

➤ Voor goed persoonlijk document-beheer hoef je niet naar de cloud ◀

van het OCR-proces toont. Zie **screenshot 2**, waar de muis over het woord 'monumental' zweeft.

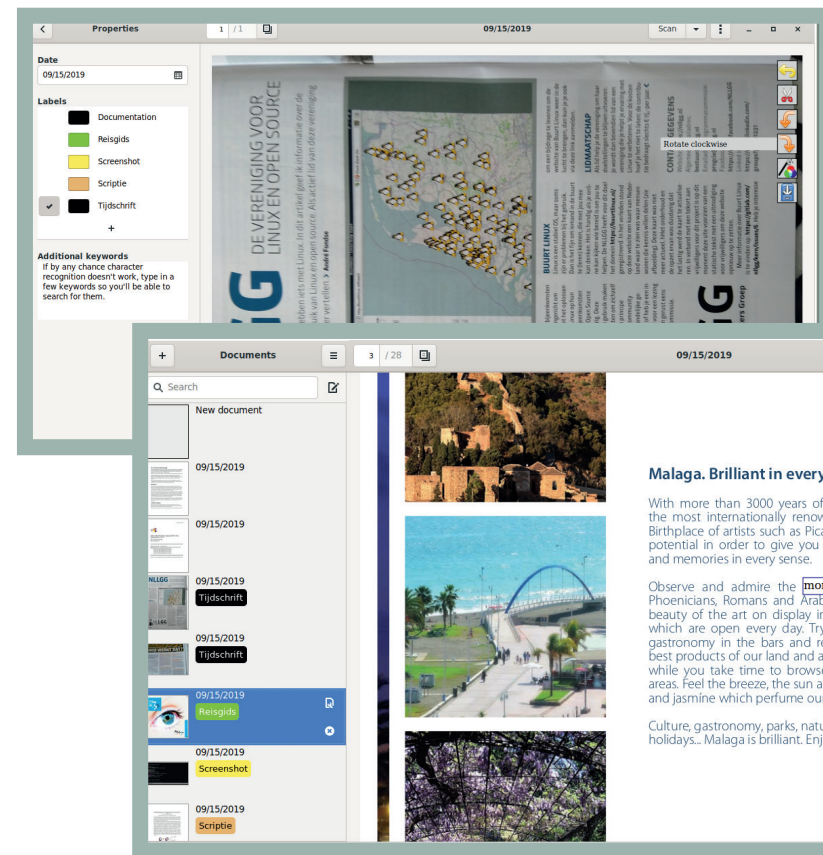
In de edit-modus beheer je de meta-data van het document. Je opent de edit-modus door op het edit-icoontje in het blauwe blokje te klikken. Je kunt nu desgewenst de document-datum aanpassen, labels toevoegen of verwijderen, en aanvullende sleutelwoorden intypen. Een document kan meerdere labels hebben. Bij het toevoegen van een label kies je voor een bestaand label of je maakt een nieuw label aan. Bij het aanmaken van een nieuw label kies je ook de kleur waarin dit label getoond wordt.

Ook kun je in de edit-modus enkele aanpassingen aan het document uitvoeren, zoals roteren. Zie **screenshot 4**.

DIRECTORIES IMPORTEREN

Paperwork biedt de mogelijkheid om een complete directory te importeren. Alle PDF-documenten worden in Paperwork geïmporteerd, waarbij reeds eerder geïmporteerde documenten worden overgeslagen. Wanneer de directory onderliggende sub-directories heeft, dan worden deze ook bezocht. Je kunt zo dus een complete directory-tree in één keer in Paperwork importeren. Na afloop krijg je een notificatie-windowtje met het aantal geïmporteerde documenten en het totaal aantal geïmporteerde pagina's. Ook nu krijg je de vraag, of je de geïmporteerde bestanden wilt verwijderen.

Nadat je al je documenten in Paperwork hebt opgeslagen, komt er een moment dat je iets wilt terugvinden. Bovenaan de linkerkolom staat hiervoor een zoekveld. Hier vul je een of meer woorden in. De linkerkolom toont vervolgens de documenten, die deze zoekwoorden bevatten. Wanneer je zo'n document opent, worden in de tekst de zoekwoorden met een groen kadertje gehighligt. Wanneer het document in grid-mode



◀ Screenshot 4: Paperwork in edit-modus.

✓ Screenshot 3: Popup met ge-OCR-ed woord.

is getoond, dan zijn de pagina's waarin een zoekwoord voorkomt, voorzien van een groen kader.

In de bovenbalk zie je in welke pagina je van het actieve document staat, bijvoorbeeld '3 of 13', met daar omheen een aantal pijl-icoontjes. We bespreken deze pijlen als paren, van buiten naar binnen. De buitenste twee pijl-icoontjes bevatten een dubbele pijl, hiermee ga je naar het vorige resp. volgende document. Daarbinnen staan twee pijlen, hiermee ga je naar de allereerste plek in het document waar een zoekwoord voorkomt, resp. naar de allerlaatste. Met de pijl-icoontjes daarbinnen ga je naar het vorige resp. volgende gevonden zoekwoord binnen het document.

Via de export-functie kun je documenten exporteren, in PDF formaat. Naast het hele document kun je ook een losse pagina exporteren, je hebt daarbij de keuze uit de formaten PNG en JPEG.

OPSLAG

Al je documenten komen in één enkele directory, in `~/papers`. Hierin vind je per document een subdirectory. De naam van deze subdirectory is afgeleid van de datum en tijd waarop je het document hebt gescand of geïmporteerd. In de subdirectory vind je het document,

een klein tekst-bestandje met informatie over het label of de labels die aan het document zijn toegevoegd, en een thumbnail afbeelding van het document. Bij geïmporteerde documenten is de oorspronkelijke bestandsnaam verdwenen, alle PDF documenten heten 'doc.pdf', geïmporteerde afbeeldingen zoals PNG-bestanden zijn omgezet naar JPEG-formaat en heten 'paper.1.jpg'.

Bij PDF-documenten is de hOCR informatie in het PDF-document opgenomen, bij afbeeldingen kan dit niet. Daar zie je een extra bestandje met de herkende woorden, met als naam 'paper.1.words'.

Het is belangrijk dat je deze papers-directory niet kwijtraakt, deze bevat immers al je documenten, inclusief de OCR-uitkomst en de meta-data. Op de website van Paperwork worden enkele back-up mogelijkheden genoemd, zoals het gebruik van SparkleShare (welke, al dan niet met end-to-end-encryptie, directories synchroniseert naar een Git-repository), Syncthing, enzovoorts. Welke methode je ook kiest, zorg voor een goede back-up.

Flatpack heeft op de achtergrond in je homedirectory de verborgen directory `~/var` gemaakt. Wanneer je hierin afdaald, vind je het persoonlijke configuratiebestand

voor Paperwork. Hierin kun je desgewenst een andere directory dan `~/papers` kiezen.

SMARTPHONE

Bij gebrek aan een scanner hebben wij naast PDF-documenten, voor deze review gebruik gemaakt van reeds eerder gescande documenten. Daarnaast hebben wij met onze smartphone twee foto's gemaakt in het Linux magazine, een van een enkele pagina, en een van twee opeenvolgende pagina's. Verder maakten we ook een foto van een deel van een pagina uit de ANWB Kampioen. Alle foto's werden correct in tekst omgezet. Verder hebben wij ook enkele screenshots geïmporteerd, zowel van terminal-vensters met tekst als van webpagina's, eveneens met een goed OCR-resultaat.

Paperwork is een desktopapplicatie en is uitgevoerd als een single user applicatie. Dit betekent dat je zonder authenticatie het programma opent en bij de gescande en geïmporteerde documenten komt. Het is dus echt een persoonlijk document beheersysteem. Voor goed persoonlijk documentbeheer hoef je dus niet naar de cloud. Documenten die je in Paperwork stopt, vind je op basis van een of enkele zoekwoorden razendsnel terug. ◀