

JUMBO BLOCK ENGINEERING SUITE

JUMBO BLOCK Convert

User Manual · Version 0.3.2

© 2026 ZANNI GROUP · <https://jumboblock.org>

Contents

1. Overview
2. Getting Started
3. How the Conversion Works
4. Choosing a Save Location
5. Limitations, Notes and Best Practices
6. Language (German/English)
7. Quick Reference of Symbols and Buttons

1. Overview

JUMBO BLOCK Convert is the JUMBO BLOCK Engineering Suite's tool for turning PDF files into Word documents. Many suite tools deliberately produce only PDF reports; Convert closes that gap for anyone who needs to keep working on the content in Microsoft Word.

The program runs entirely locally in the browser. There is no cloud connection, no user account, no cookies, and no transmission of the file to a server — processing takes place solely in the memory of the open browser tab.

Core principle of the conversion: priority is given to producing a result that is well editable in Word, not to an exact pixel-for-pixel visual copy of the PDF original. Clean, editable flowing text is the goal, even if that means accepting minor visual differences from the original.

2. Getting Started

Operation is deliberately reduced to three steps, with no preview and no additional settings:

- “Select PDF” opens the operating system's local file picker (only PDF files are accepted).
- “Convert” becomes active only after a valid file has been selected and triggers the full conversion.
- The save dialog then appears, or the file is downloaded automatically (see chapter 4).

The status text beneath the buttons always shows the current state: “No PDF selected”, “Converting ...”, “Done”, “Cancelled”, or “Conversion failed” (full overview in chapter 7).

3. How the Conversion Works

Clicking “Convert” runs the file through several automatic processing steps:

- Text extraction: for every page, all text characters are read from the PDF file together with their position, font, and font size.
- Header/footer detection: text lines that recur identically on at least two pages in the top or bottom margin area, or that match a typical page-number pattern (e.g. “Page 3”), are recognized and placed into a real Word header or footer instead of appearing in the body text.
- Paragraph reconstruction: individual PDF text lines are merged into coherent, editable paragraphs based on line spacing, indentation, and font changes — instead of carrying over every single PDF line as its own, needlessly fragmented paragraph.
- Table detection: sequences of at least three consecutive lines with matching column positions are recognized as a table and rebuilt as a real, editable Word table.
- Images: embedded raster graphics (e.g. photos, logos stored as pixel images) are extracted from the PDF file and embedded into the Word document.
- Font mapping: PDF font names are mapped to the closest standard Word font (e.g. Helvetica to Arial, Times to Times New Roman); unknown or embedded fonts fall back to Calibri.

4. Choosing a Save Location

Once the conversion is finished, and provided the browser supports the File System Access API (e.g. Chrome, Edge), the operating system's native "Save as" dialog opens, letting the destination folder and file name be chosen freely.

If the browser does not support this feature (e.g. Firefox, Safari), the file is instead downloaded automatically to the browser's download folder — without a dialog, but likewise without any further transmission.

The suggested file name matches the PDF file's name with the extension ".docx" in place of ".pdf".

5. Limitations, Notes and Best Practices

- No pixel-perfect 1:1 copy: the goal is a result that is easy to edit in Word, not a visually identical reproduction of the PDF original.
- Multi-column layouts (e.g. newspaper-style layouts with several text columns side by side) are not specifically recognized; text may appear in an unexpected reading order.
- Table detection requires at least three consecutive, column-aligned lines. Shorter tables (one or two lines) are carried over as plain flowing text with visible spacing, not as a real Word table.
- Headers and footers are only recognized when a line recurs identically on at least two pages, or matches a page-number pattern. One-off or per-page-varying headings near the page margins (e.g. chapter titles) are deliberately kept in the body text.
- Vector graphics (embedded as paths rather than pixel images) are not converted — only embedded raster graphics are.
- Scanned PDF files without an embedded text layer (pure image documents) cannot be converted, since Convert relies on the PDF's text layer and does not perform text recognition (OCR).

6. Language (German/English)

The language switch is handled by the suite's global menu (menue.js) above the tool. Convert switches all labels and status messages between German and English. The language setting has no effect on the content of Word documents already produced.

7. Quick Reference of Symbols and Buttons

Button/Status	Meaning
Select PDF	Opens the local file picker for a PDF file
Convert	Starts text extraction, paragraph/table reconstruction, and the save dialog
No PDF selected	Initial state before a file has been selected
Converting ...	The conversion is currently running
Done	Conversion completed successfully and the file was saved or downloaded
Cancelled	The save dialog was closed without saving
Conversion failed	An error occurred during processing