

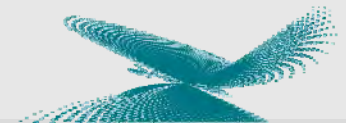


Internationalized Help Systems

Unicode in HTML Help Systems

Petra Thiemann/Christine Platz

02. November 2006



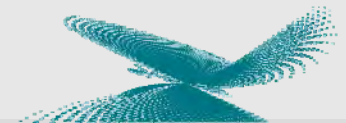


Content

- Unicode Basics
 - What is Unicode?
 - How are characters defined?
 - How are characters encoded?
 - Unicode and HTML

- Unicode in HTML Help Systems
 - How to: Preferences
 - How to: Steps
 - How to: System steps
 - How to: With Flare

- Further Information



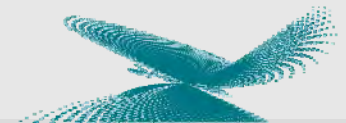


What is Unicode?

- A character set
- An international standard which is designed to integrate all characters known worldwide into one character set
- A character set which allows you to display multiple languages in one and the same text

Unicode has been developed by

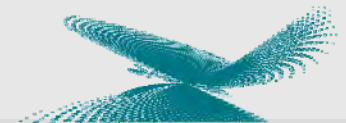
- the International Standardisation Organisation (ISO) and
- the Unicode consortium





How are characters defined?

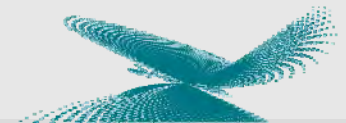
- A Code Point is assigned to each character
- Unicode 4.0 contains 1.114.112 Code Points
- A Code Point is listed in hexadecimal form following the prefix „U+“
- Examples:
 - U+0041 - Latin capital letter A
 - U+0423 - Kyrillic capital letter U
 - U+FB2A - Hebrew letter ψ (Shin)





How are characters encoded?

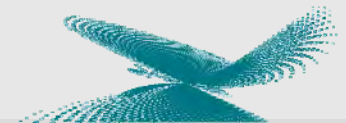
- The Unicode Standard defines two mapping methods:
 - UTF (Unicode Transformation Format) encoding
UTF-8, UTF-16, UTF-32, UTF-EBCDIC
 - UCS (Universal Character Set) encoding
UCS-2, UCS-4
- UTF-8 uses one to four bytes per character.
- Only one byte is needed to encode the 128 US-ASCII characters (Unicode range U+0000 to U+007F).





Unicode and HTML

- HTML uses Numeric Character References/Character Entities
- A character entity has the form “&#N;”
- N represents a decimal or hexadecimal number for the Unicode Code Point.
- Example: U+03A3
Σ Σ Σ Σ
Greek capital letter Σ (Sigma)

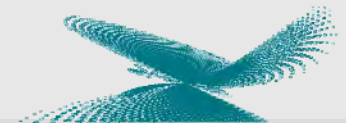




How to: Preferences

- Unicode font is installed on the machine.
- HTML Help Workshop is installed.
- The translated text is available.

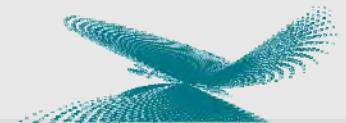
- Use always the same language settings in the help project e.g. for russian always kyrilic (russian) not kyrilic (mongolian).





How to: Steps

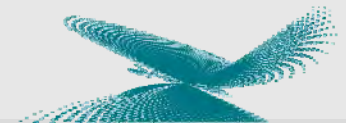
1. Create new project.
2. Set language for help compiler.
3. Assign the relevant font to the help project.
4. Import the translated HTML files.
5. Check or insert the language meta tag within HTML files.
6. Create TOC and Index.
7. Assign the relevant font to TOC and Index.
8. If you use a style sheet, ensure the language settings match.
9. Enter TOC or Index entries in the correct language





How to: System steps

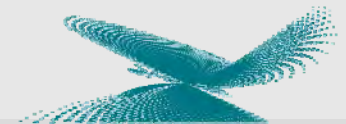
- Change the keyboard setting for entering TOC and Index entries
 - Set the system locale
 - Set the user locale
 - Compile and hope the best...
-
- After a successful compilation do not forget to reset system and user locale!!





How to: With Flare

- Import unicode project
 - Check or insert keywords
 - Check or insert TOC entries
 - Compile and hope the best
 - Change system locale if necessary
 - Compile again
-
- After a successful compilation do not forget to reset system and user locale!!





Further Informationen

- <http://en.wikipedia.org/wiki/Unicode>
- <http://en.wikipedia.org/wiki/Utf-8>
- Unicode Consortium:
<http://www.unicode.org>
- Code Charts of the Unicode Consortium:
<http://www.unicode.org/charts/>
- Unicode in XML and other Languages:
<http://www.unicode.org/unicode/reports/tr20/>
- Converting Table for UTF-8 and Unicode:
<http://www.utf8-zeichentabelle.de/>

