



DelphiLive!

The Delphi Developer Conference

Marco Cantù | Wintech Italia Srl

Unicode and Delphi 2009

marco.cantu@gmail.com

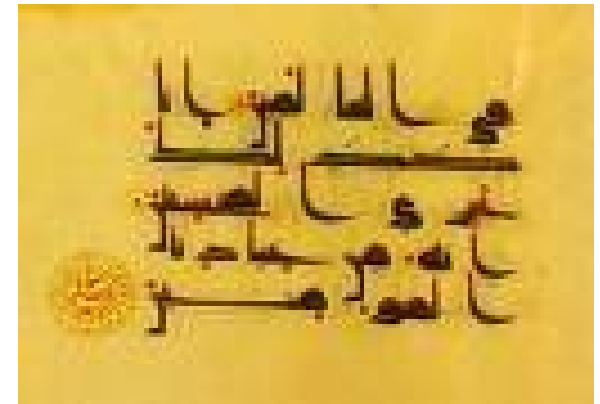
Marco Cantù

- Delphi books author
 - Mastering Delphi series
 - Essential Pascal
 - Delphi 2007 Handbook
 - Delphi 2009 Handbook
- Consultant, trainer
 - <http://blog.marcocantu.com>
 - <http://www.delphideveloperdays.com>
 - <http://twitter.com/marcocantu>



Overview

- Understanding Unicode
- WideChar and UnicodeString
- Encodings
- Porting Existing Code
- More Unicode Support



Understanding Unicode

- “Characters from the Past”
- Unicode Consortium
 - www.unicode.org
 - ASCII, ISO Encodings, Unicode
- Concepts
 - Code Points
 - Graphemes
 - UTF Encodings and BOMs
 - Descriptions

Unicode and Win32

- Windows API calls
 - Wide and Ansi versions
 - Delphi used to call Ansi, 2009 calls Wide
 - No Change: PChar now PWideChar
- Exceptions
 - WinExec, GetProcAddress (Ansi only)
- Font Substitution Issues

Char as WideChar

- Char and WideChar
 - PChar and PWideChar
 - UCS4Char (32-bit)
- The new Character unit
 - Conversions and tests

UnicodeString

- String = UnicodeString
 - Full UTF16, not UCS2
- The new internal layout
 - For all string types, including AnsiString

Bytes	CodeP.	Ref.Co.	Len	s[1]	s[2]
-------	--------	---------	-----	------	------

AnsiString Types

- AnsiString
- Defining custom string types
 - AnsiString as a type constructor
- UTF8String
- RAWByteString

Converting String

- Automatic Conversions
 - Watch out for speed issues
 - The new string conversion warnings
- The Ensure calls
 - Avoidable in your code

String Literals

- Remapping of €
- \$HIGHCHARUNICODE
- Warning
 - String literals in concatenation

Stream Encoding

- The new TEncoding class
 - Streams and encoding
- TString to and from files
 - Default loading based on BOM
- Writing custom encodings

More RTL Support

(Only a summary)

- TStringBuilder
- TTextReader and TTextWriter
- Unicode TDataSet

Unicode Source Code

- Unicode in Pascal source code files
 - The files are encoded, you decide how
- Unicode in DFM files
 - The code points are escaped

Porting Existing Code

- Set of Char
 - CharInSet (but slow)
- FillChar
 - StringOfChar
- Turn on conversion warnings
- Move and strings and buffers...

Porting Existing Code

- Using TBytes and RawByteString
- Dataset bookmarks
- Calling Ansi functions and the new AnsiStrings unit
 - Unicode only and Ansi only cases

Porting Existing Code

- Pointer Math
 - PChar and PByte
 - The new POINTERMATH directive
- Variants
 - varUString
- Variant records
 - vtUnicodeString

Conclusion

- One of the most relevant updates to a Delphi core feature strings
- High compatibility, along with many issues
- Extra Unicode complexity, often worth facing