

ASCII-Dateien zur Volltextindexierung

Mit der Volltextindexierung können Sie **alle Wörter Ihres Dokuments suchfähig** machen, nicht nur die Wörter der Dokumentbeschreibung, die Sie im Formulareditor erstellen. Alle Wörter in reinen Textdateien (*.txt) werden in den Volltext übernommen. Beachten Sie bitte, dass z.B. DOC- oder PDF-Dateien nicht reinen Text, sondern zusätzlich noch Formate und graphische Elemente enthalten. Auch die Auszeichnungen in HTML- oder Latex-Quell-Dateien sind kein Text in diesem Sinne. Textdateien haben im allgemeinen das Suffix ".txt" und enthaltenen nur Inhaltstext, der aus Ziffern, Buchstaben, Satzzeichen und Zeilenwechselln besteht. Ihr Textverarbeitungsprogramm erlaubt Ihnen die Speicherung einer geöffneten Datei als reine Textversion.

Problematisch sind dabei die nicht im lateinischen Alphabet enthaltenen Zeichen, also auch die deutschen Umlaute und das ß. Windows nutzt den so genannten ANSI-Zeichensatz, andere Betriebssysteme wiederum arbeiten mit dem ASCII-Zeichensatz. Für eine korrekte Suche mit solchen Zeichen ist der **ASCII-Zeichensatz** vorausgesetzt.

Angenommen, Ihr Dokument besteht aus 3 HTML-Dateien, von denen eine als Startdatei Links auf die beiden anderen enthält, und aus einer Reihe von GIF-Dateien für die Bilder. Um Volltextsuche zu ermöglichen, erstellen Sie mit Ihrem Textverarbeitungsprogramm aus den 3 HTML-Dateien jeweils reine Textversionen. Diese 3 Textdateien legen Sie in dasselbe Verzeichnis wie die anderen Dateien des Dokuments. Diese Dateien werden mit den anderen Dateien in die DBT kopiert, sind aber nicht über die Startdatei erreichbar. Allerdings sind sie in der Detailanzeige und in der automatisch erstellten gepackten ZIP-Datei des Derivats enthalten. Die DBT fügt alle Wörter der Textdateien in den Volltextindex ein.