

Zeichencodierung

In [wikipedia.de](http://de.wikipedia.org)¹ findet man zu dem Begriff „Code“ folgende Erläuterung:

„Unter einem Code [...] wird eine Vorschrift verstanden, in der Symbole einer Darstellung in solche einer anderen übertragen werden. Beispielsweise stellt der Morsecode eine Übertragungsvorschrift zwischen Buchstaben und einer Abfolge kurzer und langer Tonsignale her. Im allgemeinen Sprachgebrauch versteht man unter einem Code meist einen geheimgehaltenen Code, der zur Verschlüsselung von Botschaften verwendet wird.“

In der Realität gibt es zahlreiche Codes, die jeweils nach bestimmten Gesichtspunkten (z.B. Fehlererkennung und -korrektur) entwickelt wurden.

BCD-Code

BCD steht für Binary Coded Decimal. Das bedeutet, dass die Dezimalziffern 0 bis 9 binär codiert werden. Dieser Code (auch 8-4-2-1-Code genannt) codiert die die zehn Dezimalziffern mit jeweils vier Bit.

Da sich mit vier Bit jedoch 16 verschiedene Zustände codieren lassen, bleiben sechs Zustände ungenutzt. Diese sechs zur Codierung unbenutzten Zustände bezeichnet man als sogenannte **Pseudotetraden**².

0		8	
1		9	
2		10	
3		11	
4		12	
5		13	
6		14	
7		15	

Mit dem BCD-Code lassen sich beliebig lange Zahlen darstellen, da jede einzelne Ziffer durch die zugehörige Bitkombination ersetzt wird. Allerdings ist dieser Code für Berechnungen sehr unhandlich.

ASCII-Code

Mit einem 4-Bit Code lassen sich maximal 16 Zeichen darstellen. Daher wurde 1968³ in der noch heute gültigen Fassung ein Code zur Darstellung von Schriftzeichen festgelegt, der diese mittels 7 Bit codiert. Mit diesen sieben Bit lassen sich 2^7 Zeichen (das sind _____ Zeichen) codieren.

Die ersten 32 Zeichen dieses Codes sind sogenannte Steuerzeichen. Steuerzeichen sind **keine** darstellbaren Schriftzeichen, sondern dienen der Steuerung eines Gerätes. So hat z.B. das Zeichen 13_D die Bedeutung des Wagenrücklaufes (Carriage Return) oder das Zeichen 12_D die Bedeutung des Seitenvorschubes (Form Feed).


Die darstellbaren Zeichen des ASCII-Codes beginnen mit dem Zeichen 32_D ($=20_H$). Dieses Zeichen ist das Leerzeichen. Die Zeichen 32_D bis 127_D (20_H bis $7F_H$) enthalten im Wesentlichen alle wichtigen Satzzeichen, Groß- und Kleinbuchstaben sowie die Ziffern 0 bis 9.

Da Computer auch in früherer Zeit bereits Datenmengen von mehr als 7 Bit gleichzeitig verarbeiten konnten (8-Bit Prozessoren) wurde das 8.te Bit zur Darstellung von weiteren 128 Sonderzei-

¹ am 20. September 2005

² Tetrade bedeutet „Vierheit“

³ <http://de.wikipedia.org/wiki/ASCII> am 20. September 2005

Arbeitsblatt Nr. 4	Q1 Technologie: Digitaltechnische Grundlagen		B S G G
Datum:	Thema: Zeichencodierung		
Seite 2 von 2	Name:		

chen genutzt. Diese Sonderzeichen waren jedoch nicht festgelegt.

ANSI-Code

„American National Standard Institute. ANSI-Code ist ein genormter Satz aus 256 Zeichen (Buchstaben, Ziffern, Interpunktionszeichen und Sonderzeichen). Jedem der 256 darstellbaren Zeichen ist eine Zahl zwischen 0 und 255 zugewiesen. Der ANSI-Code wird von Windows und Windows-Programmen zur Zeichendarstellung verwendet. Nicht-Windows-Programme verwenden zur Zeichendarstellung dagegen meist den ASCII-Zeichensatz. Die Belegung des ANSI- und des ASCII-Zeichensatzes ist für die Zeichen mit den Nummern 32 bis 127 identisch und enthält das lateinische Alphabet sowie Satz- und Währungszeichen. In den Werten 128 bis 255 finden sich etliche Sonderzeichen wie deutsche Umlaute, französische Akzentzeichen oder das spanische N.“⁴

Der ANSI Zeichensatz basiert also ebenfalls auf dem ASCII-Zeichensatz. Nun sind jedoch die Zeichen 128_D bis 255_D (80_H bis 7F_H) normiert und damit unabhängig von einer bestimmten Zeichensatztafel.

Unicode

Auch hier noch ein Zitat aus Wikipedia:

„Unicode ist ein internationaler Standard, in dem langfristig für jedes sinntragende Zeichen bzw. Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt wird. Er will das Problem der verschiedenen inkompatiblen Kodierungen in den unterschiedlichen Ländern beseitigen. Herkömmliche Computer-Zeichensysteme umfassen einen Zeichenvorrat von 128 (7 Bit) wie der sehr bekannte ASCII bzw. 256 (8 Bit) Zeichen, wie z. B. ISO Latin-1, wovon nach Abzug der Steuerzeichen 96 bzw. 192–224 als Schrift- und Sonderzeichen darstellbar sind. Diese Zeichenkodierungen erlauben die gleichzeitige Darstellung von nur wenigen Sprachen im selben Text, wenn man sich nicht damit behilft, in einem Text verschiedene Schriften mit unterschiedlichen Zeichensätzen zu verwenden. Dies behinderte den internationalen Datenaustausch erheblich. In Unicode finden sämtliche Zeichen bestehender Zeichensätze nach Industriestandards, sogenannten Codepages, und nationalen sowie ISO-Normen eine 1:1-Entsprechung.“⁵

Für die Speicherung eines einzelnen Zeichens wurden jetzt anstatt 8 Bit (1 Byte) nun 16 Bit (2 Byte) verwendet. Damit lassen sich 2^{16} Zeichen codieren.

Da jedoch auch diese Zeichenmenge nicht ausreichte, definierte man einen Unicode, der mit 32 Bit (4 Byte) arbeitet. Dieser wird jedoch nur dann verwendet, wenn entsprechende Zeichen (z.B. darzustellen sind.

Unter <http://www.unicode.org/charts/> lassen sich für die jeweiligen Schriftzeichen die Kodierungen als PDF-Dokumente laden

Übung

1. Stellen Sie die Zahl 1989 im Binärsystem und BCD codiert dar.
2. Ermitteln Sie anhand einer ASCII-Code-Tabelle die Codes für Ziffern, Klein- und Großbuchstaben. Um wie viele Positionen unterscheiden sich die Codes der Klein- und Großbuchstaben?
3. Um welche Art von Zeichen handelt es sich bei den Codenummern 0 bis 31? Nennen Sie die Bedeutung der Zeichencodes 0x0A_H 0x0D_H 0x0C_H 0x07_H
4. Nennen Sie drei weitere in der EDV verwendete Codes.

⁴ <http://www.evodion.de/opencms/export/evodionIT/General/Glossar/a17.html> am 20. September 2005

⁵ <http://de.wikipedia.org/wiki/Unicode> am 20. September 2005