Tidy Documentation

Release 0.0.1

Xiaoyu Wei

May 23, 2018

Contents:

Tidy is a flexible document manager.

Table of contents

1.1 Getting started with Tidy

Using tidy.

1.1.1 Installation and configuration

Tidy can be installed from PyPI with `pip`: pip install tidy

Tidy follows the XDG Base Directory Specification. The user configuration file goes into `$XDG_CONFIG_HOME/tidy/config.json`, which by default is `~/.config/tidy/config.json`.

1.1.2 Import vs index vs clone

There are three ways to adding in files that serve different needs:

1. *Index*: Files added via index command are processed for fulltext search, and the user can assign logical collection / group / tags to them. However Tidy does not control the actual files added in this manner. Instead all indexed files remain where they were, and tidy only keep symbolic links to them.

2. *Clone*: Besides being indexed, Tidy also make a copy of the files to its data directory. The files can be then deleted without affecting Tidy's functionalities. Cloned directories preserve their structures so that relative paths remain functional.

3. *Import*: All files added via `import` command also get renamed and reorganized for better scalability and performance. FWIW, imported files are stored in a similar fashion to how Git stores its object files.

A rule of thumb is to always use `import` unless you have specific reasons to do otherwise. For example, for a system directory for which you want Tidy to keep its index updated, you might want to use `index`; while if you are adding a Git repository or some IDE's project folder, `clone` might be a good idea.

1.2 Dependencies

Tidy depends on several python packages as described in `setup.py`.

Besides, it depends on:

- `tesseract` for OCR (also you need some language pack for `tesseract` installed). For Arch, install with

```
$ sudo pacman -S tesseract tesseract-data-eus
```

- `imagemagick` for image manipulation, such as metadata generation.

```
$ sudo pacman -S imagemagick
```

1.3 Collections

Collections are databases.

- genindex
- modindex