uTidylib Documentation

Release 0.10

uTidylib contributors

Mar 05, 2024

CONTENTS

1 Installing **3**

2 Contributing **5**

3 Running testsuite **7**

4 Building documentation **9**

5 License **11**

6 Changes **13**
- 6.1 0.10 . . . 13
- 6.2 0.9 . . . 13
- 6.3 0.8 . . . 13
- 6.4 0.7 . . . 14
- 6.5 0.6 . . . 14
- 6.6 0.5 . . . 14
- 6.7 0.4 . . . 14
- 6.8 0.3 . . . 14

7 History **15**

8 Indices and tables **17**

Python Module Index **19**

Index **21**

The Tidy wrapper.

I am the main interface to TidyLib. This package supports processing HTML with Tidy, with all the options that the tidy command line supports.

For more information on the tidy options, see the reference. These options can be given as keyword arguments to parse and parseString, by changing dashes (-) to underscores(_).

For example:

```
>>> import tidy
>>> from __future__ import print_function
>>> print(tidy.parseString(
...     '<Html>Hello Tidy!',
...     output_xhtml=1, add_xml_decl=1, indent=1, tidy_mark=0,
...     doctype='transitional'
... ))
<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title></title>
  </head>
  <body>
    Hello Tidy!
  </body>
</html>
```

For options like newline and output_encoding, which must be set to one of a fixed number of choices, you can provide either the numeric or string version of the choice; so both tidy.parseString('<HTML>foo</html>', newline=2) and tidy.parseString('<HTML>foo</html>', newline='CR') do the same thing.

There are no plans to support other features of TidyLib, such as document-tree traversal, since Python has several quality DOM implementations. (The author uses Twisted's implementation, twisted.web.microdom).

`tidy.parse`(*filename: str*, ***kwargs: OPTION_TYPE*) → *Document*

Open and process filename as an HTML file.

Returning a processed document object.

Parameters

- **kwargs** – named options to pass to TidyLib for processing the input file.
- **filename** – the name of a file to process

Returns

a *Document* object

`tidy.parseString`(*text: bytes | str*, ***kwargs: OPTION_TYPE*) → *Document*

Use text as an HTML file.

Returning a processed document object.

Parameters

- **kwargs** – named options to pass to TidyLib for processing the input file.
- **text** – the string to parse

Returns
a `Document` object

class tidy.**Document**(*options: OPTION_DICT_TYPE*)

Document object as returned by `parseString()` or `parse()`.

get_errors() → list[*ReportItem*]

Return list of errors as a list of `ReportItem`.

gettext() → str

Unicode text for output returned by tidy.

getvalue() → bytes

Raw string as returned by tidy.

write(*stream: BinaryIO*) → None

Parameters
stream – Writable file like object.

Writes document to the stream.

class tidy.**ReportItem**(*err: str*)

Error report item as returned by tidy.

col: int | None

Column where error was fired (can be None)

err: str

Whole error message as returned by tidy

full_severity: str

Full severity string

line: int | None

Line where error was fired (can be None)

message: str

Error message itself

severity: str

D, W, E or C indicating severity

exception tidy.**TidyLibError**

Generic Tidy exception.

exception tidy.**InvalidOptionError**

Exception for invalid option.

exception tidy.**OptionArgError**

Exception for invalid parameter.

CHAPTER

ONE

INSTALLING

To use uTidylib, you need to have HTML tidy library installed. Check <http://www.html-tidy.org/> for instructions how to obtain it.

Once you have installed the library, install uTidylib:

```
pip install uTidylib
```

CHAPTER

TWO

CONTRIBUTING

You are welcome to contribute on GitHub, we use it for source code management, issue tracking and patches submission, see <https://github.com/nijel/utidylib>.

CHAPTER

THREE

RUNNING TESTSUITE

The testsuite can be exececuted using pytest:

```
pytest tidy
```

CHAPTER

FOUR

BUILDING DOCUMENTATION

To build the doc, just run:

```
make -C docs html
```

This requires that you have Sphinx installed.

The API documentation will be built in the `docs/_build/html/` directory.

CHAPTER

FIVE

LICENSE

The MIT License

Copyright (c) 2003 Cory Dodt <corydodt@twistedmatrix.com>

Copyright (c) 2014-2018 Michal Čihař <michal@cihar.com>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

CHAPTER

SIX

CHANGES

6.1 0.10

- Dropped support for Python 3.7.
- Added support for Python 3.12.
- Added type hints.
- Improved documentation.
- Always call CleanAndRepair after parsing.
- Fixed handling char_encoding argument.

6.2 0.9

- Dropped support for Python 3.6.
- Added support for Python 3.10 and 3.11.
- Compatibility with html-tidy 5.8.0.
- Added support for specifying library full path using TIDY_LIBRARY_FULL_PATH.
- Added getTidyVersion to get libtidy version.

6.3 0.8

- Code cleanups.
- Fixed typo in 0.7 release notes.

6.4 0.7

- Dropped support for Python 2.

6.5 0.6

- First official release PyPI.

6.6 0.5

- Fixed compatibility with Debian patched libtidy5deb1.

6.7 0.4

- Compatibility with html-tidy 5.6.0.
- Added support for Python 3.

6.8 0.3

- Initial release under new maintainer.
- Incorporated Debian patches.
- Various compatiblity fixes (eg. with 64-bit machines).
- Various code cleanups.
- New test suite.
- New documentation.
- Support for new HTML 5 tidy library.

CHAPTER

SEVEN

HISTORY

This is fork of the original uTidylib with permission with original author. Originally it incorporated patches from Debian and other distributions, now it also brings compatibility with recent html-tidy versions and works with Python 3.

The original source code is still available at https://github.com/xdissent/utidylib/.

CHAPTER

EIGHT

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

t

tidy, **??**

INDEX

C

col (*tidy.ReportItem attribute*), 2

D

Document (*class in tidy*), 2

E

err (*tidy.ReportItem attribute*), 2

F

full_severity (*tidy.ReportItem attribute*), 2

G

get_errors() (*tidy.Document method*), 2
gettext() (*tidy.Document method*), 2
getvalue() (*tidy.Document method*), 2

I

InvalidOptionError, 2

L

line (*tidy.ReportItem attribute*), 2

M

message (*tidy.ReportItem attribute*), 2
module
    tidy, 1

O

OptionArgError, 2

P

parse() (*in module tidy*), 1
parseString() (*in module tidy*), 1

R

ReportItem (*class in tidy*), 2

S

severity (*tidy.ReportItem attribute*), 2

T

tidy
    module, 1
TidyLibError, 2

W

write() (*tidy.Document method*), 2