

# Unicode-Integration in FrameMaker?

tekomp-Tagung 22.-23.4.2004, Aachen

FrameMaker Tutorial TF6

Dipl.-Ing. (Univ.) Michael Müller-Hillebrand, Erlangen

## *Inhalt*

Was kann FrameMaker offiziell?

Was fehlt?

Was braucht man für weitere Sprachen?

Font mit den gewünschten Zeichen

Wie kommt FrameMaker an die Zeichen?

Codepage-Prinzip aus alten Tagen

Eingabe der Zeichen

Anzeige der Zeichen in Dialogboxen

Aber XML braucht UTF-8?

Die bessere Option

Detaillierte Informationen

Rückfragen?

## **Was kann FrameMaker offiziell?**

FrameMaker kann offiziell Texte in die chinesische, japanisch, koreanisch (CJK) und in den sogenannten westeuropäischen Sprachen verarbeiten (MacRoman bzw. Windows-1252 bzw. ISO Latin-1).

## **Was fehlt?**

Unterstützung für Texte in den Sprachen der folgenden gegenwärtigen oder zukünftigen EU-Mitgliedsländer fehlt: Griechisch, Tschechisch, Slowakisch, Polnisch, Ungarisch, Estnisch, Lettisch, Litauisch.

Darüberhinaus natürlich auch: Arabisch, Hebräisch, Indisch (diverse), Thailändisch, Türkisch, Vietnamesisch, Russisch u.a. kyrillisch schreibende Sprachen, und fast alle Unicode-Unterbereiche (siehe Windows Systemsteuerung > Zeichentabelle).

## **Was braucht man für weitere Sprachen?**

Von hinten her:

- Font mit den gewünschten Zeichen
- Wie kommt FrameMaker an die Zeichen?
- Codepage-Prinzip aus alten Tagen

- Eingabe der Zeichen
- Anzeige der Zeichen in Dialogboxen

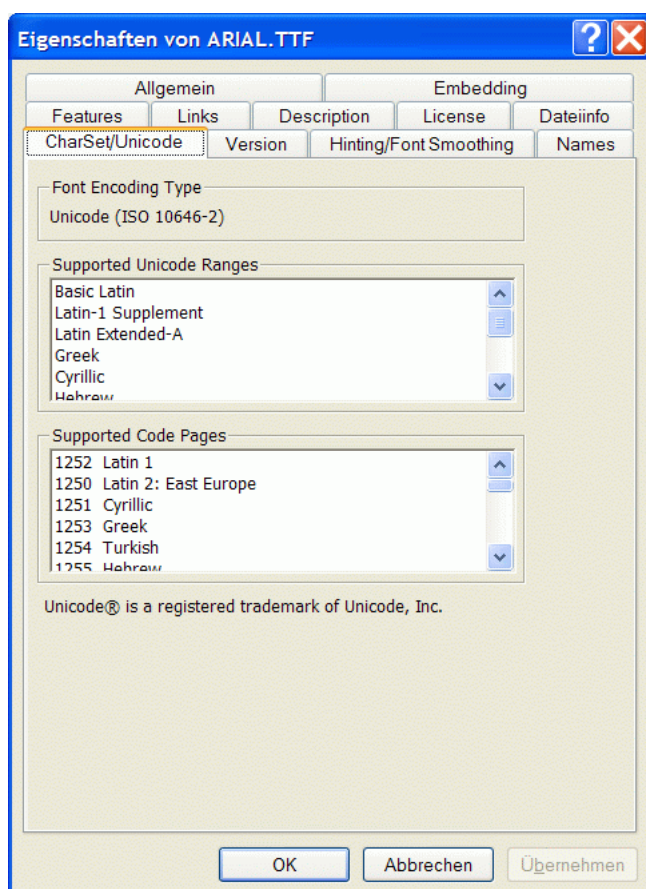
### Font mit den gewünschten Zeichen

Speziell angepasste Fonts, die z.B. an Stelle eines ä einen kyrillischen Buchstaben enthalten, sind wenig kompatibel, z.B. bei Font-Wechsel oder Export in andere Programme.

Dagegen enthalten OpenType- oder TrueType-Fonts mit erweitertem Zeichensatz (Windows Glyph List, WGL) die benötigten Zeichen in Unicode-Kodierung.

Sichtbar wird der Umfang der Zeichen mit der Microsoft Font Properties Extension:

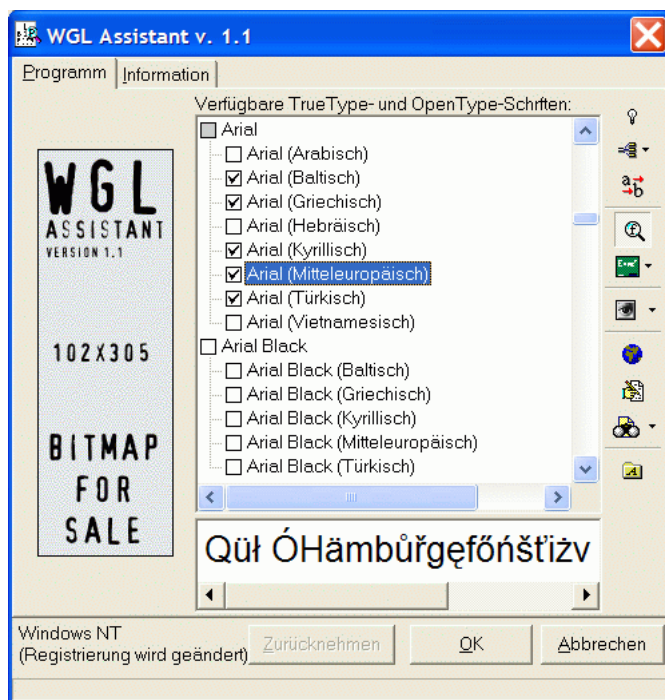
<http://www.microsoft.com/typography/property/property.htm?fname=%20&size>



### Wie kommt FrameMaker an die Zeichen?

Über den in Windows eingebauten Mechanismus "FontSubstitutes". Durch Einstellung in der Registry können Untermengen der Zeichen eines Unicode-kodierten Fonts unter eigenem Namen verfügbar gemacht werden. Der virtuelle Font "Arial Greek" enthält neben den ASCII-Zeichen eben auch die griechischen Zeichen an den Positionen 128...255.

Hilfsmittel zur Aktivierung der virtuellen Fonts: WGL Assistant <http://wgl.typ.pl/>



Aber in welcher Ordnung erscheinen die Zeichen?

### Codepage-Prinzip aus alten Tagen

Eine Codepage = eine Zeichenauswahl (ASCII plus 128 frei wählbare)

Der Zeichenvorrat einer Codepage umfasst in der Regel eine Sprachgruppe.

Verschiedene Codepages verwenden einen Bytewert für unterschiedliche Zeichen.

Windows kennt folgende Single Byte Character Set (SBCS) Codepages (in Klammern das Zeichen an Position 192= \xC0 der jeweiligen Codepage mit dem zugehörigen Unicode-Wert):

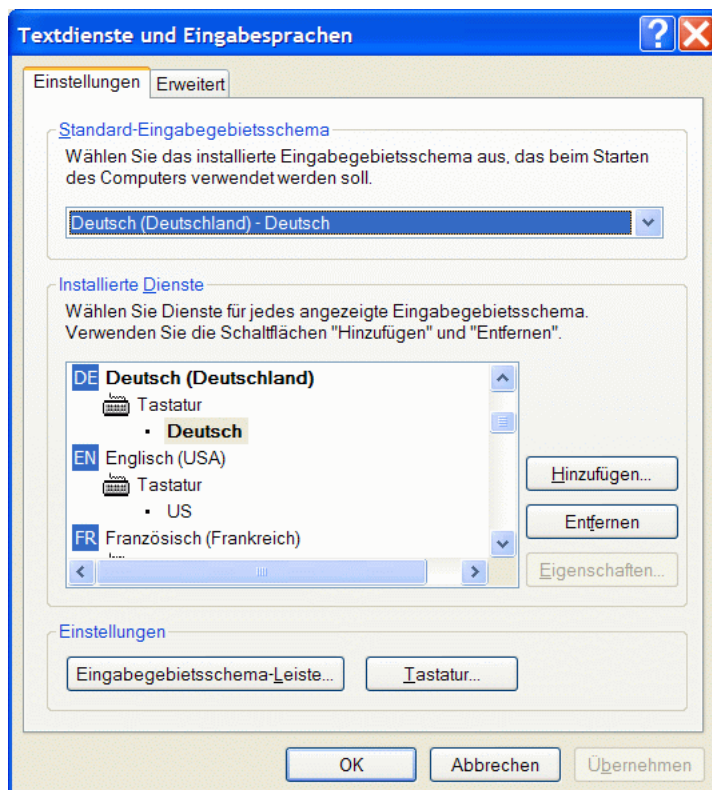
- 874 Thai (THAI CHARACTER PHO SAMPHAO \u0E20)
- 1250 Central Europe (LATIN CAPITAL LETTER R WITH ACUTE \u0154)
- 1251 Cyrillic (CYRILLIC CAPITAL LETTER A \u0410)
- 1252 Western (LATIN CAPITAL LETTER A WITH GRAVE \u00C0)
- 1253 Greek (GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS \u0399)
- 1254 Turkish (LATIN CAPITAL LETTER A WITH GRAVE \u00C0)
- 1255 Hebrew (HEBREW POINT SHEVA \u05B0)
- 1256 Arabic (ARABIC LETTER HEH GOAL \u06C1)
- 1257 Baltic (LATIN CAPITAL LETTER A WITH OGONEK \u0104)
- 1258 Vietnam (LATIN CAPITAL LETTER A WITH GRAVE \u00C0)

Sonderfall Arabisch und Hebräisch: Die Zeichenanzeige ist prinzipiell möglich, aber durch die geänderte Schreibrichtung kaum bearbeitbar.

## Eingabe der Zeichen

Dazu benötigen Sie eine landessprachliche Tastatur aus der jeweiligen Sprachgruppe.

Windows XP: Systemsteuerung > Regions- und Sprachoptionen > Sprachen > Details > Textdienste und Eingabesprachen



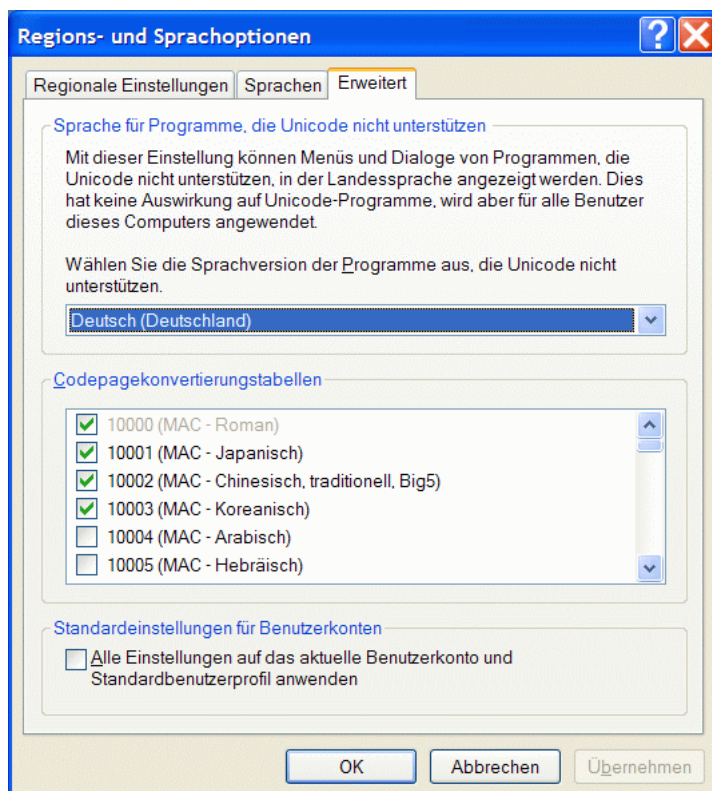
Für die Anzeige der Tastenbeschriftungen eignet sich zum Beispiel das Microsoft Visual Keyboard <<http://office.microsoft.com/germany/downloads/2000/viskeyboard.aspx>>

## Anzeige der Zeichen in Dialogboxen

Hier gilt der Windows-Systemfont und um hier korrekte Zeichen zu sehen (z.B. zur Bearbeitung von Indexeinträgen), kann das "system locale" des Betriebssystems umgestellt werden.

Windows 2000: Ländereinstellungen

Windows XP: Regions- und Sprachoptionen ("Sprache für Programme, die Unicode nicht unterstützen")



Weitere Option für Windows XP: Das Microsoft AppLocale Utility  
 <<http://www.microsoft.com/globaldev/tools/apploc.msp>>

## Aber XML braucht UTF-8?

Unicode beschreibt eine Zuordnung von Zeichen zu Zahlenwert. UTF-8 ist eine praktikable Umschrift der Unicode-Zeichenkodierung, bei der die ASCII-Zeichen unverändert bleiben, was z.B. die Lesbarkeit der XML-Tags gewährleistet.

FrameMaker kennt nur die "offiziellen" Zeichen und schreibt für Zeichen mit dem Bytewert 192 immer den Unicode-Wert `\u00C0` in eine XML-Datei. Ob das Zeichen zum Beispiel dem Font Arial Greek entstammt und deshalb `\u0390` werden müsste, bleibt unberücksichtigt.

Beim Import würde das Zeichen `\u0390` nicht sichtbar werden, sondern in einer Marke vom Typ UNKNOWNCHAR als Entity `&#x0390;` gespeichert und erst beim Speichern wieder als solches Unicode-Zeichen ausgegeben...

Um also in Verbindung mit XML/UTF-8 korrekt Text austauschen zu können, müsste eine Font-abhängige Vorverarbeitung stattfinden:

- Vor dem Speichern müssten alle unbekannte Zeichen `> 127` aus nicht-Western Fonts in passende Marken UNKNOWNCHAR umgewandelt werden
- Nach dem Import müssten alle Marken UNKNOWNCHAR in FrameMaker-Zeichen des gewünschten Fonts umgewandelt werden

## Die bessere Option

Eine derartige Zeichenzuordnung ist im Prinzip unkompliziert, die notwendigen Tabellen müssen pro Sprachfamilie (Codepage) einmal angelegt werden. Allerdings ist der Weg über die Marke UNKNOWNCHAR sehr umständlich.

Die derzeit überzeugendste Lösung ("so hätte es Adobe implementieren müssen") kommt von der Firma Fischer Computer Technik (FCT), die dies in ein eigenes XML-Import/Export-Plugin integriert hat. Damit können strukturierte Dokumente in valides XML zur Speicherung oder Weiterverarbeitung überführt werden. Und zwar auch mit älteren FrameMaker-Versionen!

<<http://www.fct.de/presse/presstext/pr1203.htm>>

## Detaillierte Informationen

Diese und viele weitere Informationen finden Sie auf <<http://one-world-publishing.de>>

## Rückfragen?

Für Rückfragen erreichen Sie den Autor unter <[mmh@cap-studio.de](mailto:mmh@cap-studio.de)>