

## Reiner Text

Eine Datei, die nur Textzeichen enthält, ist „reiner“ Text. So „rein“ ist aber der Text gar nicht, denn es gibt einige Möglichkeiten, „reinen“ Text zu speichern.

### Reiner Text im Notepad

Notepad ist ein Programm, das solche einfachen, unformatierten Texte verarbeitet. Im Bild rechts sind einige Zeichen eingegeben worden. Speichert man einen solchen Text, gibt es vier Speicheroptionen.

- ANSI
- Unicode
- Unicode Big Endian
- UTF-8

Wählt man die Option „ANSI“ und als Namen „demo-ansi.txt“, enthält das Dokument Zeichen, die in dieser Speicheroption nicht darstellbar sind, bekommt man eine Fehlermeldung (Bild rechts oben) zu sehen, die darauf hinweist. Am Bildschirm bleiben dabei die Zeichen unverändert. Würde man aber die gespeicherte Datei neu in den Editor laden, fehlen diese Zeichen.

Man sieht, nicht alle Zeichen, die im Editor darstellbar sind, können in diesem „ANSI“-Format auch gespeichert werden, denn die ANSI-Kodierung kann nur 256 Zeichen darstellen.

Im Beispieltext wäre das „Große scharfe s“ betroffen.

In den anderen Optionen (Unicode und UFT-8) gibt es dieses Problem nicht, diese verfügen über einen praktisch unbeschränkten Zeichenraum.

### Reiner Text in Word

Würde man diesen „reinen“ ANSI-Text in Word laden, hat Word die Schwierigkeit, diesem Text eine Kodierung zuzuweisen. Da man das nicht so ohne Weiteres erraten kann, fragt Word in einem eigenen Dialog nach, wie die Kodierung des Textes einzustellen ist und zeigt dabei die ersten Zeilen an und schlägt eine Kodierung vor.

Im ersten Beispiel wurde dem Text die Kodierung „MS-DOS“ zugewiesen und statt der Umlaut-Zeichen sieht man jetzt Block-Grafik-Zeichen.

Wählt man „Windows-Standard“, schaut das Ergebnis besser aus. Mit „Andere Kodierung“ könnte man auch noch entlegene Kodierungen wählen.

Diesen Dialog sehen wir daher immer dann, wenn Word die Kodierung eines „reinen Textes“ nicht bestimmen kann.

Bei allen alten „reinen Texten“ ist nicht an jeder Kodeposition dasselbe Zeichen, wobei mit „alt“ die Zeit vor 1995 gemeint ist. Heutzutage wird dieses Zuordnungsproblem hauptsächlich beim Umgang mit alten Texten auftreten. Die Kodierung, die man zuweisen muss, wird auch Codepage genannt.

Bei neueren „reinen“ Textdokumenten, die mit dem Notepad generiert worden sind, besteht dieses Verständnisproblem nicht, sofern sie mit den Optionen Unicode, Unicode BE oder UTF-8 gespeichert worden sind.

Alle diese Texte haben auch die Endung .txt aber die Editoren können an deren besonderen Aufbau die Kodierung eindeutig erkennen.

Wie, das sieht man im Kasten auf der folgenden Seite.

