django-robots Documentation

Release 6.2.dev1+g7e90577

Jannis Leidel

Sep 07, 2023

Contents

Django Robots

This is a basic Django application to manage robots.txt files following the robots exclusion protocol, complementing the Django Sitemap contrib app.

For installation instructions, see the documentation install section; for instructions on how to use this application, and on what it provides, see the file "overview.txt" in the "docs/" directory or on ReadTheDocs: https://django-robots.readthedocs.io/

1.1 Supported Django versions

- Django 4.0
- Django 3.2
- Django 3.1
- Django 2.2

For older Django versions (1.6-1.10) use `django-robots==3.0.` for Django 2 and above, use `django-robots>=4.0.0.`

1.2 Supported Python version

- Python 3.7, 3.8, 3.9, 3.10

Contents:

1.2.1 History

This is where upcoming changes will be.

6.1 (2023-09-07)

Compatibility

- Fix warning make in regards to `pkg_resources` in Python 3.11+, (#146, thank you @blag)

Maintenance

- Fix typos via typos
  - Special thanks to typo fix from @apoorvaeternity (Apoorva Pandey) in #145

CI updates:

- Python: Support Python 3.11 and 3.12 in test matrix (#146, thank you @blag)
- Django: Support Django 4.1 and 4.2 in test matrix (#146, thank you @blag)

6.0 (2023-09-07)

Breaking change: Missing `DEFAULT_AUTO_FIELD`

This sets a `DEFAULT_AUTO_FIELD` to `BigAutoField` for robots in `AppConfig`. Previously, there was no app configuration or setting for this. (#134, thank you @jan-szejko-steelseries)

Robots 5.0 included a `BigAutoField` migration file for django 3.2 support (see #112), however, this caused an extra migration to be created to set `django.db.models.AutoField` for certain users (see #124).

Breaking: Migration issues

New django sites should work fine.

Existing sites that manually created migrations for robots 5.0:

If you ran `./manage.py makemigrations` to bypass the #112 issue, you may need to resolve this manually if a custom migration you made for robots was applied in `django_migrations`.

Resources:

- https://docs.djangoproject.com/en/4.1/releases/3.2/#customizing-type-of-auto-created-primary-keys
- https://docs.djangoproject.com/en/3.2/ref/settings/#std-setting-DEFAULT_AUTO_FIELD

Maintenance

- Remove django-south migrations folder (#138)

5.0 (2022-01-08)

- Dropped support for Python 2.7, 3.5, and 3.6
- Added support for Python 3.9
- Added support for Python 3.10 (#113), thank you @Andrew-Chen-Wang
- Dropped support for Django versions 2.1 and below
- Dropped support for Django 3.0
- Added support for Django 3.1
- Added support for Django 3.2
- Added support for Django 4.0 (#112), thank you @umarmughal824
- Improve documentation for Django 4.0 (#119), thank you @sergioisidoro
- Sync files with @jazzband/.github
- Add *.pre-commit-config.yaml* and run autofixes (#122)

4.0 (2020-01-04)

- Support for Django 2.1, 2.2, 3.0 and Python 3.7 and 3.8.
- Updated the requirements for dropped support of six in Django 3.
- Restructure test setup to use setuptools-scm and more modern Python patterns.

3.1.0 (2017-12-11)

- Add this changelog file
- Support for Django 2.0 via GH-83, fixes GH-81, GH-79
- Drop support for Django 1.10 and below. For older Django versions, use django-robots 3.0.0 and below
- Fix docs to include README in index instead of duplicating

3.0 (2017-02-28)

- Dropped support for Django < 1.8
- Added support for Django 1.10 / 1.11
- Improved admin changeform
- Added support for protocol prefix to Host directive
- Added support for sitemap named views (for non standard sitemap views)
- Fixed an error which resulted in doubling the scheme for sitemap
- Fixed support for cached sitemaps

2.0 (2016-02-28)

- Dropped support for Django 1.5
- Added support for Django 1.9
- Improved code / metadata quality
- Added Host directive
- Added support to detect current site via http host var
- Added filter_horizontal for for allowed and disallowed
- Fixed error in which get_sitemap_urls modifies SITEMAP_URLS
- Url patterns marked as safe in template
- disabled localization of decimal fields in template

1.1 (2015-05-12)

- Fixed compatibility to Django 1.7 and 1.8.
- Moved South migrations into different subdirectory so South>=1.0 is needed.

1.0 (2014-01-16)

- *BACKWARDS-INCOMPATIBLE* change: The default behaviour of this app has changed to **allow all bots** from the previous opposite behavior.
- Fixed some backward compatibility issues.
- Updated existing translations (Danish, German, French, Portuguese (Brasil), Russian).
- Added Greek, Spanish (Spain), Japanese, Dutch, Slovak and Ukrainian translations.

0.9.2 (2013-03-24)

- Fixed compatibility with Django 1.5. Thanks, Russell Keith-Magee.

0.9.1 (2012-11-23)

- Fixed argument signature in new class based view. Thanks, mkai.

0.9 (2012-11-21)

- Deprecated `ROBOTS_SITEMAP_URL` setting. Use `ROBOTS_SITEMAP_URLS` instead.
- Refactored `rule_list` view to be class based. django-robots now requires Django >= 1.3.
- Stop returning 404 pages if there are no Rules setup on the site. Instead disallow access for all robots.
- Added an initial South migration. If you're using South you have to "fake" the initial database migration:

```
python manage.py migrate --fake robots 0001
```

- Added initial Sphinx docs.

1.2.2 Installation

Use your favorite Python installer to install it from PyPI:

```
pip install django-robots
```

Or get the source from the application site at:

```
http://github.com/jazzband/django-robots/
```

To install the sitemap app, then follow these steps:

1. Add `'robots'` to your INSTALLED_APPS setting.

2. Make sure `'django.template.loaders.app_directories.Loader'` is in your TEMPLATES setting. It's in there by default, so you'll only need to change this if you've changed that setting.

3. Make sure you've installed the sites framework.

4. Run the `migrate` management command

1.3 Sitemaps

By default a `Sitemap` statement is automatically added to the resulting robots.txt by reverse matching the URL of the installed Sitemap contrib app. This is especially useful if you allow every robot to access your whole site, since it then gets URLs explicitly instead of searching every link.

To change the default behaviour to omit the inclusion of a sitemap link, change the ROBOTS_USE_SITEMAP setting in your Django settings file to:

```
ROBOTS_USE_SITEMAP = False
```

In case you want to use specific sitemap URLs instead of the one that is automatically discovered, change the ROBOTS_SITEMAP_URLS setting to:

```
ROBOTS_SITEMAP_URLS = [
    'http://www.example.com/sitemap.xml',
]
```

If the sitemap is wrapped in a decorator, dotted path reverse to discover the sitemap URL does not work. To overcome this, provide a name to the sitemap instance in `urls.py`:

```
urlpatterns = [
    ...
    re_path(r'^sitemap.xml$', cache_page(60)(sitemap_view), {'sitemaps': [...]}, name=
→'cached-sitemap'),
    ...
]
```

and inform django-robots about the view name by adding the following setting:

```
ROBOTS_SITEMAP_VIEW_NAME = 'cached-sitemap'
```

Use ROBOTS_SITEMAP_VIEW_NAME also if you use custom sitemap views (e.g.: wagtail custom sitemaps).

1.3.1 Initialization

To activate robots.txt generation on your Django site, add this line to your URLconf:

```
re_path(r'^robots\.txt', include('robots.urls')),
```

This tells Django to build a robots.txt when a robot accesses /robots.txt. Then, please sync your database to create the necessary tables and create Rule objects in the admin interface or via the shell.

1.3.2 Rules

Rule - defines an abstract rule which is used to respond to crawling web robots, using the robots exclusion protocol, a.k.a. robots.txt.

You can link multiple URL pattern to allows or disallows the robot identified by its user agent to access the given URLs.

The crawl delay field is supported by some search engines and defines the delay between successive crawler accesses in seconds. If the crawler rate is a problem for your server, you can set the delay up to 5 or 10 or a comfortable value for your server, but it's suggested to start with small values (0.5-1), and increase as needed to an acceptable value for your server. Larger delay values add more delay between successive crawl accesses and decrease the maximum crawl rate to your web server.

The sites framework is used to enable multiple robots.txt per Django instance. If no rule exists it automatically allows every web robot access to every URL.

Please have a look at the database of web robots for a full list of existing web robots user agent strings.

1.3.3 Host directive

By default a Host statement is automatically added to the resulting robots.txt to avoid mirrors and select the main website properly.

To change the default behaviour to omit the inclusion of host directive, change the ROBOTS_USE_HOST setting in your Django settings file to:

```
ROBOTS_USE_HOST = False
```

if you want to prefix the domain with the current request protocol (**http** or **https** as in Host: https://www.mysite.com) add this setting:

```
ROBOTS_USE_SCHEME_IN_HOST = True
```

1.3.4 URLs

Url - defines a case-sensitive and exact URL pattern which is used to allow or disallow the access for web robots. Case-sensitive.

A missing trailing slash does also match files which start with the name of the given pattern, e.g., '/admin' matches /admin.html too.

Some major search engines allow an asterisk (*) as a wildcard to match any sequence of characters and a dollar sign ($) to match the end of the URL, e.g., '/*.jpg$' can be used to match all jpeg files.

1.3.5 Caching

You can optionally cache the generation of the `robots.txt`. Add or change the `ROBOTS_CACHE_TIMEOUT` setting with a value in seconds in your Django settings file:

```
ROBOTS_CACHE_TIMEOUT = 60*60*24
```

This tells Django to cache the `robots.txt` for 24 hours (86400 seconds). The default value is `None` (no caching).

1.3.6 Developing

To run tests, create a virtualenv:

> virtualenv –python 3.10 .venv

Source it:

> . .venv/bin/activate

> # or if you use source source .venv/bin/activate

Install:

> pip install -e . pip install -r tests/requirements.txt pip install django

Run the tests:

> env PYTHONPATH=. DJANGO_SETTINGS_MODULE=tests.settings django-admin test robots -v2

1.3.7 Bugs and feature requests

As always your mileage may vary, so please don't hesitate to send feature requests and bug reports:

> https://github.com/jazzband/django-robots/issues