The Experimental Unicode ‘EUx’ Encodings

Will Robertson    Elie Roux    Khaled Hosny

2010/05/27       v0.1h

Abstract
This file documents some experimental Unicode font encodings, used for \TeX{} and \LaTeX{}. Includes Latin Modern font definitions for version 1.6 and later of those fonts.

Contents

1 Introduction 1
2 The EU1 encoding 2
3 The EU2 encoding 2
4 The Latin Modern OpenType fonts 3
4.1 Conversion scripts 3
5 Latin Modern test file 5

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 \TeX{} came along, no work had 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 \TeX{}, due to impracticality for the former and redundancy in the latter; \TeX{}’s support of OpenType fonts obviates the need for font switching (for glyph coverage only) by necessity when changing languages, although multiple fonts will often be used for stylistic purposes.

Nonetheless, font encodings are tied inextricably into \LaTeX{}’s font selection mechanisms. For the first couple of years of \TeX{}’s life, defined font families were...
placed in the \U encoding; appropriate, given the Unicode initial, but \textit{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 \Xe\LaTeX{} 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. With the advent of
\Lua\TeX{}, this latter use has justified the approach originally taken here.

\section{The EU1 encoding}

The EU1 encoding represents the first attempt of providing a standard \LaTeX{} en-
coding 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
\texttt{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 (\texttt{fonts/opentype/public/lm}), these are selected as the
default fonts. More details on this 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 sup-
port Unicode accent characters.

Furthermore, no restrictions are made on the method used to load the font.
\Xe\LaTeX{} 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 \texttt{texmf} tree of the local directory.

\begin{verbatim}
1 \DeclareFontEncoding{EU1}{}{}
2 \DeclareErrorFont{EU1}{lmr}{m}{n}{1/zero.noslash}
3 \DeclareFontSubstitution{EU1}{lmr}{m}{n}
\end{verbatim}

The : character needs to be sanitised in case babel makes it active:

\begin{verbatim}
4 \g@addto@macro\nff\@catcodes{\@makeother:}
\end{verbatim}

\section{The EU2 encoding}

The EU2 encoding is to be used with \Lua\TeX{}. It is the same as EU1 for now; the
actual difference is in the font definition files.

\begin{verbatim}
5 \DeclareFontEncoding{EU2}{}{}
6 \DeclareErrorFont{EU2}{lmr}{m}{n}{10}
7 \DeclareFontSubstitution{EU2}{lmr}{m}{n}
\end{verbatim}

The : character needs to be sanitised in case babel makes it active:

\begin{verbatim}
8 \g@addto@macro\nff\@catcodes{\@makeother:}
\end{verbatim}
4 The Latin Modern OpenType fonts

Version 0.1g of euenc supports version 1.6 of the LM fonts, released in October 2009.

4.1 Conversion scripts

The .fd files are converted from the T1 encoding with the following shell script. Only the package maintainers need to be able to run this script and those following.

4.1.1 Encoding conversion shell script

```
9 sed -f sed-eu1lm.sed 'kpsewhich t1lmr.fd' eu1lmr.fd
10 sed -f sed-eu1lm.sed 'kpsewhich t1lmdh.fd' eu1lmdh.fd
11 sed -f sed-eu1lm.sed 'kpsewhich t1lmss.fd' eu1lmss.fd
12 sed -f sed-eu1lm.sed 'kpsewhich t1lmssq.fd' eu1lmssq.fd
13 sed -f sed-eu1lm.sed 'kpsewhich t1lmvtt.fd' eu1lmvtt.fd
14 sed -f sed-eu2lm.sed 'kpsewhich t1lmr.fd' eu2lmr.fd
15 sed -f sed-eu2lm.sed 'kpsewhich t1lmdh.fd' eu2lmdh.fd
16 sed -f sed-eu2lm.sed 'kpsewhich t1lmss.fd' eu2lmss.fd
17 sed -f sed-eu2lm.sed 'kpsewhich t1lmssq.fd' eu2lmssq.fd
18 sed -f sed-eu2lm.sed 'kpsewhich t1lmvtt.fd' eu2lmvtt.fd
19 sed -f sed-eu1lm.sed -e 's/:mapping=tex-text//g' 'kpsewhich t1lmtt.fd' eu1lmtt.fd
20 sed -f sed-eu2lm.sed -e 's/;.trep;.tlig;//g' 'kpsewhich t1lmtt.fd' eu2lmtt.fd
```

4.1.2 Encoding conversion sed scripts

The shell script above requires the following sed scripts.

**Sed scripts**

```
25 /\%.*d
26 <eu1sedlm>s/{T1}/{EU1}/g
27 <eu1sedlm>s/t1lm/eu1lm/g
28 <eu2sedlm>s/{T1}/{EU2}/g
29 <eu2sedlm>s/t1lm/eu2lm/g
30 s/ec-1mri\([0-9]+\)/lmroman\1-italic/g
31 s/ec-1mro\([0-9]+\)/lmromantlant\1-regular/g
32 s/ec-lmr\([0-9]+\)/lmroman\1-regular/g
33 s/ec-lmcso\([0-9]+\)/lmromancaps\1-oblique/g
34 s/ec-lmcsc\([0-9]+\)/lmromancaps\1-regular/g
35 s/ec-lmcsc\([0-9]+\)/lmromancaps\1-regular/g
36
```
s/ec-lmbxo\([0-9]*\)/lmroman\1-bold/g
s/ec-lmbxi\([0-9]*\)/lmroman\1-bolditalic/g
s/ec-lmb\([0-9]*\)/lmroman\1-bold/g
s/ec-lmbo\([0-9]*\)/lmromandemi\1-oblique/g
s/ec-lmb\([0-9]*\)/lmromandemi\1-regular/g
s/ec-lmdunh\([0-9]*\)/lmromandunh\1-regular/g
s/ec-lmduno\([0-9]*\)/lmromandunh\1-oblique/g
s/ec-lmss\([0-9]\{1,\}\)/lmsans\1-regular/g
s/ec-lmssso\([0-9]\{1,\}\)/lmsans\1-oblique/g
s/ec-lmssdc\([0-9]\{1,\}\)/lmsansdemicond\1-regular/g
s/ec-lmssdo\([0-9]\{1,\}\)/lmsansdemicond\1-oblique/g
s/ec-lmssbx\([0-9]\{1,\}\)/lmsans\1-bold/g
s/ec-lmssbo\([0-9]\{1,\}\)/lmsans\1-boldoblique/g
s/ec-lmssq\([0-9]\{1,\}\)/lmsansquot\1-regular/g
s/ec-lmssqo\([0-9]\{1,\}\)/lmsansquot\1-oblique/g
s/ec-lmssqbx\([0-9]\{1,\}\)/lmsansquot\1-bold/g
s/ec-lmssqbo\([0-9]\{1,\}\)/lmsansquot\1-boldoblique/g
s/ec-lmtt\([0-9]\{1,\}\)/lmmono\1-regular/g
s/ec-lmtto\([0-9]\{1,\}\)/lmmonoslant\1-regular/g
s/ec-lmtti\([0-9]\{1,\}\)/lmmono\1-italic/g
s/ec-lmtcsc\([0-9]\{1,\}\)/lmmonocaps\1-regular/g
s/ec-lmtcso\([0-9]\{1,\}\)/lmmonocaps\1-oblique/g
s/ec-lmtl\([0-9]\{1,\}\)/lmmonolt\1-regular/g
s/ec-lmtlo\([0-9]\{1,\}\)/lmmonolt\1-oblique/g
s/ec-lmtk\([0-9]\{1,\}\)/lmmonolt\1-bold/g
s/ec-lmtko\([0-9]\{1,\}\)/lmmonolt\1-boldoblique/g
s/ec-lmtlc\([0-9]\{1,\}\)/lmmonoltcond\1-regular/g
s/ec-lmtlco\([0-9]\{1,\}\)/lmmonoltcond\1-oblique/g
s/ec-lmvtt\([0-9]\{1,\}\)/lmmonoprop\1-regular/g
s/ec-lmvto\([0-9]\{1,\}\)/lmmonoprop\1-oblique/g
s/\([\w\-\ ]*\)lm[a-z0-9-]\{1,\}\)/\2:\mapping=tex-text"/g

5 Latin Modern test file

Adapted and improved from my PracT\TeX\ Journal article on the Latin Modern fonts.

\documentclass[12pt]{article}
\usepackage[margin=2.5cm]{geometry}
\newenvironment{vrb}{\begin{tabular}{@{}p{7cm}l@{}}}{\end{tabular}}
\usepackage{fancyvrb}
\DefineShortVerb{\|}
\usepackage{slantsc}
\usepackage{ifluatex}
\usepackage{ifxetex}
\ifxetex
\def\EU{EU1}
\else
\ifluatex
\usepackage{luaotfload}
\def\EU{EU2}
\else
\def\EU{T1}
\fi
\fi
\usepackage[\EU]{fontenc}
\usepackage{lmodern}
\begin{document}
\section*{Roman}
\fontfamily{lmr}\selectfont
|\rmdefault| & \textit{Latin Modern Roman} \\
|\itshape| & \textit{Latin Modern Roman Italic} \\
|\slshape| & \textit{Latin Modern Roman Oblique} \\
|\scshape| & \textit{Latin Modern Roman Small Caps} \\
\end{tabular}
125 | \slshape| & \{scshape\slshape Latin Modern Roman Oblique Small Caps} \\
126 | \bfseries| & \{bfseries Latin Modern Roman Bold Extended} \\
127 | \itshape| & \{bfseries\itshape Latin Modern Roman Bold Italic Extended} \\
128 | \slshape| & \{bfseries\slshape Latin Modern Roman Bold Oblique Extended} \\
129 | \fontseries{b}| & \{\fontseries{b}\selectfont Latin Modern Roman Bold} \\
130 | \fontseries{b}\slshape| & \{\fontseries{b}\slshape Latin Modern Roman Bold Oblique} \\
131 | \end{verb} \\
132 \section*{Dunhill} \\
133 \fontfamily{lmdh}\selectfont \\
134 \begin{verb} \\
135 |\fontfamily{lmdh}\selectfont| & \selectfont Latin Modern Dunhill \\\n136 |\slshape| & \slshape Latin Modern Dunhill Oblique \\\n137 \end{verb} \\
138 \section*{Sans} \\
139 \fontfamily{lmss}\selectfont \\
140 \begin{verb} \\
141 |\ttfamily| & \selectfont Latin Modern Typewriter \\\n142 |\itshape| & \itshape Latin Modern Typewriter Italic \\\n143 |\slshape| & \slshape Latin Modern Typewriter Oblique \\\n144 |\scshape| & \scshape Latin Modern Typewriter Small Caps \\\n145 \end{verb}
\section*{Variable-width Typewriter}

\begin{verbatim}
\ttfamily \selectfont \fixedspacing Latin Modern Typewriter Proportional
\slshape \fixedspacing Latin Modern Typewriter Proportional Oblique
\fontseries{l} \selectfont \fixedspacing Latin Modern Typewriter Proportional Light
\fontseries{l} \slshape \fixedspacing Latin Modern Typewriter Proportional Light Oblique
\fontseries{b} \selectfont \fixedspacing Latin Modern Typewriter Proportional Dark
\fontseries{b} \slshape \fixedspacing Latin Modern Typewriter Proportional Dark Oblique
\end{verbatim}

\renewcommand{\ttdefault}{lmvtt}
\ttfamily
\selectfont \fixedspacing Latin Modern Typewriter Proportional
\slshape \fixedspacing Latin Modern Typewriter Proportional Oblique
\fontseries{l} \selectfont \fixedspacing Latin Modern Typewriter Proportional Light
\fontseries{l} \slshape \fixedspacing Latin Modern Typewriter Proportional Light Oblique
\fontseries{b} \selectfont \fixedspacing Latin Modern Typewriter Proportional Dark
\fontseries{b} \slshape \fixedspacing Latin Modern Typewriter Proportional Dark Oblique
\end{verbatim}

\subsection*{Variable-width Typewriter}

\begin{verbatim}
\ttfamily \selectfont \fixedspacing Latin Modern Typewriter Proportional
\slshape \fixedspacing Latin Modern Typewriter Proportional Oblique
\fontseries{l} \selectfont \fixedspacing Latin Modern Typewriter Proportional Light
\fontseries{l} \slshape \fixedspacing Latin Modern Typewriter Proportional Light Oblique
\fontseries{b} \selectfont \fixedspacing Latin Modern Typewriter Proportional Dark
\fontseries{b} \slshape \fixedspacing Latin Modern Typewriter Proportional Dark Oblique
\end{verbatim}

\end{document}