The experimental unicode 'EUx' encodings

Will Robertson

2006/08/22   v0.1a

Abstract

This file documents some experimental unicode font encodings, initially used for XƎTEX. So far, there's not much here.

Contents

1 Introduction

LATEX's font encoding system provides a method for ensuring that characters in a document can be rendered properly by glyphs in the output. It is also a sensible system for typesetting multilingual documents while remaining within TEX's restrictions of 8-bit fonts.

When XƎTEX came along, no work had seemingly been done on using this infrastructure for unicode fonts. For one thing, the notion of a well-defined font encoding — which neatly pigeonholes a font into a category of containing a specific set of glyphs — doesn't work so well when designers have an essentially unrestricted selection of glyphs they may add to their fonts.

Both technical advantages of LATEX's font encodings — error checking, and multilingual typesetting — were no longer relevant with XƎTEX, due to impracticality for the former and redundancy in the latter; XƎTEX's support of OpenType fonts obviates the need for font switching by necessity when changing languages.

Nonetheless, font encodings are tied inextricably into LATEX's font selection mechanisms. For the first couple of years of XƎTEX's life, defined font families were placed in the U encoding; appropriate, given the unicode initial, but *inappropriate* within the framework of LATEX itself. The U encoding stands for 'undefined', to

be used for fonts that contain glyphs that are essentially arbitrary; for example, a Dingbats font.

Since X_ƎLATEX documents didn't interact with the 'regular' world so much, this overloading of the encoding didn't really matter in the long run. But eventually, it became apparent that it would be appropriate to use font encodings to represent unicode subsets and/or different methods of font loading (to be addressed later).

This document contains the first steps in this direction.

2 The `EUx` encodings

As with much of my X_ƎTEX work, nothing can be relied upon to last unchanged into the future. This is an unfortunate fact that extensions to TEX that deal with unicode (>16 bit fonts) are still many and varied (by this time next year, there will be at least four[1]). X_ƎTEX is the only one with active macro development at the present time to the best of my knowledge. It thus serves as a testbed for new ideas; furthermore, I'm learning all of this stuff too so it's possible that I'll make conceptual mistakes that require changes later on.

3 The `EU1` encoding

The EU1 encoding represents the first attempt of providing a standard LATEX encoding for unicode fonts. At present, it loads and defines nothing in terms of declaring text glyphs and accents. That rôle is played for now by Ross Moore's `xunicode` package; in the future I hope some of that package will migrate here.

Because we can assume that all current TEX distributions will contain the Latin Modern OpenType fonts (`fonts/opentype/public/lm`), this is selected as the default font. More details on this font later.

I am now defining the convention that EU1 encoded fonts have no necessary subsets of unicode. Perhaps in the future this will be restricted to fonts that support unicode accent characters.

Furthermore, no restrictions are made on the method used to load the font. X_ƎTEX allows fonts of two types: system installed fonts (also accessible via regular applications or programs); and 'local' fonts, which are specified by filename to be found either in a `texmf` tree of the local directory.

```
1 \ProvidesFile{eu1enc.def}[2006/08/13 v0.1a Experimental unicode font encod-
  ing]
2 \DeclareFontEncoding{EU1}{}{}
3 \DeclareErrorFont{EU1}{lmr}{m}{n}{10}
4 \DeclareFontSubstitution{EU1}{lmr}{m}{n}
```

---

[1] X_ƎTEX, LuaTEX, Omega/Aleph, "χ TEX, if I've counted correctly

4 The Latin Modern OpenType fonts

4.1 Conversion scripts

The .fd files are converted from the T1 encoding with the following shell script. In order to generate the Latin Modern EU1 encoding font definitions, execute the following commands:

```
latex euenc.ins
sh convert-lmfd.sh
```

The sources that allow this generation are shown following.

4.1.1 Encoding conversion shell script

```
5  sed -E -f sed-eu1lmr.sed `kpsewhich t1lmr.fd` > eu1lmr.fd
6  sed -E -f sed-eu1lmdh.sed `kpsewhich t1lmdh.fd` > eu1lmdh.fd
7  sed -E -f sed-eu1lmss.sed `kpsewhich t1lmss.fd` > eu1lmss.fd
8  sed -E -f sed-eu1lmssq.sed `kpsewhich t1lmssq.fd` > eu1lmssq.fd
9  sed -E -f sed-eu1lmtt.sed `kpsewhich t1lmtt.fd` > eu1lmtt.fd
10 sed -E -f sed-eu1lmvtt.sed `kpsewhich t1lmvtt.fd` > eu1lmvtt.fd
```

4.1.2 Encoding conversion sed scripts

The shell script above clearly requires the following quickly assembled sed scripts. No doubt they could be reduced to a single script with enough motivation. For now, they work.

Roman

```
11 s/ec-//g
12 s/lm([a-z]+)([0-9]+)/lmroman\2-\1/g
13 s/([0-9]+)-ri/\1-italic/g
14 s/([0-9]+)-ro/\1-oblique/g
15 s/([0-9]+)-r/\1-regular/g
16 s/([0-9]+)-bxi/\1-bolditalic/g
17 s/([0-9]+)-bxo/\1-boldoblique/g
18 s/([0-9]+)-bx/\1-bold/g
19 s/([0-9]+)-bo\}/\1-demioblique\}/g
20 s/([0-9]+)-b\}/\1-demi\}/g
21 s/([0-9]+)-csco/\1-capsoblique/g
22 s/([0-9]+)-csc/\1-capsregular/g
23 s/t1/eu1/g
24 s/T1/EU1/g
25 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

Dunhill

```
26 s/ec-//g
27 s/lm([a-z]+)([0-9]+)/lmroman\2-\1/g
28 s/([0-9]+)-dunh/\1-dunhill/g
29 s/([0-9]+)-duno/\1-dunhilloblique/g
30 s/t1/eu1/g
31 s/T1/EU1/g
32 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

Sans

```
33 s/ec-//g
34 s/lm([a-z]+)([0-9]+)/lmsans\2-\1/g
35 s/([0-9]+)-ssbo/\1-boldoblique/g
36 s/([0-9]+)-ssbx/\1-bold/g
37 s/([0-9]+)-ssdc\}/\1-demicondensed\}/g
38 s/([0-9]+)-ssdo\}/\1-demicondensedoblique\}/g
39 s/([0-9]+)-ssi/\1-italic/g
40 s/([0-9]+)-sso/\1-oblique/g
41 s/([0-9]+)-ss/\1-regular/g
42 s/t1/eu1/g
43 s/T1/EU1/g
44 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

Sans Quotation

```
45 s/ec-//g
46 s/lm([a-z]+)([0-9]+)/lmsansquotation\2-\1/g
47 s/([0-9]+)-ssqbo/\1-boldoblique/g
48 s/([0-9]+)-ssqbx/\1-bold/g
49 s/([0-9]+)-ssqo/\1-oblique/g
50 s/([0-9]+)-ssq/\1-regular/g
51 s/t1/eu1/g
52 s/T1/EU1/g
53 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

Typewriter

```
54 s/ec-//g
55 s/lm([a-z]+)([0-9]+)/lmtypewriter\2-\1/g
56 s/([0-9]+)-tl\}/\1-light\}/g
57 s/([0-9]+)-tlo\}/\1-lightoblique\}/g
58 s/([0-9]+)-tk\}/\1-dark\}/g
59 s/([0-9]+)-tko\}/\1-darkoblique\}/g
60 s/([0-9]+)-tlc\}/\1-lightcondensed\}/g
61 s/([0-9]+)-tlco\}/\1-lightcondensedoblique\}/g
62 s/([0-9]+)-tcso/\1-capsoblique/g
63 s/([0-9]+)-tcsc/\1-capsregular/g
```

```
64 s/([0-9]+)-tti/\1-italic/g
65 s/([0-9]+)-tto/\1-oblique/g
66 s/([0-9]+)-tt/\1-regular/g
67 s/t1/eu1/g
68 s/T1/EU1/g
69 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

Variable-width Typewriter

```
70 s/ec-//g
71 s/lm([a-z]+)([0-9]+)/lmtypewritervarwd\2-\1/g
72 s/([0-9]+)-vtl\}/\1-light\}/g
73 s/([0-9]+)-vtlo\}/\1-lightoblique\}/g
74 s/([0-9]+)-vtk\}/\1-dark\}/g
75 s/([0-9]+)-vtko\}/\1-darkoblique\}/g
76 s/([0-9]+)-vtto/\1-oblique/g
77 s/([0-9]+)-vtt/\1-regular/g
78 s/t1/eu1/g
79 s/T1/EU1/g
80 s/(>.[ ]*)(lm[a-z0-9\-]*)/\1"[\2]"/g
```

4.2   Latin Modern test file

Adapted and improved from my PracTeX Journal article on the Latin Modern fonts.

```
81 \documentclass[12pt]{article}
82 \usepackage[margin=2.5cm]{geometry}
83 \newenvironment{vrb}
84            {\begin{tabular}{@{}p{7cm}l@{}}}
85            {\end{tabular}}
86
87 \usepackage{fancyvrb}
88 \DefineShortVerb{\|}
89
90 \usepackage{slantsc}
91 \usepackage[EU1]{fontenc}
92 \usepackage{lmodern}
93 \begin{document}
94
95 \title{Testing the Latin Modern \texttt{EU1} encoding}
96 \author{Will Robertson}
97 \maketitle
98 \thispagestyle{empty}
99 \pagestyle{empty}
100
101 \section*{Roman}
102
```

5

```
103 \fontfamily{lmr}\selectfont
104 \begin{vrb}
105 |\rmdefault| & {\selectfont Latin Modern Roman} \\
106 |   \itshape| & {\itshape Latin Modern Roman Italic} \\
107 |   \slshape| & {\slshape Latin Modern Roman Oblique} \\
108 |   \scshape| & {\scshape Latin Modern Roman Small Caps} \\
109 |    \slshape| & {\scshape\slshape Latin Modern Roman Oblique Small Caps} \\
110 |   \bfseries| & {\bfseries Latin Modern Roman Bold Extended} \\
111 |       \itshape| & {\bfseries\itshape Latin Modern Roman Bold Italic Ex-
    tended} \\
112 |       \slshape| & {\bfseries\slshape Latin Modern Roman Bold Oblique Ex-
    tended} \\
113 |   \fontseries{b}| & {\fontseries{b}\selectfont Latin Modern Roman Bold} \\
114 |      \fontseries{b}\slshape| & {\fontseries{b}\slshape Latin Modern Ro-
    man Bold Oblique} \\
115 \end{vrb}
116
117 \section*{Dunhill}
118
119 \fontfamily{lmdh}\selectfont
120 \begin{vrb}
121 |\fontfamily{lmdh}\selectfont| & {Latin Modern Dunhill} \\
122 |   \slshape| & {\slshape Latin Modern Dunhill Oblique} \\
123 \end{vrb}
124
125 \section*{Sans}
126
127 \fontfamily{lmss}\selectfont
128 \begin{vrb}
129 |\sffamily| & {\selectfont Latin Modern Sans} \\
130 |   \slshape| & {\slshape Latin Modern Sans Oblique} \\
131 |   \bfseries| & {\bfseries Latin Modern Sans Bold} \\
132 |     \slshape| & {\bfseries\slshape Latin Modern Sans Bold Oblique} \\
133 |   \fontseries{sbc}\selectfont| & {\fontseries{sbc}\selectfont Latin Mod-
    ern Sans Demi Condensed} \\
134 |          \slshape| & {\fontseries{sbc}\selectfont\slshape Latin Mod-
    ern Sans Demi Condensed Oblique} \\
135 \end{vrb}
136
137 \section*{Sans Quotation}
138
139 \fontfamily{lmssq}\selectfont
140 \begin{vrb}
141 |\renewcommand\sfdefault{lmssq}| \\
142 |\sffamily      | & {\selectfont Latin Modern Sans Extended} \\
143 |   \slshape    | & {\slshape Latin Modern Sans Extended Oblique} \\
144 |   \bfseries   | & {\bfseries Latin Modern Sans Bold Extended} \\
```

```
145 |            \slshape | & {\bfseries\slshape Latin Modern Sans Bold Ex-
  tended Oblique} \\
146 \end{vrb}
147
148 \section*{Typewriter}
149
150 \fontfamily{lmtt}\selectfont
151 \begin{vrb}
152 |\ttfamily        | & {\selectfont Latin Modern Typewriter} \\
153 |   \itshape      | & {\itshape Latin Modern Typewriter Italic} \\
154 |   \slshape      | & {\slshape Latin Modern Typewriter Oblique} \\
155 |   \scshape      | & {\scshape Latin Modern Typewriter Small Caps} \\
156 |  \scshape\slshape | & {\scshape\slshape Latin Modern Typewriter Oblique Small Caps} \\
157 |   \bfseries     | & {\bfseries Latin Modern Typewriter Dark} \\
158 |    \slshape | & {\bfseries\slshape Latin Modern Typewriter Dark Oblique} \\
159 |  \fontseries{b}\selectfont | & {\fontfamily{lmtt}\fontseries{b}\selectfont Latin Mod-
  ern Typewriter Dark} \\
160 |  \fontseries{b}\slshape   | & {\fontfamily{lmtt}\fontseries{b}\slshape Latin Mod-
  ern Typewriter Dark Oblique} \\
161 |  \fontseries{l}\selectfont | & {\fontfamily{lmtt}\fontseries{l}\selectfont Latin Mod-
  ern Typewriter Light} \\
162 |  \fontseries{l}\slshape   | & {\fontfamily{lmtt}\fontseries{l}\slshape Latin Mod-
  ern Typewriter Light Oblique} \\
163 |  \fontseries{lc}\selectfont | & {\fontfamily{lmtt}\fontseries{lc}\selectfont Latin Mod-
  ern Typewriter Light Condensed} \\
164 |  \fontseries{lc}\slshape   | & {\fontfamily{lmtt}\fontseries{lc}\slshape Latin Mod-
  ern Typewriter Light Condensed Oblique} \\
165 \end{vrb}
166
167 \section*{Variable-width Typewriter}
168
169 \fontfamily{lmvtt}\selectfont
170 \def\fixedspacing{\fontdimen3\font0pt\fontdimen4\font0pt}
171 \begin{vrb}
172 |\renewcommand\ttdefault{lmvtt}| \\
173 |\ttfamily                  | & {\selectfont \fixedspacing Latin Mod-
  ern Typewriter Proportional} \\
174 |   \slshape                | & {\slshape \fixedspacing Latin Modern Type-
  writer Proportional Oblique} \\
175 |   \fontseries{l}\selectfont | & {\fontseries{l}\selectfont \fixedspac-
  ing Latin Modern Typewriter Proportional Light} \\
176 |   \fontseries{l}\slshape   | & \mbox{\fontseries{l}\slshape \fixedspac-
  ing Latin Modern Typewriter Proportional Light Oblique} \\
177 |    \fontseries{b}\selectfont | & {\fontseries{b}\selectfont \fixedspac-
  ing Latin Modern Typewriter Proportional Dark} \\
178 |   \fontseries{b}\slshape   | & \mbox{\fontseries{b}\slshape \fixedspac-
  ing Latin Modern Typewriter Proportional Dark Oblique} \\
```

```
179 \end{vrb}
180
181 \end{document}
```