

JUMBO BLOCK ENGINEERING SUITE

JUMBO BLOCK Convert

Benutzerhandbuch · Version 0.3.2

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

Inhalt

1. Überblick
2. Erste Schritte
3. Wie die Umwandlung funktioniert
4. Speicherort wählen
5. Grenzen, Hinweise und Best Practices
6. Sprache (Deutsch/Englisch)
7. Kurzreferenz der Symbole und Schaltflächen

1. Überblick

JUMBO BLOCK Convert ist das Umwandlungswerkzeug der JUMBO BLOCK Engineering Suite für den Weg von PDF nach Word. Viele Werkzeuge der Suite erzeugen ihre Berichte bewusst ausschließlich als PDF; Convert schließt die Lücke für alle, die den Inhalt anschließend in Microsoft Word weiterbearbeiten müssen.

Das Programm läuft vollständig lokal im Browser. Es gibt keine Cloud-Anbindung, kein Benutzerkonto, keine Cookies und keine Übertragung der Datei an einen Server – die Verarbeitung findet ausschließlich im Arbeitsspeicher des geöffneten Browsers statt.

Grundprinzip der Umwandlung: Priorität hat eine gute Bearbeitbarkeit des Ergebnisses in Word, nicht eine pixelgenaue optische 1:1-Kopie der PDF-Vorlage. Ein editierbarer, sauber strukturierter Fließtext ist das Ziel – auch wenn dafür kleinere optische Abweichungen vom Original in Kauf genommen werden.

2. Erste Schritte

Die Bedienung ist bewusst auf drei Schritte reduziert, ohne Vorschau oder zusätzliche Einstellungen:

- „PDF auswählen“ öffnet die lokale Dateiauswahl des Betriebssystems (nur PDF-Dateien werden akzeptiert).
- „Konvertieren“ ist erst nach einer gültigen Dateiauswahl aktiv und stößt die vollständige Umwandlung an.
- Anschließend erscheint der Speichern-Dialog bzw. die Datei wird automatisch heruntergeladen (siehe Kapitel 4).

Der Statustext unterhalb der Schaltflächen zeigt jederzeit den aktuellen Zustand: „Keine PDF ausgewählt“, „Wird konvertiert ...“, „Fertig“, „Abgebrochen“ oder „Konvertierung fehlgeschlagen“ (vollständige Übersicht in Kapitel 7).

3. Wie die Umwandlung funktioniert

Beim Klick auf „Konvertieren“ durchläuft die Datei mehrere automatische Verarbeitungsschritte:

- Textextraktion: Für jede Seite werden alle Textzeichen mit Position, Schriftart und Schriftgröße aus der PDF-Datei gelesen.
- Kopf-/Fußzeilen-Erkennung: Textzeilen, die identisch auf mindestens zwei Seiten im oberen bzw. unteren Randbereich wiederkehren, oder die einem typischen Seitenzahl-Muster entsprechen (z. B. „Seite 3“), werden erkannt und als echte Word-Kopf- bzw. -Fußzeile angelegt statt im Fließtext zu erscheinen.
- Absatz-Rekonstruktion: Einzelne PDF-Textzeilen werden anhand von Zeilenabstand, Einzug und Schriftwechsel zu zusammenhängenden, editierbaren Absätzen zusammengefasst – statt jede PDF-Zeile als eigenen, unnötig zerstückelten Absatz zu übernehmen.

- **Tabellenerkennung:** Folgen von mindestens drei aufeinanderfolgenden Zeilen mit übereinstimmenden Spaltenpositionen werden als Tabelle erkannt und als echte, bearbeitbare Word-Tabelle wiederaufgebaut.
- **Bilder:** Eingebettete Rastergrafiken (z. B. Fotos, Logos als Pixelbild) werden aus der PDF-Datei extrahiert und in das Word-Dokument eingebettet.
- **Schriftart-Zuordnung:** PDF-Schriftnamen werden auf die nächstliegende Standard-Word-Schriftart abgebildet (z. B. Helvetica auf Arial, Times auf Times New Roman); unbekannte oder eingebettete Schriftarten fallen auf Calibri zurück.

4. Speicherort wählen

Nach Abschluss der Umwandlung öffnet sich – sofern der Browser die File System Access API unterstützt (z. B. Chrome, Edge) – der native „Speichern unter“-Dialog des Betriebssystems, in dem Zielordner und Dateiname frei gewählt werden können.

Unterstützt der Browser diese Funktion nicht (z. B. Firefox, Safari), wird die Datei stattdessen automatisch in den Download-Ordner des Browsers heruntergeladen – ohne Dialog, aber ebenfalls ohne weitere Übertragung.

Der vorgeschlagene Dateiname entspricht dem Namen der PDF-Datei mit der Endung „.docx“ anstelle von „.pdf“.

5. Grenzen, Hinweise und Best Practices

- **Keine pixelgenaue 1:1-Kopie:** Ziel ist ein in Word gut bearbeitbares Ergebnis, keine optisch identische Nachbildung der PDF-Vorlage.
- **Mehrspaltige Layouts** (z. B. Zeitungslayout mit mehreren nebeneinanderliegenden Textspalten) werden nicht gesondert erkannt; der Text kann dabei in einer unerwarteten Lesereihenfolge erscheinen.
- **Tabellenerkennung** setzt mindestens drei aufeinanderfolgende, spaltenartig ausgerichtete Zeilen voraus. Kürzere Tabellen (ein bis zwei Zeilen) werden als gewöhnlicher Fließtext mit sichtbaren Abständen übernommen, nicht als echte Word-Tabelle.
- **Kopf-/Fußzeilen** werden nur erkannt, wenn eine Zeile auf mindestens zwei Seiten identisch wiederkehrt oder einem Seitenzahl-Muster entspricht. Einmalige oder von Seite zu Seite wechselnde Überschriften im Randbereich (z. B. Kapitelüberschriften) bleiben bewusst im Fließtext erhalten.
- **Vektorgrafiken** (als Pfade statt als Pixelbild eingebettet) werden nicht umgewandelt – nur eingebettete Rastergrafiken.
- **Gescannte PDF-Dateien** ohne eingebettete Textebene (reine Bilddokumente) können nicht umgewandelt werden, da Convert auf der Textschicht der PDF-Datei aufbaut und keine Texterkennung (OCR) durchführt.

6. Sprache (Deutsch/Englisch)

Die Sprachumschaltung erfolgt über das globale Suite-Menü (menue.js) oberhalb des Werkzeugs. Convert wechselt dabei sämtliche Beschriftungen und Statusmeldungen zwischen Deutsch und Englisch. Die Sprache hat keinen Einfluss auf den Inhalt bereits erzeugter Word-Dokumente.

7. Kurzreferenz der Symbole und Schaltflächen

Schaltfläche/Status	Bedeutung
PDF auswählen	Öffnet die lokale Dateiauswahl für eine PDF-Datei
Konvertieren	Startet Textextraktion, Absatz-/Tabellenrekonstruktion und den Speichern-Dialog
Keine PDF ausgewählt	Ausgangszustand vor der Dateiauswahl
Wird konvertiert ...	Die Umwandlung läuft gerade
Fertig	Umwandlung erfolgreich abgeschlossen und Datei gespeichert bzw. heruntergeladen
Abgebrochen	Der Speichern-Dialog wurde ohne Speichern geschlossen
Konvertierung fehlgeschlagen	Während der Verarbeitung ist ein Fehler aufgetreten