# ASCII und andere

Wenn man mit dem Computer und dem Internet arbeitet, begegnet man des Öfteren der Möglichkeit, die Zeichencodierung auswählen zu können, so z.B. im Browser oder im E-Mail-Programm. Die Wenigsten kennen überhaupt den Unterschied, daher stelle ich einmal die verschiedenen Codierungen vor:

Zeichensatzcodierungen sind im Grunde nur Tabellen, die bestimmten Byte-Werten konkrete Zeichen zuordnen. Anfangs waren dies nur alphanumerische Zeichen, heutzutage umfassen Zeichensätze z.B. auch Runen oder kyrillische Zeichen.

## ASCII

Der ASCII (American Standard Code for Information Interchange) entstand in den 60er Jahren, ist 7-Bit codiert, nutzt also nur die ersten 128 Kombinationsmöglichkeiten. Er umfasst alphanumerische, Sonder- und Steuerzeichen und ist von den verwendeten Zeichen her auf die englische Sprache ausgerichtet.

Neben dem Standard-ASCII-Code gibt es noch den erweiterten ASCII-Code, bei dem nicht nur die ersten 128 Möglichkeiten genutzt werden, sondern auch noch die "letzten" 128. Die Varianten des erweiterten ASCII-Codes nennt man ANSI und ISO-8859.

## ANSI

ANSI ist eine Erweiterung des ASCII-Codes, er ist genormt vom American National-Standards Institute (daher der Name) und hat sich als Standard auf den Windows- und Macintosh-Betriebssystemen durchgesetzt.

## ISO-8859

Bei diesem Zeichensatz handelt es sich erstmal um eine Normserie mit länderspezifischen Zeichensätzen der ISO (International Organization for Standardization). Der Zeichensatz ISO-8859-1 (Latin) ist der erste von insgesamt 16 ländereigenen Zeichensätzen und ist für Westeuropa gemacht, beinhaltet also z.B. deutsche, französische oder skandinavische Sonderzeichen.

## Unicode

Entstanden Ende der 80er Jahre mit dem Ziel alle Sprachen der Welt in einem Zeichensatz zu vereinen, ist der Unicode der größte und umfassendste Zeichensatz. Anfangs 16-Bit codiert, allerdings 2001 umgestellt auf 32-Bit beinhaltete Unicode 4 im Jahre 2003 ca. 100000 verschiedene Zeichen. Der Unicode vereint tote wie auch lebende Sprachen, so sind z.B. auch Runen Bestandteil.

## UTF-8

Im Grunde eine Unicode-Variante mit hohem ASCII-Anteil, die Abkürzung bedeutet Unicode Transformation Format 8-Bit. Dieser Zeichensatz ist besonders im Internet weit verbreitet. Es ist angestrebt, dass er sich zum Standard entwickelt, indem alle neuen Internetkommunikationsprotokolle sich auf UTF-8 verstehen sollen. Da dies allerdings keine Verpflichtung, sondern nur eine Empfehlung oder "Bitte" ist, sieht es mit diesem Standard aktuell (März 2007) eher mäßig aus.