

Dealing with scanned documents

LumiFind has a feature called: "Enable OCR fallback for PDFs".

If this feature is enabled under [Settings](#) -> [Analysis Settings](#), the PDFs found under the "Included Folders" will be further scanned with OCR tool to extract the text from them. This will significantly prolong the analysis time. Therefore, it is first recommended to start without this feature on.

For Mac and Linux users, this will work out-of-the-box. However, Windows users need to install third-party tools to achieve that.

On Windows:

1. Open a PowerShell and type: `winget install -e --id UB-Mannheim.TesseractOCR` (This tool is provided by Library of University Mannheim: https://ub-mannheim.github.io/Tesseract_Dokumentation/)
2. Download Ghostscript and install from: <https://ghostscript.com/releases/gsdnld.html> (Unfortunately, this needs to be done manually)

Both tools are free and open-source. You can check their requirements on their websites.

See also: [Dealing with documents as images](#)

Revision #5

Created 2025-11-22 19:51:04 UTC by MIDILLI

Updated 2026-03-01 09:54:01 UTC by MIDILLI