

Warum wir Word-Dateien vor dem Hochladen in die Wissensdatenbank umwandeln

Eine KI-gestützte Wissensdatenbank soll euch schnelle, präzise Antworten liefern – aber nur, wenn sie die Inhalte eurer Dokumente auch wirklich versteht. Genau hier liegt das Problem mit klassischen Word-Dateien.

Das Problem: optische Formatierung ist für die KI "Rauschen"

Word-Dokumente enthalten viele Elemente, die für das menschliche Auge hilfreich sind, für eine KI aber Störfaktoren darstellen:

- **Tabellen** – besonders verschachtelte oder mit verbundenen Zellen – werden beim Einlesen oft in eine unlogische Reihenfolge zerlegt. Eine Tabelle mit Produktname, Preis und Verfügbarkeit kann so zu einer Zeile werden, in der Preis und Produktname vertauscht erscheinen.
- **Kopf- und Fußzeilen** (z. B. Seitenzahlen, Firmenname, Datum) werden teils mitten im Fließtext eingefügt und unterbrechen einen zusammenhängenden Absatz.
- **Textfelder** liegen technisch "über" dem eigentlichen Text. Viele Auswerteprogramme überspringen sie komplett – wichtige Inhalte, die in einer Infobox oder einem Kasten stehen, gehen dadurch schlicht verloren.
- **Verschachtelte Aufzählungen** (Listen mit mehreren Ebenen, Unterpunkten, Einrückungen) verlieren beim Einlesen häufig ihre Hierarchie – aus "Hauptpunkt mit drei Unterpunkten" wird eine flache Liste ohne erkennbare Zusammengehörigkeit.

- **Eingebettete Bilder und Sonderzeichen** können die Textstruktur durcheinanderbringen, etwa wenn ein Bild mitten im Satz platziert ist und der Text danach an einer unerwarteten Stelle "wieder einsetzt".

Für Menschen ist das alles selbstverständlich lesbar. Eine KI muss diese Struktur aber erst technisch entschlüsseln – und tut das nicht immer zuverlässig.

Die Lösung: Umwandlung in ein einfacheres Format (z. B. Markdown)

Wird ein Dokument vorher in ein reduziertes Format wie Markdown umgewandelt, bleibt nur der reine Inhalt übrig: klare Überschriften, saubere Absätze und eindeutig getrennte Abschnitte. Die KI kann dadurch viel besser erkennen,

- wo ein Thema anfängt und wo es aufhört,
- was die wichtigsten Punkte sind,
- und wie einzelne Informationen inhaltlich zusammenhängen.

Praxisbeispiel: Eine mehrseitige Word-Datei mit Firmenlogo im Kopfbereich, einer verschachtelten Tabelle zu Ansprechpartnern und einem Textfeld mit "Wichtiger Hinweis" liefert nach der Umwandlung einen klar gegliederten Text mit Überschriften pro Abschnitt, einer einfachen Tabelle ohne Formatierungscode und dem Hinweistext direkt im Fließtext, statt in einer separaten, möglicherweise unsichtbaren Box.

Das Ergebnis für eure Fragen an die Wissensdatenbank

Wenn später jemand eine Frage stellt, findet die KI die passende Antwort schneller und genauer, weil sie nicht erst durch Formatierungs-Ballast "hindurchsuchen" muss. Zusätzlich verhindert die Umwandlung typische Fehlerquellen:

- Tabellen, die durch die Umwandlung unübersichtlich werden,
- Textfelder oder Kopf-/Fußzeilen, die komplett ignoriert werden,
- Sonderzeichen oder eingebettete Bilder, die den Lesefluss der KI unterbrechen.

Nicht jedes Dokument ist gleich betroffen

Ein einfaches Word-Dokument mit reinem Fließtext, wenigen Überschriften und ohne Tabellen, Textfelder oder komplexe Layouts bereitet in der Regel auch ohne Umwandlung keine größeren Probleme. Der Umwandlungsschritt lohnt sich vor allem bei Dokumenten mit:

- Tabellen,
- mehrspaltigem Layout,
- Kopf-/Fußzeilen mit relevantem Inhalt,
- Textfeldern oder Infoboxen,
- oder stark verschachtelten Aufzählungen.

Fazit

Der zusätzliche Umwandlungsschritt kostet ein paar Minuten, sorgt aber dafür, dass eure Dokumente in der Wissensdatenbank so ankommen, wie sie gemeint sind – klar strukturiert und für die KI optimal lesbar. Der konkrete Nutzen für euch: weniger falsche oder unvollständige Antworten, weil die KI eure Inhalte von Anfang an richtig erfasst.

Version #2

Erstellt: 2026-09-11 13:57:47 UTC von Frank Böttger

Zuletzt aktualisiert: 2026-09-11 13:58:37 UTC von Frank Böttger