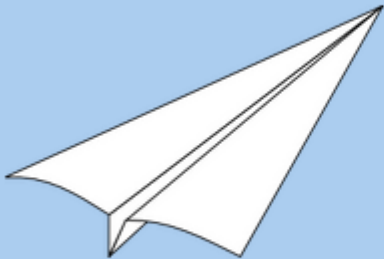


Paperwork

Sorting documents is a machine's job.



With *Paperwork*, scan or import your documents
and find them back in a snap

[Download for GNU/Linux](#)

[Download for Windows](#)

Summary

| | |
|-------------|---------------|
| Open source | Y |
| Language | Python |
| DB | N |
| OCR | Y (tesseract) |

Competitors

- Gnome documents (too simplistic)
- OpenKM (Java based large system, bad docs)
- Mayan-EDMS (too large)
- Alfresco (too large)
- LogicalDOC (use for Medium/Large businesses)
I use this at work, handles millions of documents, lots of features \$\$\$.
- Krystal DMS (Basic features, bit clunky, large business)
- PDFcat (my out of date document management system)

Use cases

- Small office
- Home filing

Pros

- Easy/Quick
- OCR built in
- Can import directories
- Accepts multiple formats (PDF + image files)
- Intelligent tagging (learns as you scan more)
- Open formats
- Easy to backup
- Great documentation
- Multi-OS (Linux/Windows)

Cons

- No web frontend
- No advanced features such as;
 - direct email
 - Rotate/Split PDFs

Installing from repository

- Debian (testing) / Devuan Unstable

```
apt install paperwork-gtk
```

- Arch

```
pacman -S paperwork
```

- Gentoo - **paperwork**

Installing from Source

For Debian Stable/Ubuntu I had to install manually.

Install dependencies

```
sudo apt install python3-pip python3-setuptools  
sudo apt install python3-dev python3-pil
```

PyEnchant dependency

```
sudo apt install libenchant-dev
```

**Not a build dependency, but
recommended anyway**

```
sudo apt install python3-whoosh
```

Install paperwork

- pip3 is the python3 package manager

```
sudo pip3 install paperwork
```

Installing libinsane

To talk to the scanner (required for paperwork to run)

Ref: [libinsane install instructions](#)

- Requires git
- Compile from source
- Go with the default location

Live Demo

- tag
- import ~/papers_{testupload}
- hover over words to show ocr

Extras

Syncthing

- I use syncthing to sync my 'papers' dir PCs

ripgrep-all

- <https://github.com/phiresky/ripgrep-all>
- ripgrep-all allows quick searching through PDFs and other documents
- Command line tool

XournalPP

- `xournalpp` is a handwriting notetaking and pdf annotation program
- Works with wacom pads and touchscreens
- I tested Lenovo Yoga C740 2-1 tested

Libreoffice Online

- [libreoffice-online](#) looks like a full Google Docs replacement
- Haven't tested yet

Foliate

- **foliate** is a ebook reader which looks nice and handles lots of formats
- Haven't tested
- Don't think it does DRM files

Watch Files

- `watch_files` recursively watch directories and apply permissions
- I wrote this to solve a problem at work
- Not related to PDFs, but thought I mention it.
- Doesn't poll the file system uses inotify (file system based subscription).

References

ripgrep-all grep pdfs, odt, etc

xournalpp Handwriting notetaking software

libreoffice-
online GDocs replacement

paperwork personal document management
system

libinsane
install dependency for paperwork

watchfiles Watch files and apply

permissions

bmon

Bandwidth monitor (CLI)

foliate

ebook reader

Questions

Email map7777@gmail.com

Twitter [@map7](https://twitter.com/map7)

Github [github: map7](https://github.com/map7)
