
tagterm Documentation

Release 0.1

Cosmin Poiana

May 25, 2016

1	Intro	3
1.1	Tagterm	3
2	Installation	5
2.1	Preparing the system	5
2.2	Clone & install	5
2.3	Build <i>tagterm</i> executable	6
2.4	Build Java wrapper	6
3	Usage	7
3.1	Python	7
3.2	Java	7
3.3	Tags	8
4	API	9
4.1	Python	9
4.2	Java	9
5	Indices and tables	11

Contents:

Intro

Here's a quick guide on what's **tagterm** all about and how to install and/or use it alongside your project.

1.1 Tagterm

Tagterm name stands for tag terminator, which is a piece of software capable of:

- validating & repairing HTML files
- converting them into XHTML format
- removing unwanted tags while keeping the content

In order to use it, you'll have to install the current project, then to build an executable (or not) from the main script or just use the currently available ones inside the *bin* directory. For more details about this process, please read *installation* and *usage*.

Installation

The whole component is written in Python and needs this interpreter and some external libraries and packages to run and build properly. On top of a final Linux/Windows executable built upon this package, there's a Java source under the *src* directory which acts as a wrapper over the Python code/executable for doing some basic processes over (X)HTML files. More details about the API [here](#).

2.1 Preparing the system

Linux

```
sudo apt update
sudo apt install --upgrade python python-dev python-setuptools libtidy-dev
sudo -H easy_install -U pip
```

Windows

1. Download and install Python interpreter: <https://www.python.org/downloads/release/python-2711/>
2. Install PIP: <https://pip.pypa.io/en/stable/installing/>

Note: Tidy library dependency will be copied and registered by installing the package.

2.2 Clone & install

Linux

```
git clone https://github.com/cmin764/tagterm.git
cd tagterm

sudo -H pip install -Ur requirements.txt
sudo ./setup.sh
```

Windows

```
git clone https://github.com/cmin764/tagterm.git
cd tagterm

pip install -Ur requirements.txt
```

```
rem Run this as Administrator:
setup.bat
```

Note: You can also use *virtualenv(wrapper)* to install the package and related libraries.

In order to make sure that everything works as expected, refer to the *usage* examples.

2.3 Build *tagterm* executable

For this to work properly, you have to really install the package (no develop) and have the appropriate version of PyInstaller (see *requirements.txt*).

Linux/Windows

```
cd bin
pyinstaller -F tagterm
stat dist/tagterm      # details about the built executable
```

And you'll find the built executable under this path: *dist/tagterm*. You can copy it anywhere and use it as a standalone ELF/MZPE executable. Also, you can use the prebuilt ones, available under *bin* directory:

- tagterm (pure Python script)
- tagterm.elf (Linux)
- tagterm.exe (Windows)

2.4 Build Java wrapper

Linux

```
cd src
javac Main.java      # compile
java Main tagterm ../res/error.html  # run a set of examples

# Or just simple:
./main.sh tagterm ../res/error.html
```

Windows

```
cd src
javac Main.java
java Main ../bin/tagterm.exe ../res/error.html

# rem Run with Windows executable:
main.bat ../bin/tagterm.exe ../res/error.html
```

And you should see no exception trace in case everything is fine. Instead of *tagterm* you may use any of the *../bin/dist/tagterm** built executables (or prebuilt ones) suitable for your platform.

Usage

After completing the *installation* procedures, you can use the available or prebuilt **tagterm** executable alone or under the Java wrapper found under the *src* directory. Check *src/Main.java* as an example of handling the *src/tagterm* Java package.

3.1 Python

On Windows, make sure that you have added your paths of interest properly under the *PATH* system environment variable or just use *python tagterm* under *bin* instead of *tagterm* when executing it as Python code. It is recommended to use the prebuilt executable called **tagterm.exe**.

```
# List of available commands.
tagterm --help

# Validate and convert HTML file to XHTML with permissive level 1.
tagterm -v validate -i res/error.html -p
tagterm -v convert -i res/error.html -p
# Check and watch out for nonzero exit codes.
tail tagterm.log

# Now remove tags from the XHTML file.
tagterm -v remove -i res/error-convert.xhtml
cat res/error-convert-remove.xml    # everything ok
```

Note: The *-v* flag stands for *verbose* and you can also use *-o* option for putting the output in a separate path. Run with *-h* for more info, based on the chosen command.

3.2 Java

As you can see in *src/Main.java* example you can simple validate a HTML file by running this (after importing the *tagterm* package):

```
Tagterm tagterm = new Tagterm("tagterm");
tagterm.validate("res/error.html");
```

For conversion, removal and their [s]tring relatives, please consult the [API](#).

3.3 Tags

For editing the tags configuration file, you have to edit the *etc/tagterm/tags* file accordingly, then loading the new settings into your previously setup installed package to make a new build based on new settings.

1. Edit **tags** file.
2. Run *./setup.sh* or *setup.bat* again.
3. *Build* again.
4. Use the newly created *bin/dist/tagterm* executable.

4.1 Python

4.2 Java

class `tagterm.Tagterm` (*String path*)

path - Path to the main Python *tagterm* executable used for all the processes.

convert (*String file*)

Converts a validated HTML file into XHTML. Returns *String* meaning the converted file path.

converts (*String html*)

Same as *convert*, but accepts HTML and returns XHTML content.

remove (*String file*)

Removes all the tags under a XHTML file. Returns *String* meaning the removed-tags file path.

removes (*String html*)

Same as *remove*, but accepts XHTML and returns XML content.

validate (*String file*)

Validates a HTML file. Returns *boolean* as a status of the operation.

validates (*String html*)

Same as *validate*, but accepts HTML content.

Indices and tables

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

C

`convert()` (`tagterm.Tagterm` method), 9
`converts()` (`tagterm.Tagterm` method), 9

R

`remove()` (`tagterm.Tagterm` method), 9
`removes()` (`tagterm.Tagterm` method), 9

T

`tagterm.Tagterm` (built-in class), 9

V

`validate()` (`tagterm.Tagterm` method), 9
`validates()` (`tagterm.Tagterm` method), 9