

Zpravodaj Československého sdružení uživatelů TeXu

Jiang Jiang

Chinese TeX Typesetting: Past and Present

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 20 (2010), No. 3, 215–219

Persistent URL: <http://dml.cz/dmlcz/150123>

Terms of use:

© Československé sdružení uživatelů TeXu, 2010

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ*:
The Czech Digital Mathematics Library <http://dml.cz>

Abstrakt

Článek seznamuje čtenáře s přehledem sazby čínštiny v T_EXu od prvopočátků až do současnosti.

Klíčová slova: CCT, TianYuan, CJK, C_T_EX, X_Y_T_EX, xcp.py, zhspacing, xeCJK, ctex, ctex-kit, ctex-doc, sazba čínštiny, Google code.

doi: 10.5300/2010-3/215

Introduction

Knuth developed T_EX as a 7-bit system at the time it was invented, internationalization became a tough issue ever since then, yet CJK (Chinese, Japanese and Korean) typesetting is probably the toughest part. The biggest problem is that Eastern language systems like Chinese has a lot of characters compare to western languages like Latin, English and French.

There are more than 100 thousands of ideograph based characters and more than 10 thousand of which are used on daily basis. It means the typesetting system should be able to support such a large range of characters and use fonts accordingly. Unfortunately under a 7-bit or 8-bit system like Knuth-T_EX, there is no trivial or elegance solution to support that.

CCT and TianYuan

Since the original T_EX system does not provide such a capability, workarounds are invented. Early T_EX adopters in China are mostly mathematicians, in 1990s, Zhang Linbo invented the CCT (<http://lsec.cc.ac.cn/~zlb/>) system while Xiao Gang and Chen Zhijie invented TianYuan system (TY for short; <http://wims.math.ecnu.edu.cn/ty/>).

Both of them are based preprocessors – using a preprocessing tool to convert from the original T_EX source file in raw Chinese to a version that T_EX can recognize, most importantly, the DVI generated in such methods can be further processed to use Chinese fonts. This mechanism was not very easy to use but worked fairly well. However, there are still many limitations:

1. preprocessing based solutions are cumbersome, you need an extra wrapper to the T_EX executable otherwise files will be messed up;

2. font support is very limited, in both systems mentioned above, you can only use the fonts predefined, it is very difficult for them to support arbitrary fonts.

CJK package

Things changed a lot since Werner Lemberg introduced the CJK macro package (<http://cjk.ffii.org/>), its a pure T_EX macro package so you don't need to preprocess the T_EX source. Werner also defined a font system for CJK font selection, based on L^AT_EX font selection scheme. To support fonts that contains thousands of glyphs, this system will need users to convert TTF files into multiple “sub-font” files (since T_EX can support 256 font files at most, each of them can have 256 glyphs, so theoretically we can use 65536 different glyphs in one single T_EX document), each one covers 256 different glyphs, combining this sub-font scheme and a PS/PDF generation driver that supports this scheme, we can correctly generate the resulting PDF (or PostScript). At that moment, DVIPDFM_x (<http://project.ktug.or.kr/dvipdfmx/>) is such an PDF driver.

A typical CJK document looks like this:

```
\documentclass{article}
\usepackage{CJK}
\begin{document}
\begin{CJK}{GBK}{song}
你好，这里是一段中文内容。
\end{CJK}
\end{document}
```

As you can see, the text wrapped between `\begin{CJK}...` and `\end{CJK}` can be CJK characters, but everything outside cannot be. Obviously that is not very convenient since you will like to define macros containing CJK characters in the preamble from time to time.

This method works pretty well and proved to be quite stable, so more and more Chinese typesetters adopted it. Zhang Linbo also converted CCT to use a similar method, compare to CJK, CCT have more features focused on Chinese typesetting, such as white space adjustment, punctuation width adjustment, etc.

C_T_EX

As one memorable mark of T_EX typesetting in China, Wang Lei and Wu Lingyun created a community called “C_T_EX” (<http://www.ctex.org/>), which later be-

came the most popular and most active community for T_EX learning and development in China. C_T_EX was originally a mailing list, but converted to a Web forum later.

With the help of many others, Wu Lingyun packaged all mature Chinese T_EX typesetting solutions with the most popular T_EX distribution on Windows – MiK_T_EX – to an individual distribution called C_T_EX. C_T_EX became the most popular T_EX distribution in China ever since then.

X_qT_EX

However, with the above solutions, installing fonts for T_EX usage is still very much a pain. It requires the users to convert fonts from TTF to TFM and edit mapping file very carefully. Thus, new T_EX users consider font installation a very difficult part and try to avoid it as much as possible.

Something revolutionary happened since the introduction of X_qT_EX created by Jonathan Kew (<http://scripts.sil.org/xetex>): processing Unicode documents became so simple and straightforward, especially for CJK documents: you can use any CJK characters in Unicode, not just limited to certain local character sets like GBK or Big5 or Shift-JIS. Best of all, you can use any system font directly, without doing any clumsy “subfont” conversion or carefully creating mapping files for PS/PDF drivers.

However, X_qT_EX was only available to Mac users when it was first developed, because only a few T_EX users use Mac at that time, it was not widely recognized. Soon after Jonathan Kew ported it to Linux and Akira Kakuto provided the Windows port (through `w32tex` and T_EX Live), X_qT_EX became very popular among Chinese users.

The low-level facility provided by X_qT_EX is enough to typeset regular Chinese documents, but typesetters always want better results: switch fonts automatically between CJK and Latin text, adjust the width of punctuations to achieve better line breaking, better support for vertical layout are some of the common requests.

The `xcp.py` script and `zhspacing` package

Among these request, automatic font switching is especially urgent. At that time, most Latin glyphs from Chinese fonts are so badly designed that it is almost impossible to use. Naturally a typesetter would like to use the high quality Latin fonts for non-Chinese texts.

At first, Jiang Jiang designed a preprocessing script called `xcp.py` (<http://code.google.com/p/xcp/>), it simply wrap all the text in CJK with a macro to switch font. This solution is not very convenient but worked amazingly well for most Chinese documents, yet does not impose much performance penalty.

Since more and more users are requesting this feature, Jonathan Kew and Miyata Shigeru worked together to implement it as a build-in of \XeTeX . Basically, you can assign a “character class” to each Unicode code point with \XeTeXcharclass , then define the command to invoke when switching between two specific class with $\text{\XeTeXinterchartoks}$.

With the sample code provided by Miyata Shigeru, Yin Dian implemented a package called `zhspacing` (<http://code.google.com/p/zhspacing/>), which has complete support to assign and switch fonts automatically for Chinese documents. `zhspacing` is a big step forward compare to `xcp.py`, and it is well documented in both English and Chinese.

xeCJK

Soon after the release of `zhspacing`, Sun Wenchang provided another package called `xeCJK` (<http://ctan.org/pkg/xecjk/>) with exactly the same purpose. `xeCJK` is more like a direct descendant of the `CJK` package by Werner Lemberg, users can port their `CJK` documents into `xeCJK` without much efforts, yet features like `CJKnumber` and `CJKpunct` remains compatible. `xeCJK` soon became the most popular solution for Chinese typesetting.

A typical document in `xeCJK` looks like this:

```
\documentclass{article}
\usepackage{xecjk}
\setmainfont{Palatino}
\setCJKmainfont{SimSun}

\begin{document}
你好，这里是一段中文内容。
\end{document}
```

While it is totally fine using `xeCJK` in your documents directly, it is still a bit inconvenient if you need to include other functions like captions (“Table of Contents”, “Chapter”, “Bibliography”, and so on) in Chinese, punctuations in Chinese, `hyperref` with `CJK` bookmarks, etc. Different packages may have different options, which should be carefully tuned for each \TeX engine and output drivers (\pdfTeX / \XeTeX , `dvipdfmx`, `dvips`).

ctex, ctex-kit and ctex-doc

To simplify this process, Wu Lingyun initiated a higher level package called `ctex` (<http://ctan.org/pkg/ctex/>), which contains several document classes and macro packages designed for Chinese typesetting with \LaTeX .

ctex has multiple backends to support traditional CJK and pdf \TeX based engine, or dvipdfmx based PDF output, while it also supports xeCJK and X \TeX based engine. Fonts are also pre-defined in several schemes like “winfonts” and “adobefonts”. A typical ctex document looks like this:

```
\documentclass{ctexart}
```

```
\begin{document}
```

你好，这里是一段中文内容。

```
\end{document}
```

The most recent effort of \TeX community is a project called ctex-kit (<http://code.google.com/p/ctex-kit/>), we organized all Chinese related \TeX packages into a code repository on Google code, then we package and submit these packages to CTAN regularly. With this totally open source approach, these packages are now maintained in a more active state, and users can retrieve them easily from CTAN or through the package manager of \TeX Live 2009.

Besides ctex-kit, there is another project called ctex-doc (<http://code.google.com/p/ctex-doc/>), which organized all of the translations (lshort-cn for instance) or original documents related to \TeX , \LaTeX , \ConTeXt and graphics tools like METAPOST.

Conclusion

With the collaboration by the whole Chinese \TeX community, we have every reason to expect an even more bright future for Chinese \TeX typesetting.

Summary: Chinese \TeX Typesetting: Past and Present

The article introduces and overviews Chinese \TeX typesetting from its early beginnings to the present day.

Keywords: CCT, TianYuan, CJK, \CTEX , X \TeX , xcp.py, zhspacing, xeCJK, ctex, ctex-kit, ctex-doc, Chinese typesetting, Google code.

*Jjgod Jiang, gzzjgod@gmail.com
6 South Kexueyuan Road
Beijing, 100080, China*