

möglich zu sein, weil pro Zeichen immer zwei Bytes reserviert sind. Es stimmt aber leider nicht ganz, denn der Unicode kennt die Besonderheit der diakritischen Zeichen, die es erlauben Akzente oder Umlaute und andere diakritische Zeichen mit anderen Zeichen zu kombinieren. Dabei wird im Zeichenstrom das diakritische Zeichen dem anderen Zeichen nachgestellt. Das aber wieder bedeutet, dass bei Verwendung diakritischer Zeichen die Anzahl der Bytes in einer UTF-16-Textdatei nicht automatisch doppelt so groß ist wie die Anzahl der Zeichen. Leider gilt das auch für die UTF-32-Kodierung, wo dann sogar 8 Bytes für ein kombiniertes Zeichen verbraucht werden.

Eingabe der Unicode-Zeichen

Möchte man ein Zeichen in einem Text eingeben, das auf der Tastatur nicht verfügbar ist, empfiehlt sich als anschauliches und programm-unabhängiges Auswahlwerkzeug die Zeichentabelle (Bild rechts oben).

Innerhalb von Office-Programmen erreicht man die Zeichen über Einfügen -> Symbol -> Weitere Symbole (Bild rechts unten).

Die Eingabe von Zeichen in Office erweist sich interessanterweise flexibler und umfassender als jene über die Zeichentabelle. Man kann nämlich die Zeichen auch als CodePoint eingeben und außerdem werden bei seltenen Fonts auch tatsächlich alle Unicode-Bereiche angezeigt, was bei der Zeichentabelle nicht der Fall ist.

Direkte Eingabe in Word

Es gibt aber auch noch eine unmittelbare Eingabemöglichkeit in Word.

Man gibt den hexadezimalen Zahlenwert des CodePoint ein, zum Beispiel 5d0 (oder auch U+5d0) und drückt danach Alt-c. Damit wandelt sich der Code-Wert in das betreffende Zeichen.

Das funktioniert auch in der Gegenrichtung. Man stellt den Cursor rechts vor irgend ein Zeichen, dessen CodePoint man wissen will und klickt auf Alt-c.

Frühere Eingabemöglichkeiten

Die Eingabemöglichkeit über die Tastenkombination Alt und Eingabe des Kode über den Ziffernblock ist wegen der kleiner werdenden Tastaturen nicht mehr allgemein anwendbar.

Unicode Fonts

Die üblichen Hauptschriftarten in Windows wie Courier New, Times New Roman, Arial, Verdana und Tahoma können schon sehr viele der Unicode-Zeichen darstellen. Inzwischen ist aber der Zeichenvorrat des Zeichenraums von Unicode auf weit über 100.000 Zeichen angewachsen und ein True Type Font kann nur 65536 Glyphen kodieren. Die richtige Antwort ist daher: keiner.

Man muss aber unterscheiden zwischen früheren Fonts, die im Zeichenbereich bis 255 Zeichen darstellen und echten Unicode-Fonts, die diese Zeichen im jeweils richtigen UnicodeBlock darstellen.

Schmankerl

Hier ein „Schmankerl“ aus der *Supplementary Code Plane*, ein paar altägyptische Hieroglyphen, gezeichnet durch den Unicode-Font *Aegyptus*.

