

2.9 Unicode

Definition

Unicode ist ein internationaler Standard zur Darstellung von Zeichen.

Unicode kann Zeichen aus vielen verschiedenen Sprachen und Schriftsystemen darstellen.

Bedeutung

Unicode wurde entwickelt, um die Einschränkungen älterer Zeichencodierungen wie **ASCII** zu überwinden.

ASCII kann nur **128 Zeichen** darstellen.

Unicode kann deutlich mehr Zeichen darstellen.

Beispiele für Unicode-Zeichen

Zeichen	Bedeutung
A	lateinischer Großbuchstabe
ä	deutscher Umlaut
€	Euro-Zeichen
Ω	griechischer Buchstabe Omega

UTF-8

UTF-8 ist eine häufig verwendete Codierung für Unicode-Zeichen.

Eigenschaften:

- kompatibel zu ASCII
 - variable Zeichenlänge
 - weit verbreitet im Web
-

Vergleich ASCII und Unicode

ASCII	Unicode
7 Bit	deutlich größerer Zeichenvorrat
128 Zeichen	viele Schriftsysteme

ASCII	Unicode
vor allem englische Zeichen	internationale Zeichen
keine Umlaute im ursprünglichen ASCII	Umlaute und Sonderzeichen möglich

Merksatz

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

Prüfungshinweis

Unicode und UTF-8 dürfen nicht gleichgesetzt werden.

- **Unicode** = Zeichensatz / Standard
 - **UTF-8** = Codierung
-

Quellen

- IT-Handbuch für Fachinformatiker
 - The Unicode Standard
-