

SHORT INTRODUCTION TO XETEX

RICHARD KOCH

1. SOME T_EX HISTORY

T_EX was invented before several of the important computing developments of the twentieth century. In 1978 there was no Macintosh and no generally available computer with a graphical interface. Personal computers printed output using a single bitmapped font in which all characters had the same width; some computers only printed upper case letters.

Consequently, Knuth had to invent outline fonts at the same time that he invented T_EX. In his documentation, Knuth always talks about *two* programs, the typesetting program T_EX and the outline font program METAFONT.

Later in the decade, other players introduced their own outline fonts using different formats: Adobe's Type 1 Postscript fonts, Apple's TrueType fonts, and Microsoft's OpenType fonts. The history of these fonts is rather complicated; suffice it to say that each is now supported on the Mac and available on other operating systems as well. A vast number of beautiful fonts are available in these new formats.

Adapting these fonts for use in T_EX is a complicated task, which has only been successfully completed in a small number of cases. One of the most important complaints about T_EX is the limited number of available fonts.

2. X_ƎL^AT_EX AND FONTS

X_ƎL^AT_EX completely solves the font problem. It can use Knuth's fonts, of course, but it can mix these with system fonts in Adobe Type 1, TrueType, and OpenType formats. These fonts need not be adapted for use in T_EX; instead they are immediately available.

3. MORE HISTORY: UNICODE

T_EX source manuscripts contain standard characters available on any typewriter. Indeed, originally T_EX could only accept the 128 standard ASCII input characters. Later this was expanded to 256 characters. This is enough for standard English and for Western European Languages, but it is certainly not enough for Japanese, Chinese, and a multitude of other languages. Over the years, many projects have attempted to adopt T_EX so it can typeset these other languages.

Later the same problem arose in the entire computer industry, since a large number of sales occur in countries which do not use Western scripts. The solution invented by the industry is Unicode, an open standard which can theoretically encode the characters of all languages on earth. Most computer manufacturers have adopted this standard. For example, the edit class in Cocoa (and consequently TeXShop's editor) accepts arbitrary Unicode characters.

4. X_QL^AT_EX AND UNICODE

X_QL^AT_EX modifies T_EX to accept any Unicode character; source documents are saved in UTF-8 Unicode format. In conjunction with full support for OpenType fonts, this makes it possible to write T_EX documents in virtually any language on earth.

5. HOW DO I USE X_QL^AT_EX?

Simple. Take any standard L^AT_EX document. Somewhere close to the top, say within the first ten lines, include the lines

```
%!TEX TS-program = xelatex
%!TEX encoding = UTF-8 Unicode
```

These lines begin with the T_EX comment character, so T_EX will ignore them. However, TeXShop and TeXworks recognize the lines and do something special when you open, save, or typeset the file. The first line tells these programs to call `xelatex` rather than `pdflatex` when you ask it to typeset the document. The second line tells the programs to load and save the file with **UTF-8 Unicode** encoding, rather than the standard **Mac OS Roman** encoding.

Since X_QL^AT_EX supports standard T_EX fonts and standard T_EX commands, typesetting will proceed without change.

6. HOW DO I SWITCH TO ALTERNATE FONTS IN X_QL^AT_EX?

A style package written by Will Robertson makes this task simple. Add the following lines to the T_EX preamble:

```
\usepackage{fontspec,xltxtra,xunicode}
\defaultfontfeatures{Mapping=tex-text}
\setromanfont{Hoefler Text}
```

```
\setsansfont[Scale=MatchLowercase]
           {Gill Sans}
\setmonofont[Scale=MatchLowercase]
           {Andale Mono}
```

Change the three font names to any system fonts you'd like to use. The first font is the default serif font, the second is the default sans serif font, and the third is the default monospaced font for typewriter-like output. If your entire document is in a serif font, it is enough to change the first font.

To find appropriate font names, open Apple's Font Book application, look at the font samples, and choose one of the font names listed by the application.

7. HOW DO I INSERT UNICODE IN THE SOURCE MANUSCRIPT?

Go to Apple's System Preferences and choose the International panel. Select the Input Menu tab. A large number of keyboards layouts are provided, as well as more complicated input methods for Chinese and Japanese.

Choose, for example, Arabic, Nepali, and Russian. A flag icon will appear at the right side of the menu bar. Open a blank TeXShop or TeXworks document. Start with a US or Western European flag and type some standard text.

Then select the Russian flag and type some more. Notice that you are inputting Cyrillic characters. Select Nepali and type another line, noticing the distinct characters.

Finally select Arabic and type again. Arabic is written from right to left; notice that the characters you type in the editor appear from right to left. In Arabic, character shapes change when the character ends a

word, Notice that previous characters tend to change shape as you add additional characters.

None of this is very useful unless you know one of these languages!

8. HOW DO I SEE THESE CHARACTERS IN THE T_EX OUTPUT?

Ah, that is the tricky step. Some Macintosh fonts appear to contain only standard Roman characters, but have been extended to contain additional Unicode characters. You must typeset in a font which contains characters for the language you have selected.

For example, the Geeza Pro font contains Arabic characters, the Lucida Grande font contains Hebrew characters, and the Osaka font contains Japanese characters. Some experimentation may be needed to find a font which contains characters for the language you want to use.

Suppose, for example, that you want to output Arabic. The following line defines a font family which contains Arabic characters:

```
\newfontfamily{\A}{Geeza Pro}
```

To insert Arabic, type the following and insert Arabic characters in the spot currently containing dots:

```
This is Arabic text: {\A ....}
```

as in {\A السلام عليكم}.

Incidentally, this document was typeset with X_ƎL^AT_EX. In ordinary L^AT_EX, the last clause of the previous paragraph would have been difficult to typeset.

9. SAMPLE X_ƎL^AT_EX DOCUMENTS

The remaining pages show sample XeT_EX source manuscripts, and the resulting X_ƎL^AT_EX output. This gives some feel for the advantages of X_ƎL^AT_EX in the general community.