

**"Unicode & Linguists"**  
**Thursday Linguistics Seminar**  
**03 September 2009**  
**presented by Russ Nekorчук**

**Abstract**

“The Unicode Standard is the universal character encoding standard for written characters and text. It defines a consistent way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation for global software. As the default encoding for HTML and XML, the Unicode Standard provides a sound underpinning for the World Wide Web and new methods of business in a networked world. Required in new Internet protocols and implemented in all modern operating systems and computer languages such as Java and C#, Unicode is the basis of software that must function all around the world.”

-Chapter 1, Introduction, The Unicode Standard, Version 5.0

Computers were originally designed to work in English, using Latin characters, Arabic numerals, and American punctuation. As computers proliferated throughout the world, so did the techniques of representing other languages and characters. Unicode was developed as a single standard and is identical to the specifications of the International Standards Organization and the International Electrotechnical Commission, ISO/IEC 10646. It is now possible for computers to handle every known written language of the world, past and present.

Unicode, fonts, and character rendering software all interact to display written language data. Anyone involved with languages that have writing systems beyond the basic Latin characters should have some familiarity with these features and their interaction.