

Workflow - Conversion of PDF/scanned images to Microsoft Word format using ABBYY Finereader



1. Import document into Finereader

Note:

For scanned PDFs, you can improve image quality using Image Editor. This must be done for each page before Step 2. Useful tools include Brightness and Contrast, Deskew, Correct Trapezoid Distortion, Straighten Text Lines, etc.

2. Mark and read recognition areas

Notes:

For simple layouts, you can do page analysis automatically to get faster results. Afterwards, remove redundant recognition areas, e.g. repeated headers/footers. Modify area size, change area type if appropriate, then re-read updated areas.

3. Check for recognition issues

- low-confidence characters
- spelling errors
- division into paragraphs (one big paragraph instead of many)

4. Save Document / Save Selected Pages

Notes:

If some pages need to be skipped, either delete them from Pages panel or Shift-select pages you want to convert and then use Save Selected Pages command.

5. Check text recognition errors in Microsoft Word

Check and fix

- low-confidence characters (highlighted)
- spelling mistakes

6. Automatically fix / check for formatting issues in Microsoft Word

TransTools – Document Cleaner – Autoformat / Quick Actions

- Identify textboxes and selectively move them into main story
- Convert multi-column text to single column
- Convert wrong tables of contents to text
- Replace tab characters between words with spaces
- Convert wrong numbered list markers to text
- Correct mixed font size/color/font in same paragraph

Remove redundant page breaks and section breaks.

TransTools - Unbreaker

- Remove wrong paragraph breaks/line breaks

7. Manually format document

- Paragraph spacing
- Paragraph and list indentation, left/center/right alignment
- Incorrect numbering in numbered lists
- Put text back into columns where appropriate
- Check for merged paragraphs and split them if appropriate
- Put text into headers/footers
- Merge multi-page tables.

8. Remove excessive tags (“tag soup”)

TransTools – Document Cleaner – Tag Cleaner tab

9. Check inline formatting against PDF

10. Prepare images containing text

- Overlay textboxes on top of images.
- Mark text areas on top of picture area to let Finereader create textboxes for you.
- Translate image using Paint.net or another image editor.
- Use Infix Editor / TransPDF to translate text over images, then copy large-scale screenshot from Adobe Reader to Microsoft Word.

11. Pre-flight check

- Visual check against PDF
- Test import into CAT tool to check for excessive tags and incorrect segmentation