



A_DATEN_CODIEREN_TXT

Der ASCII-Code

Der ASCII-Code (American Standard Code for Information Interchange) ist in seiner Ursprungsversion eine 7-Bit-Zeichenkodierung und wurde bereits damals eingesetzt, wo noch Textnachrichten per Fernschreiber (Telex) übermittelt wurden. Die druckbaren Zeichen umfassen das lateinische Alphabet in Gross- und Kleinschreibung, die zehn arabischen Ziffern sowie einige Interpunktionszeichen (Satzzeichen, Wortzeichen) und andere Sonderzeichen. Der Zeichenvorrat entspricht weitgehend dem einer Tastatur oder Schreibmaschine für die englische Sprache. Die nicht druckbaren Steuerzeichen enthalten Ausgabezeichen wie Zeilenvorschub oder Tabulator, Protokollzeichen wie Übertragungsende oder Bestätigung und Trennzeichen wie Datensatztrennzeichen.

ASCII-Erweiterung gemäss ISO 8859

ISO: International Organization for Standardization / Internationale Organisation für Normung

ASCII belegte ursprünglich 7 Bit pro Character (0...127) und wurde später um ein Bit auf 8 Bit (0...255) erweitert. Um den verschiedenen Sprachen gerecht zu werden, wurde der ISO-Standard 8859 definiert, der nun im zweiten Teil des ASCII-Zeichensatz (128...255) deren 16 Sprachzusätze unterscheidet.

Bsp.: ISO-Standard 8859-1 = Latin-1, Westeuropäisch oder ANSI-ASCII.

Der Unicode

Der ASCII-Code mit seinen 256 Zeichen genügt den heutigen Anforderungen nicht mehr. Es fehlen z.B. die chinesischen Schriftzeichen oder wie wär's mit einem Violinschlüssel? Es muss ein umfangreicherer Zeichencode her, nämlich der Unicode.

- Ein Unicode kann max. 8 Byte lang sein (64 Bit): U+XXXX'XXXX, wobei Unicode V2.0 bisher erst 1'114'112 verschiedene Zeichen U+0000'0000 bis U+0010'FFFF nutzt.
- UTF ist die verbreitetste Unicode-Kodierungsform (UTF=Universal-Coded-Character-Set Transformation Format)
- UTF-8: Je nach Zeichen beträgt die Codelänge von UTF-8 ein bis vier Bytes. UTF-8 ist in den ersten 128 Zeichen (Indizes 0–127) deckungsgleich mit ASCII.
- UTF-16: Je nach Zeichen beträgt die Codelänge von UTF-16 zwei oder vier Bytes. Zusätzlich muss mit der Byte-Order-Mark BOM (=Bytereihenfolge) angegeben werden, ob nach BigEndian BE oder LittleEndian LE verfahren wird. (Bsp. Datum : BE wäre yyyy.mm.dd, LE wäre dd.mm.yyyy)
- UTF-32: Ein einzelnes Zeichen belegt immer exakt 32 Bit.
- Notation bei HTML: &#x<unicode>; (hexadezimale Notation des Unicodes)
- Eingabe bei Windows-Word: U+<unicode> gefolgt von der Tastenkombination Alt+C.
- Wenn das Unicode-Zeichen im gewählten Font-Satz (Arial, CourierNew etc.) nicht implementiert ist, wird auf dem Bildschirm oder am Drucker ein Stellvertreter-Zeichen oder ein Leerzeichen dargestellt.

Hinweis: Notepad++ versteht nebst ANSI-ASCII auch UTF-8 und UTF-16 mit LE-BOM und BE-BOM.



Hier folgen Aufgaben zum Thema. Siehe separates Aufgabenblatt.

Zusammengesetzte Codierung

Mit zusammengesetzter Codierung ist ein Datensatz (Record, Tupel etc.) gemeint. Ein Datensatz ist eine Gruppe von inhaltlich zusammenhängenden und zu einem Objekt gehörenden Datenfeldern. (Z.B. Artikelnummer, Artikelname, Farbe, Länge, Breite etc.) Datensätze entsprechen einer logischen Struktur, die bei der Softwareentwicklung (Datenmodellierung) festgelegt wurde. In der Datenverarbeitung werden zu Datensätzen zusammengefasste Daten in Datenbanken oder in Dateien gespeichert. Sie sind Gegenstand der Verarbeitung von Computerprogrammen und werden von diesen erzeugt, gelesen, verändert und gelöscht.

Barcodes



- **EAN-13:** Das Bildchen mit den verschiedenbreiten schwarzen und weissen Balken, wie man es heutzutage auf allen Food- und Non-Food-Artikeln antrifft, repräsentiert eine 13-stellige Zahl. Diese 13 Zahlen sind vom Produktheersteller oder einer Organisation weiter aufgeschlüsselt wie z.B. in Ländercode, Produktcode, Lotnummer, Datum, Prüfziffer etc.
- **QR-Code:** Der QR-Code ist im Gegensatz zum EAN-13-Code ein zweidimensionaler Code. Dank einer "eingebauten" Fehlerkorrektur können bis zu 30% der QR-Grafik beschädigt oder verschmutzt sein, ohne die Lesbarkeit zu beeinträchtigen (gilt für Fehlerkorrektur-Level H). Diese Eigenschaft wird oft zur Platzierung von Werbebildchen und Logos missbraucht. Der maximale Informationsgehalt (177x177 Elemente, max. Verlust von 7% der Daten bzw. Fehlerkorrektur-Level L) beträgt knapp 3kB. Das würde dann ca. 7000 Dezimalziffern oder 4300 alphanumerische Zeichen ergeben. Vorsicht: Weil der Inhalt eines QR-Codes nicht auf den ersten Blick ersichtlich ist, kann man einem Fake-Link auf schädliche Webseiten aufsitzen oder das Smartphone führt Malware aus.



Hier folgen Aufgaben zum Thema. Siehe separates Aufgabenblatt.