
django-html5-appcache Documentation

Release 0.4.1-dev

Iacopo Spalletti

May 16, 2017

Contents

1	Install	3
2	Usage	9
3	Indices and tables	15
	Python Module Index	19

This document refers to version 0.4.1-dev

Application to manage [HTML5 Appcache Manifest](#) files for dynamic Django web applications.

While handy and quite simple in its structure, manifest files is quite burdensome to keep up-to-date on dynamic websites.

`django-html5-appcache` try to make this effortless, exploiting the batteries included in Django to discover pages and assets as they are updated by the users.

See [Basic Concepts](#) for further details.

Installation

Requirements

- django>=1.4
- lxml
- html5lib

Installation

To get started using `django-html5-appcache` install it with `pip`:

```
$ pip install django-html5-appcache
```

If you want to use the development version install from github:

```
$ pip install git+https://github.com/nephila/django-html5-appcache.git#egg=django-  
→html5-appcache
```

Requirements will be automatically installed.

Run migrate command to sync your database:

```
$ python manage.py migrate html5_appcache
```

Warning: Migrations have been added in 0.3.0. Don't skip this if you are upgrading from 0.2.

Basic configuration

- Add `html5_appcache` to `INSTALLED_APPS`.
- Include in your `URLCONF`:

```
urlpatterns += patterns('',
    url('^', include('html5_appcache.urls')),
)
```

Warning: on Django 1.4+ (or django CMS 2.4+) you may need to use `urlpatterns` instead of `patterns` above, depending on you project layout.

- Enable appcache discovery by adding the lines below in `urls.py`:

```
import html5_appcache
html5_appcache.autodiscover()
```

- Add the middleware just below `django.middleware.cache.UpdateCacheMiddleware`, if used, or at the topmost position:

```
'html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse'
```

- Insert `appcache_link` template tag in your templates:

```
{% load appcache_tags %}
<html {% appcache_link %} >
<head>
...
</head>
<body>
...
</body>
</html>
```

- Enable the cache for your project. Refer to Django [CACHES](#) configuration.

django CMS integration

See *django CMS installation*.

Advanced configuration

While no specific configuration is needed to run `html5-appcache`, you can customize its behavior for your own needs with the following parameters:

HTML5_APPCACHE_DISABLE

If you want to keep `django-html5-appcache` installed but you want to disable it temporarily (for debug purposes, for example), set this parameter to `True`: it makes the `manifest` view return a non-caching manifest file and disables `appcache_link` templatetag.

New in version 0.3.0.

Defaults: False

HTML5_APPCACHE_ADD_WILDCARD

If `True` a wildcard entry is added in network section to allow browser to download files not in the `CACHE` section.

New in version 0.3.0.

Defaults: True

HTML5_APPCACHE_CACHE_KEY

Name of the cache key.

Defaults: `html5_appcache`

HTML5_APPCACHE_CACHE_DURATION

Duration of the cache values.

Default: 86400 seconds

HTML5_APPCACHE_USE_SITEMAP

`django-html5-appcache` can leverage the `sitemap` application of `django` to discover the cacheable urls. If you want to disable, you must provide a urls list.

Default: True

HTML5_APPCACHE_CACHED_URL

It's possible to provide a list of urls to include in the manifest file as cached urls, if it's not discoverable by the `django` application (e.g.: it's not managed by `django` or not linked to any page).

Default: []

HTML5_APPCACHE_EXCLUDE_URL

It's possible to exclude specific url from being cached by using this parameter. Contrary to [HTML5_APPCACHE_NETWORK_URL](#) URLs will be excluded by cached urls but are **not** set in the **NETWORK** section of the manifest. This way you can mask out *private* URLs or URLs that are not meant to be known.

Warning: This is **not** a security feature. *Security through obscurity* is broken by design. This parameter is intended only to have a cleaner and more concise manifest.

New in version 0.4.

Default: []

HTML5_APPCACHE_NETWORK_URL

You can exclude specific url from being cached by using this parameter. URLs will be excluded by cached urls and set in the **NETWORK** section of the manifest.

Default: []

HTML5_APPCACHE_FALLBACK_URL

It's possible to provide a dictionary of urls to be included in the **FALLBACK** section. Key is the *original* url, value is the *fallback* url.

Default: {}

HTML5_APPCACHE_OVERRIDE_URLCONF

When using **django CMS** apphooks, you must provide an alternative urlconf for `django-html5-appcache` to be able to traverse the application urls, due to way apphooks works. See the **django CMS integration** section to know more (WiP)

Default: False

HTML5_APPCACHE_OVERRIDDEN_URLCONF

This is used internally by `django-html5-appcache` and should remain to its default value.

Default: False

Changelog

0.4.1 (2014-01-05)

- Django 1.6 support

0.4 (2013-06-26)

- Manifest file is now authentication-sensitive (see docs)
- Add data-appcache optin parameter
- Add HTML5_APPCACHE_EXCLUDE_URL setting
- Fix HTML5_APPCACHE_NETWORK_URL and HTML5_APPCACHE_FALLBACK_URL settings

0.3.1 (2013-06-08)

- Fix view-generated manifest

0.3.0 (2013-06-06)

Warning: 0.3.0 introduces migrations. Run `migrate html5_appcache` on upgrade

- Special permissions for management views
- `Templatetag` to show the cache status and update the manifest (see *Web interface*)
- `HTML5_APPCACHE_DISABLE` parameter to disable manifest file (see *Advanced configuration*)

0.2.2 (2013-06-02)

- Fixes issue with Google Chrome

0.2.0 (2013-06-02)

- Initial release

Basic Concepts

`django-html5-appcache` leverages `django cache`, `test` and `signals` frameworks to browse project urls and assets and generate an appcache manifest file.

Manifest file generation

The manifest file is generated collecting all the cached urls and exploring them using the test client to extract asset urls and including them in the manifest itself.

As this can be quite resource intensive, the manifest file is saved in the project cache; the view that actually delivers the file manifest to the browser can thus use the project cache to serve it with no performance impact.

The manifest file can be generated out-of-band using a *django command*, to execute the command manually or in a cron job, or using *Web interface* (since version 0.3.0).

Cache invalidation

Whenever a registered model is saved or deleted (see *Enabling caching in your application* on how to enable this for your application), manifest cache is marked as **dirty**; this has no immediate effect on the manifest file served, as the outdated copy is still served.

To actually refresh the manifest file served to users, it's necessary to regenerate it (see above).

URL discovery

Using sitemap

`django-html5-appcache` uses the sitemap as a primary mean to discover urls in the web application.

This is a two steps process:

1. get the sitemap and extract the urls declared
2. scrape each url and extract the asset urls

In the scraping phase, the actual HTML of each page is generated and analyzed.

Currently `django-html5-appcache` extracts data from `img`, `script` and `link` tags. See [AppCacheAssetsFromResponse](#) for more in depth details.

See *In the Markup* on how to customize the assets extraction in your markup.

Customizing urls

Additional to the sitemap method above, you can define your own *custom url list*; in this case, it's your duty to define the list of assets in those urls.

Enabling caching in your application

`django-html5-appcache` will automatically include your application urls in the manifest file the if you have a sitemap-enabled application; however, to enable cache invalidation, is strongly advised to explicitly enable appcache support in your application.

Basic support

For basic appcache support, you must create a `appcache.py` in your application directory (along `models.py` file), write an `AppCache` class and register it:

```
from html5_appcache import appcache_registry
from html5_appcache.appcache_base import BaseAppCache

from .models import MyModel, AnotherModel

class MyModelAppCache(BaseAppCache):
    models = (MyModel, AnotherModel)

    def signal_connector(self, instance, **kwargs):
        self.manager.reset_manifest()
appcache_registry.register(MyModelAppCache())
```

This code declare support for `MyModel` and `AnotherModel` and hooks `MyModelAppCache`. `signal_connector` with `post_save` and `post_delete` signals.

Anytime you save or delete an instance of `MyModel` and `AnotherModel` cache will be marked as **dirty**.

Custom urls support

If you don't have a sitemap or you just want to customize the urls in the manifest file, you can add methods to the basic `AppCache` class above:

```
class MyModelAppCache(BaseAppCache):
    ...

    def _get_urls(self, request):
        ...
```

```

    return urls

def _get_assets(self, request):
    ...
    return urls

def _get_network(self, request):
    ...
    return urls

def _get_fallback(self, request):
    ...
    return urls

```

- `_get_urls(self, request)`: returns a list of urls to be included in the `CACHE` section of the manifest file;
- `_get_assets(self, request)`: returns a list of asset urls to be included in the `CACHE` section of the manifest file; if you add urls in `_get_urls` method, you have to return the assets in the above urls in this method;
- `_get_network(self, request)`: returns a list of urls to be included in the `NETWORK` section of the manifest file;
- `_get_fallback(self, request)`: returns a dictionary of urls to be included in the `FALLBACK` section of the manifest file; the dictionary key is used as the leftmost url in each manifest row, the value as the rightmost (i.e: the manifest instruct browser to substitute `key` url with `value` url when offline).

`request` object is passed for convenience

django CMS plugins

See *django CMS plugins*.

django CMS

django-html5-appcache supports django CMS out-of-the-box.

django CMS integration delivers support for all the the default plugins; to enable your own plugins see *django CMS plugins* below.

Installation

Plugins

To enable, add the following to `INSTALLED_APPS`:

- `html5_appcache.packages.cms`
- `html5_appcache.packages.filer` (if you use `django-filer`)
- `html5_appcache.packages.cmsplugin_filer` (if you use `cmsplugin_filer`)

Apphooks

If you use applications hooked to django CMS **AppHooks**, you have to write the *AppCache* class; if you use the `sitemap` method to discover the urls, you **must** add conditional urls loading to the main `urls.py`.

As the scraping uses the internal testserver to deliver the contents, **Apphooks** are not *hooked* so you have to provide an alternate method to attach the urls.

For this purpose use the following snippet:

```
if getattr(settings, 'HTML5_APPCACHE_OVERRIDDEN_URLCONF', False):
    urlpatterns += patterns('',
        url(r'^my-url', include("my-app.urls")),
        ...
    )
```

Where *my-url* is the url where the apphook is attached to, and *my-app.urls* is the urlconf of you application. Repeat for every attached apphook and for every slug they are attached to.

django CMS plugins

To enable cache invalidation for your own plugins, you must create an *AppCache* class for your plugin models too.

The example below is the implementation of an appcache for django CMS text plugin:

```
from html5_appcache import appcache_registry
from html5_appcache.appcache_base import BaseAppCache
from cms.plugins.text.models import Text

class CmsTextAppCache(BaseAppCache):
    models = (Text, )

    def signal_connector(self, instance, **kwargs):
        self.manager.reset_manifest()
appcache_registry.register(CmsTextAppCache())
```

Managing urls in the manifest

Sometimes you don't want urls to be cached for various reasons (they can pull content from external sites with no way to invalidate the local cache, or they are just non meant to be available offline), or you want to insert non-discoverable urls in it.

As there is not one-size-fit-all in managing urls in manifest, `django-html5-appcache` offers different methods to get the urls **in** or **out** of the manifest file to meet as many usecases as possible.

Via configuration

To include urls in the manifest, use `HTML5_APPCACHE_CACHED_URL`, to exclude them use `HTML5_APPCACHE_EXCLUDE_URL`.

To insert a URL in **NETWORK** see `HTML5_APPCACHE_NETWORK_URL`; for **FALLBACK** see `HTML5_APPCACHE_FALLBACK_URL`.

Via AppCache class

In the `AppCache` classes, is it possible to override methods to fine-tune the urls in each section of the manifest file:

- `BaseAppCache._get_urls`: to add urls to the **CACHE** section
- `BaseAppCache._get_assets`: to add assets to the **CACHE** section
- `BaseAppCache._get_network`: to add urls to the **NETWORK** section
- `BaseAppCache._get_fallback`: to add urls to the **FALLBACK** section

In the Markup

When using *sitemap discovery*, by default every relative URL is considered to be cached, while external URLs are not cached. It's possible to control the behavior of each url by using custom attributes in your tags.

For each `img`, `script` and `link` tag, you can add data-attributes to control how each referenced url is considered:

- `data-appcache='appcache'`: the referenced url is added to the **CACHE** section
- `data-appcache='noappcache'`: the referenced url is added to the **NETWORK** section
- `data-appcache-fallback=URL`: the referenced url is added in the **FALLBACK** section, with *URL* as a target

Web interface

Since 0.3.0 `django-html5-appcache` has a small web interface to check the cache status and update the manifest file.

The `appcache_icon` templatag show the cache status icon and hooks it to an ajax call that trigger the manifest update.

Badges

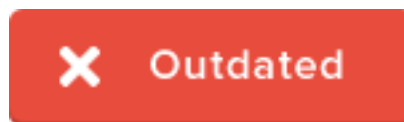


Fig. 2.1: Outdated cache status badge



Fig. 2.2: Up-to-date cache status badge

Installation

Add the following lines to any template you want the cache status badge to appear:

```
{% load appcache_tags %}
...
...
{% appcache_link %}
```

Permissions

Both the view that shows the cache status and the view to update the manifest are subject to specific permissions:

- `can_view_cache_status`: required to access the view that show the cache status
- `can_update_manifest`: required to trigger the manifest update

You need to explicitly add these permissions to any user who manages the appcache.

Both the views and the templatetag checks this permissions, so you can actually write your own code to call the views and your code will still be safe.

Command line usage

`django-html5-appcache` define two commands to control the manifest cache:

`update_manifest`

`update_manifest` updates the manifest cache:

```
$ python manage.py update_manifest
```

`clear_manifest`

Mostly a debugging tool, `clear_manifest` wipe the manifest cache completely.

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

Autodoc

class `html5_appcache.appcache_base.AppCacheManager`

Main class.

`_fetch_url (client, url)`

Scrape a single URL and fetches assets

`_get_sitemap ()`

Pretty ugly method that fetches the current sitemap and parses it to retrieve the list of available urls

`_setup_signals ()`

Loads the signals for all the models declared in the appcache instances

`add_appcache (appcache)`

Adds the externally retrieved urls to the internal set.

appcache is a dictionary with *cached*, *network*, *fallback* keys

`extract_urls ()`

Run through the cached urls and fetches assets by scraping the pages

`get_cached_urls ()`

Create the cached urls set.

Merges the assets, the urls, removes the network urls and the external urls

See `BaseAppCache.get_urls ()`, `get_network_urls ()`

get_fallback_urls ()

Creates the fallback urls set.

get_manifest (*update=False*)

Either get the manifest file out of the cache or render it and save in the cache.

get_network_urls ()

Create the network urls set.

* (wildcard entry) is added when `ADD_WILDCARD` is `True` (default)

get_urls ()

Retrieves the urls from the sitemap and `BaseAppCache.get_urls()` of the appcache instances

get_version_timestamp ()

Create the timestamp according to the current time.

It tries to make it unique even for very short timeframes

reset_manifest ()

Clear the cache (if clean)

setup (*request, template*)

Setup is required wen updating the manifest file

setup_registry ()

Setup the manager bootstrapping the appcache instances

class `html5_appcache.appcache_base.BaseAppCache`

Base class for Appcache classes

_add_language (*request, urls*)

For django CMS 2.3 we need to manually add language code to the urls returned by the appcache classes

Returns list of urls

_get_assets (*request*)

override this method to customize asset (images, files, javascripts, stylesheets) urls.

Returns list of urls

_get_fallback (*request*)

override this method to define fallback urls.

Returns dictionary mapping original urls to fallback

_get_network (*request*)

override this method to define network (non-cached) urls.

Returns list of urls

_get_urls (*request*)

override this method to define cached urls.

If you use a sitemap-enabled application, it's not normally necessary.

Returns list of urls

get_assets (*request*)

Public method that return assets urls. Do not override, use `_get_assets()`

Returns list of urls

get_fallback (*request*)

Public method that return fallback urls. Do not override, use `_get_fallback()`

Returns dictionary mapping original urls to fallback

get_network (*request*)

Public method that return network (non-cached) urls. Do not override, use `_get_network()`

Returns list of urls

get_urls (*request*)

Public method that return cached urls. Do not override, use `_get_urls()`

Returns list of urls

signal_connector (*instance, **kwargs*)

You **must** override this method in you AppCache class.

class `html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse`

Extracts appcache assets from the rendered template.

Currently supports the following tags:

- `img`: extracts the data in the `src` attribute
- `script`: extracts the data in the `src` attribute
- `link`: extracts the data in the `href` attribute if `rel==stylesheet`

It supports custom data-attribute to exclude assets from caching:

- `data-appcache='noappcache'`: the referenced url is added to the NETWORK section
- `data-appcache='appcache'`: the referenced url is added to the CACHE section
- `data-appcache-fallback=URL`: the referenced url is added in the FALLBACK section, with `URL` as a target

handle_a (*tag, attrib*)

Extract assets from the a tag (only for opt-in link)

handle_img (*tag, attrib*)

Extract assets from the img tag

handle_link (*tag, attrib*)

Extract assets from the link tag (only for stylesheets)

handle_script (*tag, attrib*)

Extract assets from the script tag

process_response (*request, response*)

This method is called only if `appcache_analyze` parameter is attached to the querystring, to avoid overhead during normal navigation

walk_tree (*tree*)

Walk the DOM tree

h

`html5_appcache.appcache_base`, [15](#)
`html5_appcache.middleware.appcache_middleware`,
[17](#)

Symbols

_add_language() (html5_appcache.appcache_base.BaseAppCache method), 16
 _fetch_url() (html5_appcache.appcache_base.AppCacheManager method), 15
 _get_assets() (html5_appcache.appcache_base.BaseAppCache method), 16
 _get_fallback() (html5_appcache.appcache_base.BaseAppCache method), 16
 _get_network() (html5_appcache.appcache_base.BaseAppCache method), 16
 _get_sitemap() (html5_appcache.appcache_base.AppCacheManager method), 15
 _get_urls() (html5_appcache.appcache_base.BaseAppCache method), 16
 _setup_signals() (html5_appcache.appcache_base.AppCacheManager method), 15
 get_cached_urls() (html5_appcache.appcache_base.AppCacheManager method), 15
 get_fallback() (html5_appcache.appcache_base.BaseAppCache method), 16
 get_fallback_urls() (html5_appcache.appcache_base.AppCacheManager method), 15
 get_manifest() (html5_appcache.appcache_base.AppCacheManager method), 16
 get_network_urls() (html5_appcache.appcache_base.AppCacheManager method), 16
 get_urls() (html5_appcache.appcache_base.AppCacheManager method), 16
 get_urls() (html5_appcache.appcache_base.BaseAppCache method), 16
 get_urls() (html5_appcache.appcache_base.BaseAppCache method), 17
 get_version_timestamp() (html5_appcache.appcache_base.AppCacheManager method), 16

A

add_appcache() (html5_appcache.appcache_base.AppCacheManager method), 15
 AppCacheAssetsFromResponse (class in html5_appcache.middleware.appcache_middleware), 17
 AppCacheManager (class in html5_appcache.appcache_base), 15
 handle_a() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17
 handle_img() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17

B

BaseAppCache (class in html5_appcache.appcache_base), 16
 handle_link() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17
 handle_script() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17

E

extract_urls() (html5_appcache.appcache_base.AppCacheManager method), 15
 HTML5_APPCACHE_ADD_WILDCARD setting, 5
 HTML5_APPCACHE_CACHE_DURATION setting, 5

G

get_assets() (html5_appcache.appcache_base.BaseAppCache method), 16
 HTML5_APPCACHE_CACHE_KEY setting, 5
 HTML5_APPCACHE_CACHED_URL setting, 5

HTML5_APPCACHE_DISABLE

setting, 4

HTML5_APPCACHE_EXCLUDE_URL

setting, 5

HTML5_APPCACHE_FALLBACK_URL

setting, 6

HTML5_APPCACHE_NETWORK_URL

setting, 5

HTML5_APPCACHE_OVERRIDDEN_URLCONF

setting, 6

HTML5_APPCACHE_OVERRIDE_URLCONF

setting, 6

HTML5_APPCACHE_USE_SITEMAP

setting, 5

P

process_response() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17

R

reset_manifest() (html5_appcache.appcache_base.AppCacheManager method), 16

S

setting

HTML5_APPCACHE_ADD_WILDCARD, 5

HTML5_APPCACHE_CACHE_DURATION, 5

HTML5_APPCACHE_CACHE_KEY, 5

HTML5_APPCACHE_CACHED_URL, 5

HTML5_APPCACHE_DISABLE, 4

HTML5_APPCACHE_EXCLUDE_URL, 5

HTML5_APPCACHE_FALLBACK_URL, 6

HTML5_APPCACHE_NETWORK_URL, 5

HTML5_APPCACHE_OVERRIDDEN_URLCONF,
6

HTML5_APPCACHE_OVERRIDE_URLCONF, 6

HTML5_APPCACHE_USE_SITEMAP, 5

setup() (html5_appcache.appcache_base.AppCacheManager method), 16

setup_registry() (html5_appcache.appcache_base.AppCacheManager method), 16

signal_connector() (html5_appcache.appcache_base.BaseAppCache method), 17

W

walk_tree() (html5_appcache.middleware.appcache_middleware.AppCacheAssetsFromResponse method), 17