

Paperwork Handbuch

Inhaltsverzeichnis

1	Einleitung	3
2	Definitionen	3
2.1	Arbeitsverzeichnis	3
2.2	Dokument	4
2.3	Seite	4
2.4	Indexierung und Schlüsselwörter	4
2.5	Labels und zusätzliche Schlüsselworte	5
3	Einstellungen	5
3.1	Zugriff auf die Einstellungen	5
3.2	Arbeitsverzeichnis	6
3.3	Scanner	7
3.3.1	Gerät	7
3.3.2	Scanmodus	7
3.3.3	Scan-Auflösung	8
3.3.4	Scanner-Kalibrierung	9
3.4	OCR (optical character recognition, englisch für Texterkennung)	9
3.4.1	Sprachen hinzufügen	10
3.4.2	OCR deaktivieren	10
3.5	Updates	11
4	Neues Dokument	11
5	Scannen	11
6	Import	12
6.1	Bilder	12
6.2	PDF	13
6.3	Viele PDFs auf einen Schlag	13
7	Schlagwörter	13
7.1	Neue Schlagwörter einrichten	13
7.2	Schlagwörter für Dokumente setzen	14
7.3	Ändern der Schlagwortfarbe	14
7.4	Ändern eines Schlagwortnamens	14
7.5	Löschen eines Schlagwortes	14
7.6	Automatische Verschlagwortung	15
8	Suchen	15
8.1	Einfache Suche	15
8.2	Erweiterte Suche	15
9	Anzeigen	16
9.1	Zoomstufe	16
9.2	Seiten im Raster anordnen	16
9.3	Seiten in Listenform anordnen	16
9.4	Alle Wörter hervorheben	17

10 Verschieben von Seiten	17
10.1 Innerhalb eines Dokuments	17
10.2 Zwischen Dokumenten	17
11 Kopieren von Text	17
12 Bearbeiten einer Seite	18
13 Zurücksetzen einer Seite	18
14 Löschen	18
15 Export	18
16 Drucken	18
17 Backup	19
18 Synchronisation zwischen mehreren Computern	19
18.1 USB-Stick / USB-Platte	19
18.2 Anwendungen zur Dateisynchronisation	19
18.2.1 Gemeinsamer Ordner	20
19 Verschlüsselung	20
19.1 GNU/Linux	20
19.1.1 cryptsetup	20
19.1.2 Encfs	20
19.2 Windows	21
20 Tastenkürzel	21
20.1 Dateispeicherorte bei Paperwork	21
20.2 Innere Struktur des Arbeitsverzeichnisses	21
20.2.1 Globale Struktur	22
20.2.2 hOCR Dateien	23
20.2.3 Schlagwort-Dateien	23
21 Unterstützung bekommen	23
22 Probleme melden	24
22.1 Fehlerverfolgung	24
22.2 Automatischer Fehlerbericht	24
22.2.1 ZIP Datei	24
22.2.2 Automatische Übertragung	25
23 Deinstallation	25
23.1 GNU/Linux	25
23.2 Windows 10	25

1 Einleitung



Die meisten privaten Dokumente sind ziemlich wiederholend: Gehaltsabrechnungen, Mietabrechnungen, Stromrechnungen usw. Für die meisten unorganisierten Menschen ist es bestenfalls beunruhigend, diese später wiederfinden zu müssen. Für die meisten organisierten Menschen ist das Benennen und Sortieren so mühsam, wie Farbe beim Trocknen zuzusehen.

Die Grundidee hinter Paperwork ist es, dass die Verwaltung von Dokumenten ein Computerjob ist. Menschen sollten so wenig wie möglich tun, während Maschinen den Großteil der Arbeit erledigen. Das Ziel ist hier "scannen & vergessen".

Wenn Sie eine Software suchen, mit der Sie jedes Dokument einzeln benennen, in einer komplexen Hierarchie organisieren, jedes Mal manuell markieren, kleinere OCR-Probleme beheben usw. können, dann ist Paperwork nichts für Sie.

2 Definitionen

2.1 Arbeitsverzeichnis

Paperwork speichert alle Ihre Dokumente in einem einzigen Verzeichnis: dem Arbeitsverzeichnis. In diesem Verzeichnis hat jedes Dokument sein eigenes Unterverzeichnis.

Dies erschwert zwar die Verwendung von Paperwork mit anderen Programmen, hat aber einen großen Vorteil: Sie müssen sich nicht mehr um Dateinamen und Verzeichnisstrukturen kümmern.

2.2 Dokument



In Paperwork ist ein Dokument eine Ansammlung von Seiten. Auf der Festplatte kann dies entweder eine Ansammlung von JPEG-Dateien oder eine PDF-Datei sein.

Dokumente werden nur durch ein Datum identifiziert. Es kann entweder das Datum sein, an dem Sie sie importiert haben (voreingestellter Standard), oder ein von Ihnen gewähltes Datum.

Sie werden auf der linken Seite des Hauptfensters angezeigt (grüner Teil auf dem Screenshot oben).

2.3 Seite

In Paperwork ist eine Seite nur ein Bild und Wortpositionen auf diesem Bild.

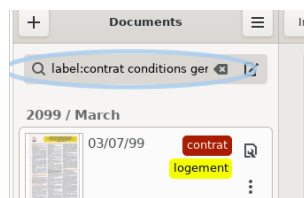
Bilder können von einem Scanner kommen oder importiert werden. In diesen Fällen wird es als JPEG-Datei gespeichert und der Text wird mit OCR (Optical Character Recognition, englisch für Zeichnerkennung) extrahiert. OCR ist ein ziemlich langwieriger Prozess. Er kann bis zu ein paar Minuten pro Seite dauern. Daher wird der aus den Bildern extrahierte Text in hOCR-Dateien neben den JPEG-Dateien gespeichert.

Seiten können auch die Seiten aus einer PDF-Datei sein. In diesem Fall speichert Paperwork standardmäßig nur eine Kopie der PDF-Datei.

Paperwork kümmert es nicht, ob eine Seite Vorder- oder Rückseite ist. Paperwork berücksichtigt auch nicht das Papierformat einer Seite (A4, Letter, usw.).

Die Seiten werden auf der rechten Seite des Hauptfensters angezeigt (blauer Teil auf dem Screenshot oben).

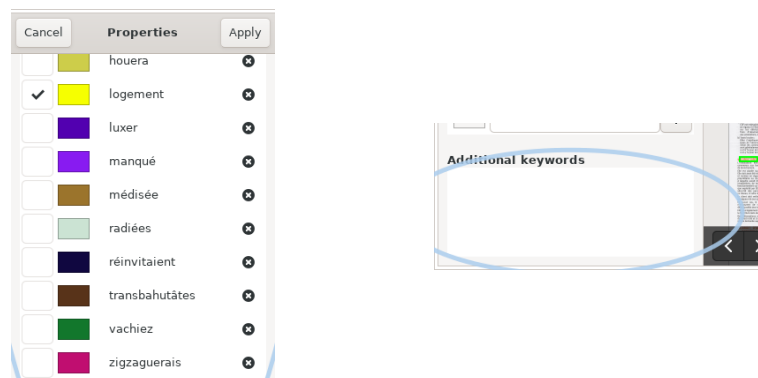
2.4 Indexierung und Schlüsselwörter



Natürlich brauchen Sie eine Möglichkeit, Ihre Dokumente wiederzufinden. Paperwork verwaltet einen Index mit allen Schlagwörtern, die in Ihren Dokumenten gefunden wurden.

Geben Sie einfach ein paar Schlüsselwörter ein, und Sie erhalten Ihre Dokumente zurück.

2.5 Labels und zusätzliche Schlüsselwörter



Leider enthalten Dokumente manchmal nicht die erforderlichen Schlüsselwörter, um sie wiederzufinden. Außerdem ist OCR kein perfekt zuverlässiger Prozess und funktioniert möglicherweise nicht.

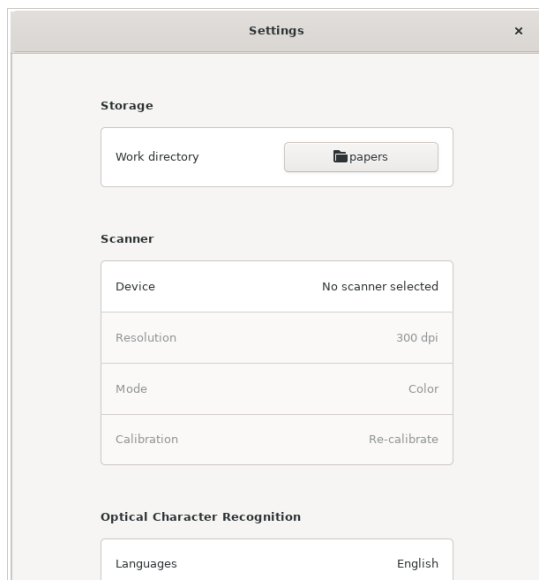
Um diese Probleme abzumildern, können Sie Ihre Dokumente mit Etiketten (oder Tags) versehen und zusätzliche Schlüsselwörter angeben. Beides wird in den Index aufgenommen.

Etiketten werden neben den Dokumenten angezeigt. Zusätzliche Stichwörter werden fast nie angezeigt.

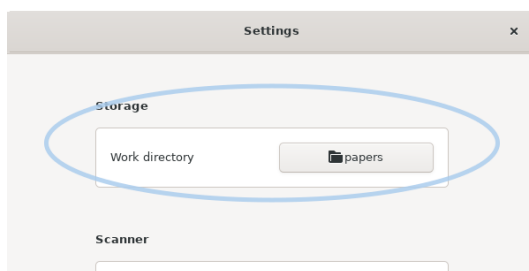
3 Einstellungen

3.1 Zugriff auf die Einstellungen





3.2 Arbeitsverzeichnis

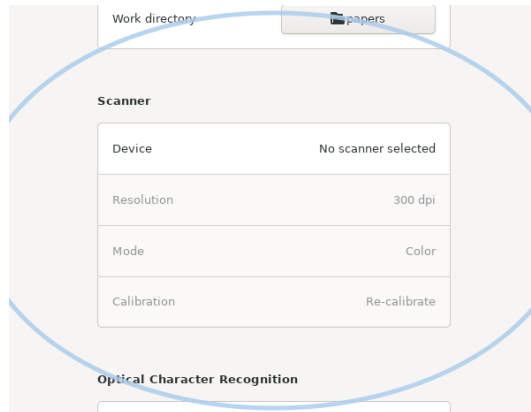


Das Arbeitsverzeichnis ist das Verzeichnis, in dem Sie alle Ihre Dokumente speichern möchten. Das kann ein normaler Ordner, ein über mehrere Computer synchronisierter Ordner oder eine Netzwerkfreigabe sein.

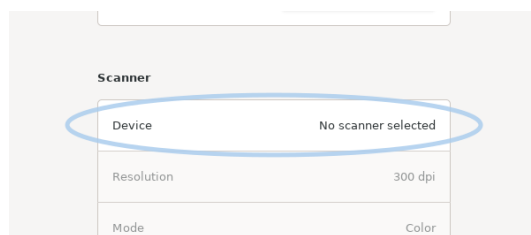
Nachdem Sie den Einstellungsdialog geschlossen haben, wird das Arbeitsverzeichnis gescannt und der Paperwork-Index wird entsprechend seinem Index aktualisiert.

Jedes Mal, wenn Paperwork startet, sucht es nach Änderungen in diesem Ordner und synchronisiert seinen Index entsprechend.

3.3 Scanner



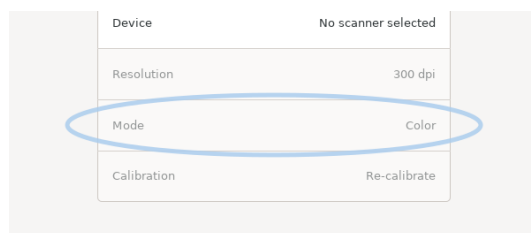
3.3.1 Gerät



Beim Start sucht Paperwork nach Scannern. Der zu verwendende Scanner kann in den Einstellungen ausgewählt werden.

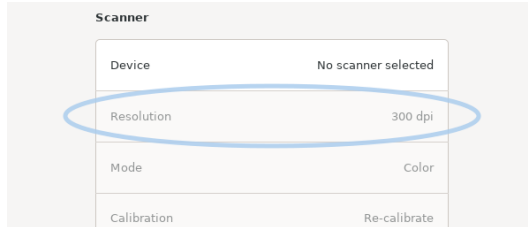
Webcams, Dateispeicher usw. können nicht verwendet werden. Nur Papierfresser.

3.3.2 Scanmodus



Die meisten modernen Scanner scannen in Farbe in einer vertretbaren Zeit. Einige ältere Scanner scannen jedoch viel schneller in Graustufen oder sogar in Schwarz-Weiß. Hier können Sie den zu verwendenden Modus auswählen.

3.3.3 Scan-Auflösung



Die Scan-Auflösung legt fest, wie detailliert die von Ihrem Scanner kommenden Bilder sein müssen.

Höhere Auflösungen bedeuten

- längere Scans,
- längere OCR,
- mehr Zeit für die Anzeige,
- mehr Speicherplatz auf der Festplatte verwendet,
- aber auch eine bessere OCR.

Niedrigere Auflösungen bedeuten

- kürzere Scans,
- kürzere OCR,
- weniger Zeit für die Anzeige,
- weniger Speicherplatz auf der Festplatte,
- aber auch eine minderwertige OCR,
- und ein möglicherweise unlesbares (auch von einem Menschen) Bild.

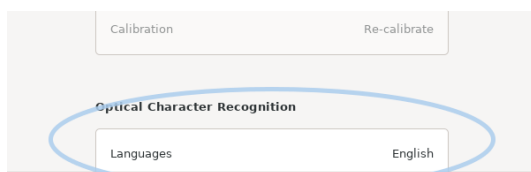
300 dpi wird als guter Kompromiss angesehen. Auf langsamen Computern sollten Sie eventuell auf 200 dpi reduzieren.

3.3.4 Scanner-Kalibrierung



Scanner neigen dazu, Bilder zu liefern, die größer sind als die gescannten Seiten. Da Sie in den meisten Fällen immer Seiten mit der gleichen Größe (normalerweise A4 oder Letter) scannen werden, bietet Paperwork eine Option namens Scannerkalibrierung. Die Scannerkalibrierung in Paperwork ist einfach ein Bereich, der aus den die vom Scanner kommenden Bildern jedesmal ausgeschnitten wird.

3.4 OCR (optical character recognition, englisch für Texterkennung)



Paperwork verwendet standardmäßig Tesseract für die OCR. Wenn das nicht verfügbar ist, greift es auf Cuneiform zurück.

Unter Linux wird Paperwork, wenn es mit Flatpak installiert wurde, immer mit Tesseract ausgeliefert. Unter Windows wird Paperwork immer mit Tesseract installiert.

Um bessere Ergebnisse zu erzielen, muss das OCR-Programm die in den Dokumenten verwendete Sprache kennen.

Die im Einstellungsdialog von Paperwork verfügbaren Sprachen sind diejenigen, die vom OCR-Programm beherrscht werden. Wenn Ihre Sprache nicht in der Liste enthalten ist, bedeutet dies, dass das OCR-Programm nicht über die erforderlichen Sprachdaten verfügt und, dass Sie diese installieren müssen.

3.4.1 Sprachen hinzufügen

Flatpak

```
# <langs> ist eine Liste von zweibuchstabigen Ländercodes getrennt durch ';'
# Beispiel: en;fr;de
flatpak config --user --set languages "<langs>"
flatpak update --user
```

Debian

```
# <lang> ist ein dreibuchstabiger Sprach-Code
# Beispiel: 'fra' für Französisch
$ sudo apt-get install tesseract-ocr tesseract-ocr-<lang>
```

Fedora

```
# <lang> ist ein dreibuchstabiger Sprach-Code
# Beispiel: 'fra' für Französisch
$ sudo dnf install tesseract tesseract-langpack-<lang>
```

Ubuntu

```
# <lang> ist ein dreibuchstabiger Sprach-Code
# Beispiel: 'fra' für Französisch
$ sudo apt-get install tesseract-ocr tesseract-ocr-<lang>
```

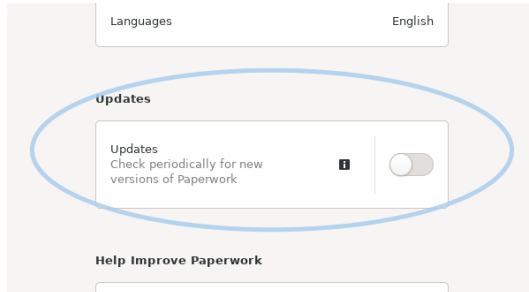
Windows Tesseract und alle seine Datendateien werden vom Installationsprogramm von Paperwork bereitgestellt. Sie können das Installationsprogramm erneut ausführen, um andere Sprachen zu installieren.

Wenn eine Sprache im Installationsprogramm nicht verfügbar ist, bedeutet dies entweder, dass sie nicht mit eingepackt wurde (in diesem Fall können Sie sie anfordern), oder es ist noch keine Datendatei für diese Sprache verfügbar.

3.4.2 OCR deaktivieren

Wenn Sie eine Seite mit Paperwork scannen, führt Paperwork sofort die OCR auf der Seite durch. Dieser Vorgang kann für jede Seite eine Weile dauern. Für den Fall, dass Sie viele Seiten schnell scannen möchten (z. B. bei der ersten Verwendung von Paperwork), kann die OCR vorübergehend deaktiviert werden. Um OCR zu deaktivieren, können Sie einfach alle OCR-Sprachen abwählen.

3.5 Updates



Wenn Sie diese Option aktivieren, sucht Paperwork beim Start nach Updates, falls es dies eine Woche oder länger nicht getan hat. Um zu prüfen, ob eine neue Version verfügbar ist, wird eine HTTPS-Anfrage an 'openpaper.work' gesendet.

Sie werden benachrichtigt, wenn ein Update gefunden wurde, aber es wird nicht installiert.

4 Neues Dokument

Standardmäßig enthält Paperwork in der Dokumentenliste ein Dokument namens "Neues Dokument". Wenn Sie es öffnen, erscheint es immer leer. Dieses Dokument existiert tatsächlich noch nicht auf der Festplatte, aber es wird existieren, sobald Sie eine Seite hineinlegen. Sie können ihm Seiten hinzufügen, indem Sie scannen, Dateien importieren oder eine Seite aus einer anderen Datei darin ablegen.

Sobald Sie einen Inhalt hineinlegen, bekommt dieses Dokument ein eigenes Datum (standardmäßig das aktuelle). In der Dokumentenliste wird "Neues Dokument" durch dieses Datum ersetzt, und ein neues "Neues Dokument" wird der Dokumentenliste hinzugefügt.



Wenn Sie gerade etwas suchen (siehe Kapitel "Suchen"), werden nur die Suchergebnisse angezeigt und daher wird dieses "Neue Dokument" nicht angezeigt. Sie können es zurückholen, indem Sie auf die Schaltfläche "+" in der linken oberen Ecke des Hauptfensters klicken.

5 Scannen



Wenn ein Scanner in den Einstellungen ausgewählt wurde, können Sie diesen zum Scannen von Seiten verwenden.

In der Kopfleiste befindet sich eine Schaltfläche zum Hinzufügen von Seiten. Der kleine Pfeil auf der rechten Seite ermöglicht den Zugriff auf mögliche Seitenquellen. Zu diesen Seitenquellen gehören Ihre Scannerquellen (Flachbett, Einzug).

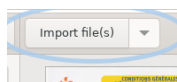
Nachdem Sie die gewünschte Scannerquelle ausgewählt haben, können Sie auf die Schaltfläche "Scannen von ..." klicken.

Dadurch wird eine Scan-Sitzung gestartet:

- Gescannte Seiten werden an das Ende des aktuellen Dokuments angehängt. Falls Sie einen automatischen Papiereinzug verwenden, scannt Paperwork Seiten, bis der Einzug leer ist.
- Die Papiere werden dann entsprechend der Scannerkalibrierung beschnitten.
- Paperwork lässt die OCR laufen
- Paperwork wird sie indizieren

Wenn in dieser Scan-Sitzung ein neues Dokument erstellt wird, versucht Paperwork, automatisch Schlagwörter im Dokument zu setzen.

6 Import



6.1 Bilder

Paperwork unterstützt eine Vielzahl von Dateiformaten. Es unterstützt JPEG, PNG, GIF, BMP, TIFF, usw.

Jede Bilddatei wird als eine Seite behandelt.

Bilder werden immer an das aktuell geöffnete Dokument angehängt. Wählen Sie einfach ein leeres Dokument ("Neues Dokument"), um beim Importieren ein neues Dokument zu erstellen.

Auf importierten Bildern wird immer eine OCR ausgeführt. Falls das importierte Bild die erste Seite eines neuen Dokuments ist, erzeugt Paperwork automatisch Schlagwörter.

Beachten Sie, dass Paperwork ein Dokumentenmanager ist. Es kann zwar Bilder mit nur sehr wenig Text oder Fotos verarbeiten, ist aber nicht dafür ausgelegt. Die automatische Verschlagwortung wird bei solchen Dokumenten nicht korrekt funktionieren.

Die OCR (Tesseract) arbeitet sehr gut mit schwarzem Text auf weißem Hintergrund. Die automatische Verschlagwortung verwendet den erkannten Text und benötigt viele Schlüsselwörter auf der ersten Seite.

6.2 PDF

PDF-Dateien werden immer als ein ganzes Dokument betrachtet. Sie werden nie an ein bestehendes Dokument angehängt. Sie werden in das Arbeitsverzeichnis kopiert und umbenannt, aber ihr Inhalt wird nicht verändert. Paperwork behält die ursprüngliche PDF-Datei immer so bei, wie sie war, auch wenn Sie einige ihrer Seiten bearbeiten: Die bearbeiteten Seiten werden neben der PDF-Datei gespeichert.

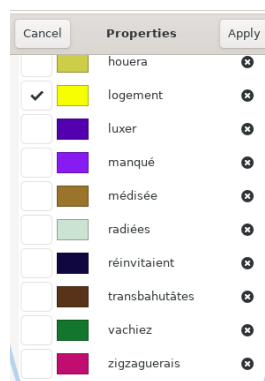
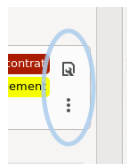
Paperwork sucht nach den Seiten, die keinen Text beigefügt haben. Auf diesen Seiten wird automatisch eine OCR durchgeführt. Nachdem alle Seiten untersucht wurden, wird automatisch verschlagwortet. Beachten Sie, dass dieser Vorgang bei großen PDF-Dateien einige Minuten dauern kann.

Wenn eine PDF-Datei bereits Teil Ihrer Dokumente ist, wird sie von Paperwork einfach ignoriert.

6.3 Viele PDFs auf einen Schlag

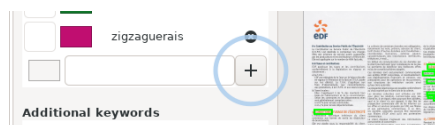
Wenn Sie beim Importieren einen Ordner auswählen, durchsucht Paperwork diesen Ordner und sucht nach zu importierenden PDF-Dateien. Bereits importierte PDF-Dateien werden einfach ignoriert. Der Ordner wird rekursiv durchsucht (alle Ordner innerhalb des Ordners werden ebenfalls abgearbeitet).

7 Schlagwörter



Derzeit gibt es eine Einschränkung in Paperwork: Jedes Schlagwort muss sich in mindestens einem Dokument finden. Andernfalls werden beim Neustart von Paperwork Schlagwörter ohne Dokumentenbezug verschwinden.

7.1 Neue Schlagwörter einrichten



Sie können auf das graue Rechteck auf der linken Seite klicken, um die Schlagwortfarbe auszuwählen. Sie können den Schlagwortnamen in das Textfeld zwischen dem grauen Rechteck und der Schaltfläche "+" eingeben.

Sobald Sie auf die Schaltfläche "+" klicken, wird das Schlagwort zum aktuellen Dokument hinzugefügt.

Das Schlagwort wird endgültig hinzugefügt, sobald Sie die Dokumenteigenschaften schließen. Paperwork wird dann seinen Index entsprechend aktualisieren.

7.2 Schlagwörter für Dokumente setzen

Wenn Sie die Dokumenteigenschaften öffnen, wird die Schlagwortliste angezeigt. Auf der linken Seite jeder Schlagwortfarbe finden Sie eine Schaltfläche. Mit dieser Schaltfläche können Sie Schlagwörter zum aktuellen Dokument hinzufügen oder davon entfernen.

Die Änderungen werden endgültig auf die Festplatte geschrieben, wenn Sie die Dokumenteigenschaften schließen. Paperwork wird dann seinen Index entsprechend aktualisieren.

7.3 Ändern der Schlagwortfarbe

Wenn Sie die Dokumenteigenschaften öffnen, können Sie auf eine Schlagwortfarbe klicken, um sie zu ändern. In einem Dialogfeld können Sie die neue Farbe auswählen.

Die Schlagwortfarbe wird endgültig auf der Festplatte geändert, wenn Sie die Dokumenteigenschaften schließen. Paperwork aktualisiert dann das Schlagwort für alle Dokumente, die dieses verwenden.

7.4 Ändern eines Schlagwortnamens

Wenn Sie die Dokumenteigenschaften öffnen, können Sie auf einen Schlagwortnamen klicken, um diesen zu ändern. In einem Dialogfeld können Sie den neuen Namen eingeben.

Der Schlagwortname wird endgültig auf der Festplatte geändert, wenn Sie die Dokumenteigenschaften schließen. Paperwork aktualisiert dann das Schlagwort für allen Dokumente, die dieses verwenden, und indiziert sie dann alle neu.

7.5 Löschen eines Schlagwortes

Rechts von jedem Schlagwort befindet sich eine weiß-schwarze Kreuztaste. Wenn Sie darauf klicken, können Sie das Schlagwort löschen.

Nachdem Sie die Dokumenteigenschaften geschlossen haben, wird das Schlagwort für alle Dokumente entfernt, die es enthalten hatten. Paperwork wird dann seinen Index entsprechend aktualisieren.

Achtung! Wenn Sie die Dokumenteigenschaften geschlossen haben, gibt es keine Möglichkeit, das gelöschte Schlagwort wieder einzufügen.

7.6 Automatische Verschlagwortung

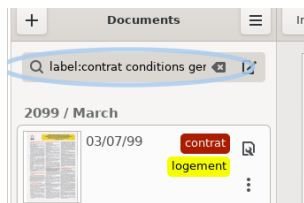
Paperwork nutzt künstliche Intelligenz. Es verwendet sogar eine ziemlich einfache Methode: Naive Bayes classifiers. Dies ist die gleiche Technologie, die von E-Mail-Clients verwendet wird, um Mails als Spam/Nicht-Spam zu klassifizieren.

Basierend auf allen Schlüsselwörtern in all Ihren Dokumenten, die ein Schlagwort haben (oder nicht), kann es eine Wahrscheinlichkeit abschätzen, dass ein Dokument, das dieselben Schlüsselwörter enthält, dasselbe Schlagwort haben oder nicht haben müsste. Wenn die Wahrscheinlichkeit hoch genug ist, wird das Dokument beim Importieren oder Scannen automagisch mit dem Schlagwort versehen.

Natürlich bedeutet dieser Ansatz, dass Paperwork genügend Muster benötigt, um zuverlässig zu arbeiten. Sie können davon ausgehen, dass dies ab etwa 100 Dokumenten funktioniert (und nur für Schlagwörter, die für mehr als 10 Dokumente oder mehr gesetzt sind).

8 Suchen

8.1 Einfache Suche

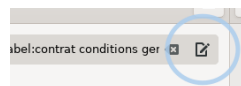


Geben Sie einfach Stichwörter in das Suchfeld ein. In wenigen Sekunden erhalten Sie alle Dokumente, die diese Stichwörter enthalten.

Paperwork führt eine "unscharfe" Suche durch: Es werden auch Dokumente mit Stichwörtern zurückgegeben, die dem von Ihnen angegebenen ähnlich sind, aber nicht identisch sind (z. B. "flech" statt "flesch").

Sie können auch die Whoosh-Abfragesprache verwenden, um komplexere Abfragen zu erstellen. Wenn Sie Beispiele brauchen, können Sie den unten beschriebenen Dialog für die erweiterte Suche verwenden.

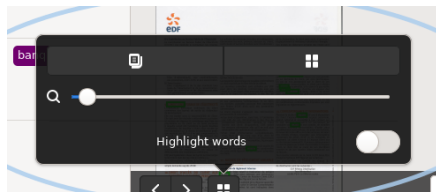
8.2 Erweiterte Suche



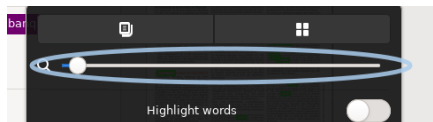
		Label	contrat	
and		Keyword(s)	conditions	Remove
and		Keyword(s)	generales	Remove
<button>Add</button>				
				<button>Cancel</button> <button>Apply</button>

Der erweiterte Suchdialog hilft bei der Erstellung komplexer Suchanfragen. Sie können verschiedene Kriterien angeben und sobald Sie auf die Schaltfläche "Anwenden" klicken, wird eine Suchanfrage für Sie generiert und sofort in das Suchfeld eingefügt. Die Suchergebnisse werden ebenfalls sofort aktualisiert.

9 Anzeigen



9.1 Zoomstufe



Mit diesem Steuerelement können Sie die Vergrößerung ändern, mit der die Seiten angezeigt werden.

9.2 Seiten im Raster anordnen



Wenn Sie auf diese Schaltfläche klicken, versucht Paperwork, die Seiten in 3 Spalten anzuzeigen. In diesem Modus können Sie Seiten per Drag'n'Drop innerhalb des Dokuments oder in ein anderes Dokument verschieben.

9.3 Seiten in Listenform anordnen



Wenn Sie auf diese Schaltfläche klicken, werden die Seiten so skaliert, dass ihre Breite der maximal zulässigen Breite des Hauptfensters entspricht. In diesem Modus können Sie Text auf der Seite auswählen (und dann kopieren).

9.4 Alle Wörter hervorheben



Mit dieser Option können Sie schnell alle von der OCR erkannten Wörter sehen. Manchmal (selten) verpasst OCR ganze Abschnitte auf einer Seite. Mit dieser Option können Sie solche Abschnitte schnell erkennen.

10 Verschieben von Seiten

10.1 Innerhalb eines Dokuments

Sie müssen die Dokumentseiten als Raster anordnen (siehe 9.2).

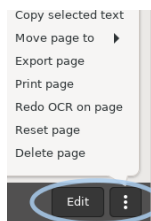
Sie können dann eine Seite greifen (halten Sie die linke Maustaste gedrückt), ziehen und an einer beliebigen Stelle im Dokument ablegen. Während des Ziehens zeigt Ihnen eine blaue Markierung an, wo die Seite landen würde, wenn Sie die linke Maustaste loslassen.

10.2 Zwischen Dokumenten

Sie müssen die Dokumentseiten als Raster anordnen (siehe 9.2).

Sie können dann eine Seite greifen (halten Sie die linke Maustaste gedrückt), ziehen und in der Dokumentenliste auf dem Dokument ablegen, in dem die Seite erscheinen soll.

11 Kopieren von Text



Sie müssen die Dokumentseiten als Liste anordnen (siehe 9.3).

Sie können dann Text auf einer Seite auswählen. Halten Sie die linke Maustaste gedrückt, um die Auswahl zu starten, bewegen Sie den Mauszeiger, um weitere Wörter auszuwählen, und lassen Sie dann die Taste los.

Sie können dann den ausgewählten Text kopieren, entweder durch Drücken von Strg-C oder über das Seitenmenü unten rechts im Hauptfenster. Nach dem Kopieren können Sie den ausgewählten Text in jeder anderen Anwendung einfügen (Strg-V).

12 Bearbeiten einer Seite



Paperwork enthält einen sehr einfachen Bildeditor. Er bietet 4 Funktionen:

- Schneiden
- Drehen der Seite um 90° (es darf mehrfach gedreht werden)
- Drehen der Seite um -90° (es darf mehrfach gedreht werden)
- Automatische Farbabgleich: Ein Algorithmus passt die Bildhelligkeit, den Kontrast und die Farben an, um das Bild so gut wie möglich lesbar zu machen.

13 Zurücksetzen einer Seite

Das Zurücksetzen einer Seite setzt diese in den Ausgangszustand zurück, in dem sie gescannt oder importiert wurde, bevor irgendeine Vorverarbeitung stattfand.

Dies kann hilfreich sein, wenn Sie eine falsche Änderung an der Seite vorgenommen haben (z. B. einen falschen Bereich beschnitten haben), wenn die Kalibrierungseinstellungen nicht passed gewählt waren oder falls Vorverarbeitungsalgorithmen die Seite verkorkst haben.

14 Löschen

Wenn Sie ein Dokument oder eine Seite löschen, werden diese in den Papierkorb Ihres Computers verschoben.

Wichtiger Hinweis zu Flatpak: Ein Fehler kann Paperwork daran hindern, Dateien in den Papierkorb zu verschieben (wir arbeiten daran). In diesem Fall wird Paperwork die Datei direkt löschen (dann ist keine Wiederherstellung möglich).

15 Export

Sie können sowohl Dokumente als auch einzelne Seiten exportieren.

In beiden Fällen können vor dem eigentlichen Exportieren verschiedene Transformationen vorgenommen werden. Zum Beispiel können Sie Farbseiten in Graustufen umwandeln, bevor Sie sie in eine neue PDF-Datei einfügen (wodurch die resultierende PDF-Datei kleiner wird).

16 Drucken

Sie können Dokumente und auch einzelne Seiten drucken.

Beachten Sie, dass die Seiten immer als Bilder an Ihren Drucker gesendet werden. Bei sehr großen Dokumenten können also einige Minuten vergehen, bevor der eigentliche Druckvorgang beginnt.

17 Backup

18 Synchronisation zwischen mehreren Computern

Paperwork ist zwar ein persönlicher Dokumentenmanager, aber keine Anwendung zur Dateisynchronisation. Es gibt Anwendungen, die sich der Dateisynchronisation widmen und das bereits sehr gut tun. Daher ist Paperwork für die Verwendung mit solchen Anwendungen (Nextcloud, Dropbox, OneDrive, SparkleShare, etc.) konzipiert.

Nach dem Start prüft Paperwork als erstes den Inhalt des Arbeitsverzeichnisses. Es sucht nach Änderungen und aktualisiert seine Dokumentenliste und auch den Index automatisch. Wenn also eine andere Instanz von Paperwork auf einem anderen Computer etwas im Arbeitsverzeichnis geändert hat und diese Änderung auf einem anderen Computer synchronisiert wurde, wird das andere Paperwork diese Änderung beim Start automatisch übernehmen.

18.1 USB-Stick / USB-Platte

Dies ist der einfachste Weg, um Dokumente gemeinsam zu nutzen. Kopieren Sie einfach Ihr Arbeitsverzeichnis auf einen USB-Stick, weisen Sie Paperwork an, es zu verwenden, und schon sind Sie fertig.

Achtung: Sie sollten Ihren USB-Stick von Zeit zu Zeit auf einem anderen sichern.

18.2 Anwendungen zur Dateisynchronisation

Solche Anwendungen synchronisieren ein lokales Verzeichnis mit einem Remote-Server (oder einer Cloud). Alle Änderungen, die Sie in Ihrem Ordner vornehmen, werden auch auf dem Server vorgenommen. Alle Änderungen, die auf dem Server angewandt wurden, werden auch auf allen verbundenen Computern angewendet. Der Server kann Ihnen oder einer anderen Person (normalerweise einer Firma) gehören.

Achtung: Wenn Sie sich dafür entscheiden, Ihre Dokumente auf einem fremden Server zu hosten (DropBox, OneDrive, usw.), kann dieser auf alle Ihre Dokumente zugreifen. Paperwork verschlüsselt diese nicht selber.

Der Paperwork wird täglich mit Nextcloud getestet. Das ist zwar nicht die einfachste Variante, aber mit Nextcloud können Sie Ihre Dateien selbst hosten. Es gibt noch andere selbst gehostete Alternativen, die existieren: SparkleShare, Syncthing, usw.

Die Verwendung von DropBox oder OneDrive könnte dann sinnvoll sein, wenn Sie nicht ganz so vertrauliche Dokumente mit anderen teilen (Vereine usw.).

18.2.1 Gemeinsamer Ordner

Wenn sich alle Ihre Computer im selben Netzwerk befinden, können Sie Ihr Arbeitsverzeichnis gemeinsam nutzen. Seien Sie jedoch sehr vorsichtig, was die Zugriffsrechte angeht. Wenn Sie zu freizügig sind, könnte ein böser Angreifer Zugriff auf alle Ihre persönlichen Dokumente erhalten! Und diese richtig einzustellen ist knifflig.

Achtung! Die Verwendung eines gemeinsamen Ordners bedeutet, dass Sie nur eine einzige Kopie Ihres Arbeitsverzeichnisses haben. Sie sollten regelmäßig Backups Ihres Arbeitsverzeichnisses erstellen.

19 Verschlüsselung

19.1 GNU/Linux

GNU/Linux-Distributionen enthalten viele Programme zum Verschlüsseln ganzer Verzeichnisse.

Für Paperwork sollten 2 Verzeichnisse zum Schutz Ihrer Privatsphäre verschlüsselt werden:

- Ihr Arbeitsverzeichnis (standardmäßig ist das `~/papers`, dies kann aber in den Einstellungen geändert werden)
- Das Cache-Verzeichnis (`~/local/share/paperwork2`, dies kann nicht geändert werden) enthält Indexdateien, aus denen der Inhalt Ihrer Dokumente teilweise wiederhergestellt werden kann

Beachten Sie: Wenn Sie sicher sein wollen, dass Ihre Daten immer verschlüsselt sind, empfiehlt es sich, wenn möglich Ihr gesamtes Home-Verzeichnis oder sogar Ihr gesamtes System zu verschlüsseln.

19.1.1 cryptsetup

Die meisten GNU/Linux-Distributionen bieten mittlerweile eine Option an, um Ihr gesamtes System oder Ihr gesamtes `/home` mit `cryptsetup` zu verschlüsseln. Dies ist die empfohlene Methode, um Ihre Dokumente zu schützen.

19.1.2 Encfs

Zur einfachen Erstellung verschlüsselter Verzeichnisse kann auch `Encfs` verwendet werden.

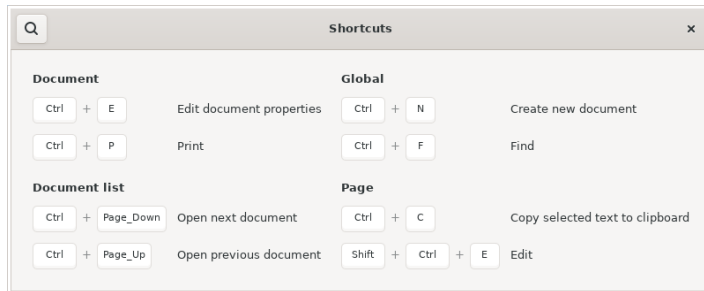
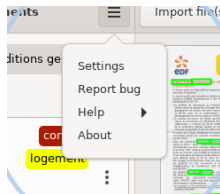
Beachten Sie, dass `Encfs` einige Sicherheitslücken zu haben scheint. Während es also wahrscheinlich ausreicht, um eine Laptop-Diebin daran zu hindern, auf Ihre Dokumente zuzugreifen, wird es wahrscheinlich nicht ausreichen, um die NSA oder die Polizei daran zu hindern ;-).

```
$ encfs ~/.local/share/.paperwork2 ~/.local/share/paperwork2
$ encfs ~/.papers ~/papers
```

19.2 Windows

Unter Windows wird zum Schutz Ihrer Dokumente dringend empfohlen, BitLocker zu aktivieren. Falls nicht verfügbar, gibt es andere Anwendungen (Veracrypt, usw.).

20 Tastenkürzel



Die Tastenkürzel können Sie anzeigen, indem Sie das Anwendungsmenü öffnen, "Hilfe" und dann "Tastenkürzel" wählen.

20.1 Dateispeicherorte bei Paperwork

Voreingestellt:

- Einstellung: `~/.config/paperwork2.conf`
- Index: `~/.local/share/paperwork2`
- Dokumente: `~/papers`

(unter Windows werden dieselben Pfade verwendet; `~` = `C:\Users[login]` ; Ordner sind ausgeblendet)

Der Index wird auf der Grundlage der Dokumente im Arbeitsverzeichnis ständig aktualisiert. Beim Start von Paperwork wird die Änderungszeit jeder Datei genutzt, um Änderungen an den Dokumenten zu erkennen.

20.2 Innere Struktur des Arbeitsverzeichnisses

`workdir|rootdir` = `~/papers` (by default)

20.2.1 Globale Struktur

Im Arbeitsverzeichnis gibt es pro Dokument einen separaten Ordner.

Die Ordnernamen sind (normalerweise) das Scan-/Importdatum des Dokuments: YYYYMMDD_hhmm_ss[_idx]. Das Suffix 'idx' ist optional und ist einfach eine Zahl, die im Falle einer Namenskollision hinzugefügt wird.

In jedem Ordner haben wir:

- Für Bilddokumente:
 - paper.<X>.jpg: Die Originalseite im JPG-Format (X beginnt bei 1)
 - paper.<X>.edited.jpg (optional): Die vom Benutzer bearbeitete Seite (X beginnt bei 1)
 - paper.<X>.words (optional): Eine hOCR-Datei, die alle per OCR auf der Seite gefunden Wörter enthält (optional, aber für die Indizierung erforderlich; kann mit der Option "Wiederhole OCR" neu erzeugt werden).
 - paper.1.thumb.jpg (optional, wird automatisch generiert): Eine Miniaturversion der Seite (lädt schneller)
 - labels (optional): eine Textdatei, die alle für dieses Dokument vergebenen Schlagwörter enthält
 - extra.txt (optional): vom Benutzer hinzugefügte zusätzliche Schlüsselwörter
- Für PDF-Dokumente:
 - doc.pdf: das Dokument
 - labels (optional): eine Textdatei, die alle für dieses Dokument vergebenen Schlagwörter enthält
 - paper.<X>.edited.jpg (optional): Die vom Benutzer bearbeitete Seite (X beginnt bei 1)
 - extra.txt (optional): vom Benutzer hinzugefügte zusätzliche Schlüsselwörter
 - paper.<X>.words (optional): Eine hOCR-Datei, die alle durch die OCR auf der Seite gefunden Wörter enthält. Einige PDFs enthalten Mist anstelle des echten Textes, so dass es manchmal nützlich sein kann, die OCR über sie laufen zu lassen.
 - passwd.txt (optional): PDF Passwort, falls das PDF passwortgeschützt ist.
 - doc.docx / doc.odt / ... (optional): Originaldatei. Konvertiert nach PDF (doc.pdf) damit Paperwork sie schneller verarbeiten und anzeigen kann.

Im folgenden ein Beispiel für eine Arbeitsverzeichnisorganisation:

```
$ find ~/papers
/home/jflesch/papers
/home/jflesch/papers/20130505_1518_00
/home/jflesch/papers/20130505_1518_00/paper.1.jpg
/home/jflesch/papers/20130505_1518_00/paper.1.thumb.jpg
/home/jflesch/papers/20130505_1518_00/paper.1.words
```

```

/home/jfflesch/papers/20130505_1518_00/paper.2.jpg
/home/jfflesch/papers/20130505_1518_00/paper.2.edited.jpg
/home/jfflesch/papers/20130505_1518_00/paper.2.words
/home/jfflesch/papers/20130505_1518_00/paper.3.jpg
/home/jfflesch/papers/20130505_1518_00/paper.3.words
/home/jfflesch/papers/20130505_1518_00/labels
/home/jfflesch/papers/20110726_0000_01f
/home/jfflesch/papers/20110726_0000_01/paper.1.jpg
/home/jfflesch/papers/20110726_0000_01/paper.1.thumb.jpg
/home/jfflesch/papers/20110726_0000_01/paper.1.words
/home/jfflesch/papers/20110726_0000_01/paper.2.jpg
/home/jfflesch/papers/20110726_0000_01/paper.2.words
/home/jfflesch/papers/20110726_0000_01/extra.txt
/home/jfflesch/papers/20130106_1309_44
/home/jfflesch/papers/20130106_1309_44/doc.pdf
/home/jfflesch/papers/20130106_1309_44/paper.1.thumb.jpg
/home/jfflesch/papers/20130106_1309_44/paper.2.edited.jpg
/home/jfflesch/papers/20130106_1309_44/paper.2.words
/home/jfflesch/papers/20130106_1309_44/labels
/home/jfflesch/papers/20130106_1309_44/extra.txt
/home/jfflesch/papers/20130106_1309_44/passwd.txt
/home/jfflesch/papers/20130520_1309_44
/home/jfflesch/papers/20130520_1309_44/doc.pdf
/home/jfflesch/papers/20130520_1309_44/doc.docx
/home/jfflesch/papers/20130520_1309_44/labels

```

20.2.2 hOCR Dateien

Mit Tesseract kann die hOCR-Datei mit folgendem Befehl erzeugt werden:

```
tesseract paper.<X>.jpg paper.<X> -l <lang> hocr && mv paper.<X>.html paper.<X>.words
```

Beispiel:

```
tesseract paper.1.jpg paper.1 -l fra hocr && mv paper.1.html paper.1.words
```

20.2.3 Schlagwort-Dateien

Beispiel für den Inhalt einer Schlagwort-Datei:

```
Rechnung,#0000b1588c61 Wohnung,#f6b6ffff0000
```

Es ist immer [label],[color]. Bei gleichem Schlagwort sollte die Farbe immer gleich sein.

21 Unterstützung bekommen

Es gibt ein Forum für Paperwork: <https://forum.openpaper.work>.

Ebenso gibt es einen IRC-Kanal für Live-Diskussionen: Freenode, channel #openpaperwork

Das sind die Anlaufstellen, wenn Sie Fragen zum Paperwork haben oder einfach nur plaudern wollen.

22 Probleme melden

Wenn Ihnen ein Fehler in Paperwork aufgefallen ist (und Sie sicher sind, dass es ein Fehler ist), können Sie einen Fehlerbericht erstellen.

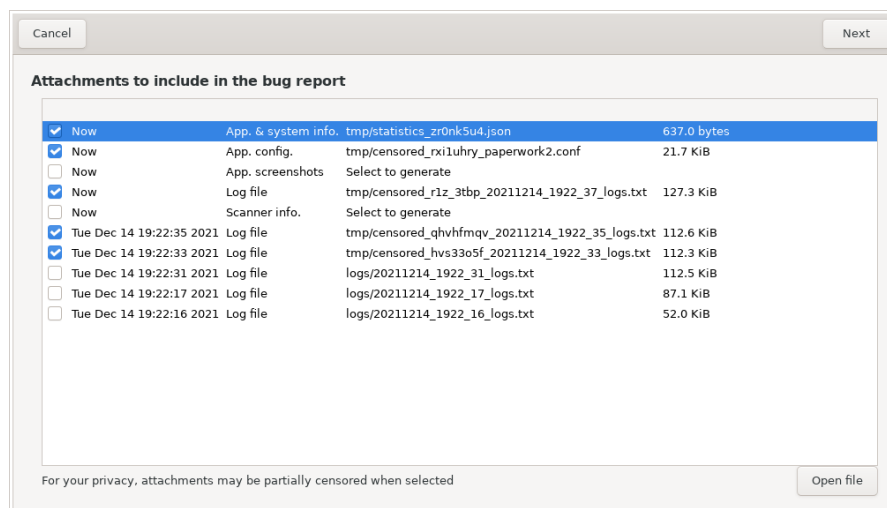
22.1 Fehlerverfolgung

Eine Möglichkeit, Fehlerberichte zu erstellen, ist das Erstellen von Tickets auf Paperwork bug tracker: <https://gitlab.gnome.org/World/OpenPaperwork/paperwork/issues>.

Dies ist der empfohlene Weg, um einen Fehlerbericht einzureichen, falls Sie ihn mit den Entwicklern von Paperwork besprechen möchten.

Um sicherzustellen, dass Sie alle erforderlichen Informationen angeben, können Sie das in Paperwork integrierte Werkzeug verwenden (siehe unten).

22.2 Automatischer Fehlerbericht



Paperwork enthält ein Werkzeug zur Erleichterung beim Melden von Fehlern. Es ermöglicht Ihnen, auf einfache Weise alle erforderlichen Informationen zusammenzustellen, um einen perfekten Fehlerbericht zu erstellen.

Um Ihre Privatsphäre zu schützen, werden alle Anhänge automatisch anonymisiert: Dokumentinhalte werden in Screenshots unscharf dargestellt und Protokolldateien werden durch Entfernen Ihres Benutzernamens anonymisiert.

Falls der zu meldende Fehler mit Scannern zusammenhängt, fügen Sie bitte "Scanner info." in den Fehlerbericht ein.

Wenn der zu meldende Fehler mit einem Darstellungsproblem zusammenhängt, fügen Sie den Fehlerberichtsdateien bitte "App. screenshots" bei.

22.2.1 ZIP Datei

Sie erhalten dann eine ZIP-Datei mit allen Daten. Bitte stellen Sie sicher, dass der Inhalt der ZIP-Datei keine privaten Informationen enthält (sollte er eigentlich nicht, aber sicher ist sicher). Dann können Sie diese ZIP-Datei einem Ticket auf Gitlab beifügen.

22.2.2 Automatische Übertragung

Sie können auch zulassen, dass das Programm den Fehlerbericht automatisch an `openpaper.work` sendet. In diesem Fall können Sie den Fehler nicht mit den Entwicklern besprechen (oder Sie müssen im Fehlerbericht eine Möglichkeit zur Kontaktaufnahme mit Ihnen hinterlassen).

Wenn Sie die automatische Übertragung verwenden, gibt Ihnen das Programm eine URL, um den übermittelten Fehlerbericht zu sehen. Diese URL ist privat und sollte nicht weitergegeben werden, bis Sie sichergestellt haben, dass der Fehlerbericht keine privaten Informationen enthält. Wenn es private Informationen gibt, können Sie die Löschung des Fehlerberichts beantragen, indem Sie eine E-Mail an `jflesch@openpaper.work` senden (bitte geben Sie die private URI in Ihrer E-Mail an, damit wir sicher sein können, dass Sie derjenige sind, der den Fehlerbericht eingereicht hat).

23 Deinstallation

Paperwork kann deinstalliert werden. Bei der Deinstallation von Paperwork werden Ihr Arbeitsverzeichnis oder Ihre Dokumente niemals entfernt.

23.1 GNU/Linux

Wenn Sie Paperwork mit dem Paketmanager Ihrer Distribution installiert haben (der empfohlene Weg), hängt die Deinstallationsmethode vom Paketmanager ab.

Beispielweise kümmert sich unter Unter GNU/Linux Debian oder GNU/Linux Ubuntu folgende Befehl darum:

```
sudo apt remove --purge paperwork-*
```

Wenn Sie es mit Flatpak installiert haben, können Sie den folgenden Befehl verwenden:

```
flatpak --user uninstall work.openpaper.Paperwork
```

23.2 Windows 10

Paperwork kann wie jede Windows-Anwendung deinstalliert werden, indem Sie in der Windows-Systemsteuerung auf "Anwendungen" klicken, Paperwork in der Liste suchen und dann auf "deinstallieren" klicken.