U

Unicode

Ethan V. Munson
University of Wisconsin-Milwaukee,
Milwaukee, WI, USA

Definition

Unicode is an international standard for representing text characters. Unicode supports the scripts of all human languages with any substantial level of use and has a flexible design that is capable of supporting all known human languages and all of their variant scripts. The development of the Unicode standard is coordinated by the Unicode Consortium.

Key Points

Unicode's development is motivated by the need to encode characters in all languages without conflicts between the encodings for different languages. Obviously, the achievement of this goal is fraught with technical and political complexities.

Unicode has several different encodings. The most widely used is the 8-bit, variable-width UTF-8 encoding, which permits the encoding of many European languages in an efficient 1-byte form and is backward compatible with both the ASCII and ISO-8859-1 character sets. UTF-16 is a 16-bit, variable-width encoding that is more suitable to languages with many characters, such as Chinese, Japanese, and Korean. UTF-32 is a 4-byte, fixed-width encoding that encompasses all Unicode characters.

The Unicode Consortium is supported by many well-known computer and software manufacturers. For the members of the consortium, the system internationalization problem is of compelling importance.

Cross-References

- [Document](#)
- [Document Representations \(Inclusive Native and Relational\)](#)
- [Unified Modeling Language \(UML\)](#)
- [XML](#)