search_public=True*, *search_private=True*, ***kwargs*)

Extract the first level domain.

Extract the top level domain based on the mozilla's effective TLD names dat file. Returns a string. May throw TldBadUrl or TldDomainNotFound exceptions if there's bad URL provided or no TLD match found respectively.

Parameters

- **url** (*str*) – URL to get top level domain from.
- **active_only** (*bool*) – If set to True, only active patterns are matched.
- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **fix_protocol** (*bool*) – If set to True, missing or wrong protocol is ignored (https is appended instead).
- **search_public** (*bool*) – If set to True, search in public domains.
- **search_private** (*bool*) – If set to True, search in private domains.

Returns String with top level domain (if `as_object` argument is set to False) or a `tld.utils.Result` object (if `as_object` argument is set to True); returns None on failure.

Return type str

`tld.`**`get_tld`**(*url*, *active_only=False*, *fail_silently=False*, *as_object=False*, *fix_protocol=False*, *search_public=True*, *search_private=True*)

Extract the top level domain.

Extract the top level domain based on the mozilla's effective TLD names dat file. Returns a string. May throw `TldBadUrl` or `TldDomainNotFound` exceptions if there's bad URL provided or no TLD match found respectively.

Parameters

- **url** (*str*) – URL to get top level domain from.
- **active_only** (*bool*) – If set to True, only active patterns are matched.
- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **as_object** (*bool*) – If set to True, `tld.utils.Result` object is returned, `domain`, `suffix` and `tld` properties.
- **fix_protocol** (*bool*) – If set to True, missing or wrong protocol is ignored (https is appended instead).
- **search_public** (*bool*) – If set to True, search in public domains.
- **search_private** (*bool*) – If set to True, search in private domains.

Returns String with top level domain (if `as_object` argument is set to False) or a `tld.utils.Result` object (if `as_object` argument is set to True); returns None on failure.

Return type str

`tld.`**`get_tld_names`**(*fail_silently=False*, *retry_count=0*)

Build the `tlds` list if empty. Recursive.

Parameters

- **fail_silently** (*bool*) – If set to True, no exceptions are raised and None is returned on failure.
- **retry_count** (*int*) – If greater than 1, we raise an exception in order to avoid infinite loops.

Returns List of TLD names

Type iterable

class tld.**Result**(*tld*, *domain*, *subdomain*)

Bases: object

Container.

domain

extension

Alias of tld.

Return str

fld

First level domain.

Returns

subdomain

suffix

Alias of tld.

Return str

tld

tld.**update_tld_names**(*fail_silently=False*)

Update the local copy of TLDs file.

Parameters **fail_silently** (*bool*) – If set to True, no exceptions is raised on failure but boolean False returned.

Returns True on success, False on failure.

Return type bool

CHAPTER 13

Indices and tables

- genindex
- modindex
- search

Python Module Index

t

Index