

2 Step

Document processing

FineReader Server processes document images via automatic recognition and document conversion, with optional verification and indexing capabilities.

Document recognition/OCR

FineReader Server's OCR process runs automatically on a dedicated workstation, the Processing Station. Using ABBYY's highly accurate OCR technology, FineReader Server supports a broad range of functions to increase recognition accuracy, including

- Image pre-processing (for example, splitting dual pages for book scans or clearing background noise)
- Print type definition (normal text, typewriter, dot-matrix, OCR-A, OCR-B, MICR E13b, and Gothic)
- Language definition (auto-recognition of more than 190 languages and historic texts in old fonts)

Depending on the document's quality and structure, processing mode can be set on either "precision" or "speed." To increase processing speed significantly — for example, to process many documents within a tight time frame — additional Processing Stations or a greater number of CPU cores can be added.

FineReader Server's scheduled processing feature enables it to process different kinds of documents at different times according to a predefined schedule.

Verification (optional)

In some cases — for example, when digitizing books — recognition results must be verified. FineReader Server's integrated Verification Station interface offers the option of correcting results, either on all documents or only on documents that did not reach a pre-defined recognition accuracy threshold.

Indexing (optional)

If required, document indexing can be done either manually — using the Indexing Station interface — or automatically using a script. Lists of index field values can be imported and synchronized with third-party systems.

3 Step

Document assembly and export

FineReader Server assembles processed pages into individual documents. Documents can be separated three different ways:

- Using blank sheets or barcode pages as separators
- According to a fixed number of pages per document
- According to a scripted rule

Assembled documents in the required formats are delivered to pre-defined output locations — such as network folders, SharePoint document libraries, and e-mail addresses — or passed to other applications connected via API.

Scripts can also be applied for intelligent routing and delivery of documents to enterprise content management systems based on document properties. FineReader Server supports a variety of output formats and enables creation of several output files at the same time.

FineReader Server can crawl individual libraries, detect non-searchable image-based documents, and convert them into searchable formats. Documents such as Microsoft Word files, PowerPoint presentations, or Excel spreadsheets, which require no processing, can be moved into the output library to the same position.

Multiple output formats:

- PDF, PDF/A-1a, PDF/A-1b, PDF/-2a, PDF/A-2b, PDF/A-2u, PDF/-3a, PDF/A-3b, PDF/A-3u, PDF/E, PDF/UA
- RTF
- DOC, DOCX
- XLS, XLSX
- TXT, CSV
- HTML
- TIFF
- JPEG, JPEG 2000
- PNG
- EPUB
- XML
- Alto XML
- FineReader internal format (FineReader Engine-compatible)