

12.11 Unicode

Definition

Unicode ist ein internationaler Standard zur Darstellung von Zeichen aus vielen Sprachen und Schriftsystemen.

Unicode erweitert die Möglichkeiten älterer Zeichencodierungen wie **ASCII**.

Ziel

Unicode soll möglichst alle weltweit verwendeten Zeichen eindeutig darstellen.

Dazu gehören:

- Buchstaben
 - Ziffern
 - Sonderzeichen
 - Symbole
 - Schriftzeichen verschiedener Sprachen
 - Emojis
-

Wichtige Begriffe

Begriff	Bedeutung
Unicode	Standard für Zeichen
Codepoint	eindeutige Nummer eines Zeichens
UTF-8	Codierung für Unicode-Zeichen
UTF-16	Codierung für Unicode-Zeichen
UTF-32	Codierung mit fester Länge

UTF-8

UTF-8 ist eine weit verbreitete Codierung für Unicode.

Eigenschaften:

- kompatibel zu ASCII
- variable Zeichenlänge
- häufig im Web verwendet
- speichereffizient bei lateinischen Zeichen

Vergleich ASCII und Unicode

ASCII	Unicode
7 Bit	deutlich größerer Zeichenvorrat
128 Zeichen	sehr viele Zeichen
hauptsächlich englische Zeichen	internationale Zeichen
keine Umlaute im ursprünglichen ASCII	Umlaute und Sonderzeichen möglich

Merksatz

Unicode ist der Standard für internationale Zeichen.
UTF-8 ist eine Codierung für Unicode.

Prüfungshinweis

In IHK-Aufgaben wird häufig zwischen **ASCII**, **Unicode** und **UTF-8** unterschieden.

Typische Fehler:

- Unicode und UTF-8 gleichsetzen.
- ASCII und Unicode verwechseln.
- Codepoint und Codierung verwechseln.
- Umlaute im ursprünglichen ASCII erwarten.

Quellen

- IT-Handbuch für Fachinformatiker
 - The Unicode Standard
-