

## Praxistipps FrameMaker: Zeichen setzen

*In den verschiedenen Versionen von Adobe FrameMaker werden Texte mit unterschiedlichen Zeichencodierungen umgesetzt und mit Fonts dargestellt, die nach Layout-Kriterien ausgewählt werden. Besonders für Projekte, in denen eine Technische Dokumentation in viele Sprachen übersetzt und über alle Sprachen ein möglichst einheitliches Layout eingehalten werden soll, ist es sinnvoll, vor der Planung die Schriftenbasis genauer unter die Lupe zu nehmen.*

**W**enn die FrameMaker-Dateien die Grundlage für verschiedene Ausgabeformate bilden, zum Beispiel PDF mit Lesezeichen, HTML- oder XML-Daten, ist es sehr hilfreich, in allen Datenformaten mit einer durchgehenden Zeichencodierung zu arbeiten. Dabei muss der Anwender von FrameMaker wissen: Erst ab Version 8 kann er auf eine Unicode-Codierung aufbauen. Versionen vor 7.2 können hingegen nicht einmal alle Zeichen von Codepages darstellen, die nicht westeuropäisch codiert sind.

### Zeichencodierung

Seit Beginn der elektronischen Datenverarbeitung entwickelten sich Zeichencodierungen parallel zu Rechnerarchitekturen und Betriebssystemen in großer Vielfalt. Beim Einsatz in FrameMaker spielen jedoch nur noch wenige eine Rolle:

**ANSI-Codierung** (1 Byte pro Zeichen) ist eine Zeichencodierung mit 218 Schriftzeichen. Sprachspezifische Codepages stellen verschiedene Zeichen ab Codepoint 128 zur Verfügung. Der in den Code-Tabellen enthaltene ASCII-Zeichensatz mit je 26 Groß- und Kleinbuchstaben des lateinischen Alphabets, Ziffern und wenigen Satzzeichen ist in allen Codepages gleich (Zeichen von 32 bis 127).

**MacRoman-Zeichencodierung** (1 Byte pro Zeichen) ist eine Zeichencodierung für Macintosh-Rechner und Mac OS (bis Version 9) mit 223 Schriftzeichen. Sie unterscheidet sich von der ANSI-

Kodierung unter anderem in der Behandlung der Umlaute und enthält teilweise Zeichen, die in der ANSI-Codierung nicht existieren, zum Beispiel das Apfel-Zeichen und spezielle Steuerzeichen. Die ANSI- und die MacRoman-Codierung wird von FrameMaker bis zur Version 7.2 zusätzlich in die **FrameRoman-Codierung** „verbogen“, um damit eine plattformübergreifende Arbeit mit direktem Dokumentenaustausch zu ermöglichen.

**Double Byte Zeichensätze** dienen zur Darstellung asiatischer Zeichen. In asiatischen Sprachen werden oft 3.000 bis 15.000 verschiedene Zeichen zur Darstellung benötigt.

Die Unicode-UTF-8-Kodierung (2-4 Byte pro Zeichen) kann von der Bitfolge „0000 0000“ bis „0010 FFFF“ insgesamt 1.114.112 verschiedene Zeichen darstellen. Die Codierung wurde entwickelt, um jedem Zeichen, das in allen Sprachen der Welt geschrieben werden kann, einen eindeutigen Code zuweisen zu können. Alle benötigten Schriftzeichen aller Sprachen können so in einer unicode-fähigen Schriftdatei enthalten sein – also sowohl Westeuropa, Mitteleuropa, Russisch, Türkisch, Arabisch und Japanisch sowie Grafikzeichen. Ein Beispiel hierfür ist die Schrift Arial Unicode MS.

### Arten von Fonts

FrameMaker nutzt die auf dem Betriebssystem installierten Schriften, um die entsprechend

codierten Zeichen mit bestimmten Glyphen darzustellen. Als Glyphe wird die grafische Darstellung eines Zeichens verstanden. Weiterhin existieren Dateien mit ähnlicher Bezeichnung für verschiedene Schnitte einer Schriftfamilie, zum Beispiel Fett, Kursiv und Fett Kursiv. Außerdem werden Schriftdateien verschiedener Dateiformate unterschieden:

- **Postscript-Fonts** überwiegend für ANSI-Codierungen
- **TrueType-Fonts** überwiegend für ANSI-Codierungen
- **OpenType-Fonts** wurden von Microsoft und Adobe entwickelt, um Font-Dateien plattformunabhängig nutzen zu können und eine bessere Unicode-Unterstützung zu ermöglichen. In einem OpenType-Font kann ein ehemaliger TrueType-Font (\*.ttf) oder auch ein ehemaliger Postscript-Font (\*.otf) stecken.

Die OpenType-Funktionalität für Ligaturen wird bisher von keiner FrameMaker-Version unterstützt. Die Nutzung von OpenType-Fonts ist keine Garantie dafür, dass

der Font auch alle Zeichen der Unicode-Palette abbildet und die Glyphen für mehrere Sprachen enthält.

Als Unicode-Schriften für die einheitliche Darstellung von Dokumenten in vielen Sprachen eignen sich so genannte **WGL4-Fonts** – **Windows Glyph List 4** mit 652 Zeichen für West- und Mitteleuropa, Kyrillisch, Griechisch und Türkisch.

**Virtuelle Fonts** nutzen einen unicode-basierten Font, um den vorhandenen Zeichenvorrat auch für ANSI-codierte Zeichen einzusetzen. Sie werden durch eine spezielle Konfiguration des Betriebssystems verfügbar gemacht. In einem virtuellen Font werden nur bestimmte, sprachspezifische Teile der Unicode-Palette angezeigt, zum Beispiel Arial CE oder Verdana Baltic.

**Kombinierte Schriften** entstehen als Programmunterstützung direkt in FrameMaker. Sie kommen zum Einsatz, wenn die asiatischen Schriftzeichen mit nicht-asiatischen Zeichen innerhalb von Absatzgrenzen kombiniert werden sollen. Kombinierte Schriften

bewirken, dass der Font-Wechsel zwischen diesen Zeichen automatisch vollzogen wird. Ohne kombinierte Schrift müsste der Autor die nicht-asiatischen Worte in asiatisch formatierten Absätzen extra mit einem Zeichenformat versehen. Der Befehl zum Einrichten von kombinierten Schriften ist im Menü **Format > Dokument** angeordnet. Um diesen Befehl in den Versionen 7.2 und 8 sichtbar zu machen, kann die Programm-

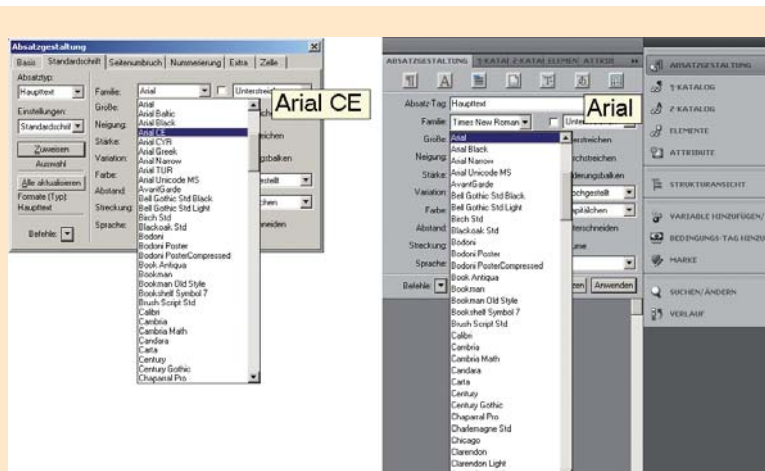


Abb. 1: Version 8 (l.) zeigt trotz Unicode für das Dokument weiterhin Arial CE an; in Version 9 (r.) werden virtuelle Fonts nicht mehr zur Auswahl angeboten.

konfiguration angepasst oder ein spezielles FrameScript genutzt werden.

### Konvertieren von 7.2 über 8 nach 9

Angenommen, ein in FrameMaker 7.2 erstelltes Dokument mit tschechischem Text wird mit einer CE-Schrift formatiert, um alle Zeichen mit den richtigen Glyphen anzuzeigen. Als CE-Schrift wird in FrameMaker 7.2 oft der virtuelle Font Arial CE genutzt. Der Text im Dokument ist ANSI-kodiert, die Schrift nutzt die Codepage 1250. Westliche Sprachen wie Deutsch

oder Französisch nutzen die Codepage 1252. Wird dieses FM 7.2-Dokument mit FM 8 geöffnet, dann konvertiert das Programm die Datei in das neue Datenformat und stellt die Codierung der Zeichen innerhalb von Absätzen automatisch auf Unicode um. Die Schrift, die den Absätzen zugewiesen wurde, zeigt jedoch unverändert in den Formaten den Font Arial CE an. Eine Änderung der Formatierung auf den Unicode-Font Arial verändert jedoch weder die Darstellung noch die Kodierung der Zeichen. Um zu gewährleisten, dass Zeichen aus Indexmarken, Variablen und Querverweis-Formaten nicht-westlicher Kodierung auch in Unicode umgewandelt werden, ist es empfehlenswert, vor der Konvertierung von 7.2 nach 8 zum Beispiel Indexmarken als speziell gekennzeichneten Text in den Textfluss einzufügen und den Text nach der Umwandlung wieder in Indexmarken einzufügen. Um dies nicht manuell tun zu müssen, kann ein FrameScript eingesetzt werden. Ist der Text eines 7.2-Dokuments nicht mit einem virtuellen Font formatiert, wird der Text nicht in Unicode gewandelt. Um die Konvertierung sicherzustellen, stellen Sie vor der Konvertierung alle Formatierungen auf einen virtuellen Font um. Dazu können Sie den Befehl zur globalen Aktualisierung aller Einträge des Absatz- und Zeichenformatkatalogs anwenden. Wird dasselbe FM 7.2-Dokument mit FM 9 geöffnet, wird der Inhalt

nicht nach Unicode konvertiert und Arial CE als fehlender Font gemeldet. Deshalb ist es notwendig FM 7.2-Dokumente zuerst in FM 8 zu speichern und dann erst mit der Version 9 zu öffnen.

### Auswirkungen von Versionen

Projekte, die in mehr als nur westeuropäische Sprachen übersetzt werden sollen, können auf Basis verschiedener Codepages mit jeder Programmversion ab 7.2 realisiert werden. In den Versionen bis 7.1 kann das Programm einige osteuropäische und kyrillische Zeichen nicht darstellen, da ihre Codierung von FrameMaker anders benutzt wird. Das ist unter anderem eine Auswirkung der FrameRoman-Codierung. Bis zur Version 7.2 wird zum effektiven Nachlayouten aller übersetzten Dokumente für jede Sprache entweder eine manuelle Umstellung auf die passenden Fonts und Sprachen benötigt oder einmal gestaltete Templates genutzt, aus denen die sprach- und fontspezifischen Formate importiert werden können. Wörterbücher und Silbentrennung sind jedoch nur für westliche Sprachen verfügbar, so dass es empfehlenswert ist, auf die Silbentrennung zu verzichten und die Rechtschreibkontrolle bereits im Übersetzungsprozess abzuschließen.

Mit der Nutzung der Unicode-Codierung ab FM 8 und entsprechenden WGL4-Fonts sind alle EU-Sprachen/Schriften in den Dokumenten integriert. Eine Umstellung der Formate in den Dokumenten muss nur noch für die aktuelle Sprache erfolgen.

### Bearbeiten von Dokumenten

Wegen der grundlegend verschiedenen Zeichen-Codierung empfiehlt es sich, in einem Projekt nicht abwechselnd mit den Versionen 7.2 und 8 zu arbeiten. Im Prinzip kann ein FM 8-Dokument mit westeuropäischer Sprache über das Sichern des Dokuments im MIF-Format zur Version 7.2 konvertiert werden. Je-

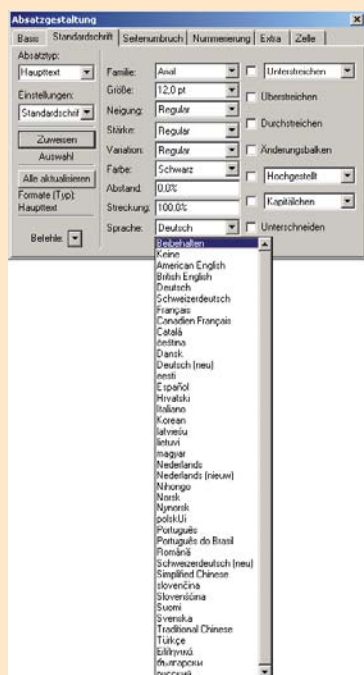


Abb. 2: Die Sprachauswahl in Version 8

doch muss auch dabei penibel auf die richtige MIF 7-Version geachtet werden, da das MIF 8-Format auch Unicode-basiert ist.

Vom wechselseitigen Bearbeiten zwischen Version 7.2 und Frame-Maker 9 ist grundsätzlich abzura-  
ten.

### Fazit

Seit Version 8 kann das Programm sowohl im Übersetzungs-Work-flow als auch beim Export in ge-tagte Textformate (HTML/XML) mit Unicode und ausreichend ausgestatteten OpenType(WGL4)-Fonts gut und ohne aufwendige,

nachträgliche Bearbeitung eingesetzt werden. Vorausgesetzt, das Layout der Dokumente ist im Hinblick auf den Textfluss robust eingerichtet.

#### Links

[www.i-frame.itl.info/framescript\\_download.htm](http://www.i-frame.itl.info/framescript_download.htm)

[www.i-frame.itl.info/i-frame-modules.html](http://www.i-frame.itl.info/i-frame-modules.html)

*Ute Mitschke*  
itl AG  
[ute.mitschke@itl.eu](mailto:ute.mitschke@itl.eu)

## Praxistipps Word: Einmal Word und zurück

*Eine der zentralen Einsatzmöglichkeiten von Word ab Version 2003 und seiner XML-Fähigkeit ist das automatisierte Erzeugen von Word-Dokumenten aus XML-Quellen mit Hilfe von XSLT. Wie läuft dieser Vorgang ab und welche Vorteile für die Technische Kommunikation sind damit verbunden?*

In der Praxis wird die Automatisierung von Word nach XML und wieder zurück dazu verwendet, um XML-Daten für Kunden aufzubereiten und als Word-Dokumente zur Verfügung zu stellen. Oder aus dem Grund, um mit anderen Technischen Redakteuren und Dienstleistern zu kommunizieren, die beispielsweise nicht über einen XML-Editor verfügen oder für die sich der Ein-  
arbeitungsaufwand nicht lohnen würde.

Wie bereits in Ausgabe 6/2008 der ‚technischen kommunikation‘ dargestellt, handelt es sich bei WordML, dem Speicherformat von Word 2003 und Word 2007, um ein XML-Format [1]. Betrachten wir ein kurzes Beispiel für einen Absatz:

```
<w:p>
  <w:r>
    <w:t>Als Gregor
    Samsa eines Morgens
    aus unruhigen Träumen
    erwachte, fand er sich
    in seinem Bett zu einem
    ungeheueren Ungeziefer
    verwandelt.</w:t>
  </w:r>
</w:p>
```

WordML verwendet hier das Element <w:p>, um einen Absatz (Paragraph) zu definieren. Innerhalb von <w:r> (Run-Text) können verschiedene einzeilige Formatierungen und Eigenschaften angelegt werden. Außerdem enthält das Element <w:r> das Element <w:t> (Text) mit dem Text des Absatzes.