**Publizieren Vorlesung 3**

**Aufgabe von Zeichenkodierungen**

Eine Zeichenkodierung beschreibt die konkrete Zuordnung eines [Codepoints](https://wiki.selfhtml.org/wiki/Codepoint) zu einem Byte oder einer Bytesequenz. [UTF-8](https://wiki.selfhtml.org/wiki/UTF-8) und UTF-16 sind beispielsweise Kodierungen für den Zeichensatz Unicode.

**Übung:**

Unicode ist ein internationaler Standard, in dem langfristig für jedes Sinn tragende Schriftzeichen oder Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt wird.

Ziel ist es, die Verwendung unterschiedlicher und inkompatibler **Kodierungen** in verschiedenen Ländern oder Kulturkreisen zu beseitigen. Unicode wird ständig um **Zeichen** weiterer Schriftsysteme ergänzt.

ISO 10646 ist die von ISO verwendete, praktisch bedeutungsgleiche Bezeichnung des Unicode Zeichensatzes; er wird dort als **Universal Coded Character Set (UCS)** bezeichnet. Der Zeichenumfang ist dazu in 17 **Ebenen** gegliedert, welche jeweils 256 Kacheln à 256 Positionen (65.536 **Zeichen**) umfassen. Sechs dieser Ebenen werden bereits verwendet, die restlichen sind für spätere Nutzung reserviert: Die **Basic Multilingual Plane** auch als Plane 0 bezeichnet, enthält hauptsächlich Schriftsysteme, die aktuell in Gebrauch sind, Satzzeichen und Symbole, Kontrollzeichen und **Surrogate-Paare**, und einen privat nutzbaren Bereich (PUA). Neben dem eigentlichen Zeichensatz sind auch eine Reihe von **Zeichenkodierungen** definiert, die den Unicode Zeichensatz implementieren und die benutzt werden können, um den vollen Zugriff auf alle Unicode-Zeichen zu haben. Sie werden **Unicode Transformation Format** genannt; am weitesten verbreitet sind zum einen **UTF-16** das sich als interne Zeichendarstellung einiger Betriebssysteme (Windows, OS X) und Softwareentwicklungs-Frameworks (Java, .NET) etabliert hat, zum anderen **UTF-8**, das ebenfalls in Betriebssystemen (GNU/Linux, Unix) sowie in verschiedenen Internetdiensten (E-Mail, WWW) eine große Rolle spielt.

**Fragen:**

Im Oktober 1991 wurde nach mehrjähriger Entwicklungszeit die Version 1.0.0 des Unicode-Standards veröffentlicht. Was trug zu einer so schnellen Entwicklung des Unicode Zeichensatzes bei?

**In den 1980er und 1990er Jahren wurde es zunehmend üblich in einem Text verschiedene Schriften mit unterschiedlichen Zeichensätzen zu verwenden. Dies erschwerte den Datenaustausch enorm**

Das Ziel von Unicode ist es:

**Den weltweiten Datenaustausch zu vereinfachen**

**Einen universalen Zeichensatz, der alle Schriften der Welt umfassen soll, zu erschaffen**

**Zukünftig auch einen Standard bei den „Emotions“ zu schaffen**

Eine Zeichenkodierung ist primär verantwortlich für:

**Eine eindeutige Nummerierung von Schriftzeichen und Symbolen innerhalb eines Zeichensatzes**

Die Default-Deklaration von XML Texten ist:

**UTF-8**

Wie viele Zeichen sind aktuell in Unicode 8 ca. enthalten?

**136.690**

Durch die Verwendung der Surrogatpositionen für die Kodierung in UTF-16 wird im Vergleich zu UTF-32 folgendes erreicht:

**Platzsparendem Abspeichern auf Festplatten**