TexLive 2007 Chinese Fonts Installation

xptao

1 Preamble

In TeXLive 2007, there is no need to use the CJK package anymore, XeTeX is good enough to handle Chinese and Open/TrueType fonts. A key flaw of XeTeX, i.e. the font switching problem, is likely to be fixed in the next release.[1]

In the rest of the article, We will describe how to install Chinese Fonts for TexLive 2007 on an `ubuntu`[2] box. We use the blue color to typeset commands you need to enter and the brown color to typeset the contents of files.

2 An Example: Installing the Font `hei`

2.1 The steps

Assume that we have installed TexLive 2007 and `fontforge` to the `ubuntu` system successfully. We will describe the steps to install Chinese font `hei` in the following.

1. Copy the files needed.

```
mkdir ~/font
cd ~/font
cp /media/sda/windows/Fonts/simhei.ttf .
cp /usr/local/texlive/2007/texmf-dist/source/latex/CJK/utils/subfonts/* ~/font/
cp /usr/local/texlive/2007/texmf/fonts/sfd/*.sfd ~/font/
```

2. Generate the fonts.[3]

```
fontforge -script subfonts.pe simhei.ttf hei Unicode.sfd
```

3. Make the `makemap` file.

   (a) create a file named `makemap` like this:

```
for i in *.tfm
do
    cat >> hei.map << EOF
${i%.tfm} ${i%.tfm} < ${i%.tfm}.pfb
EOF
done
```

[1] However, since many old documents were written using CJK/CCT/CTeX packages, one might still want to install these packages and fonts.

[2] It is a distro of linux.

[3] This step could be very slow if the `fontforge` software used is not version 2005.

(b) Run `makemap`

```
chmod +x makemap
./makemap
```

4. Create the `fd` file. Create a file named `c70hei.fd` for the CJK package with the following content:

```
% This is c70hei.fd for CJK package.
% created by Edward G.J. Lee
% modified by Yue Wang
\ProvidesFile{c70hei.fd}
\DeclareFontFamily{C70}{hei}{\hyphenchar \font\m@ne}
\DeclareFontShape{C70}{hei}{m}{n}{<-> CJK * hei}{}
\DeclareFontShape{C70}{hei}{bx}{n}{<-> CJKb * hei}{\CJKbold}
\endinput
```

5. Copy the font files into the `TEXMF` directory.

```
cd ~/.texlive2007/texmf-var
mkdir -p fonts/map/dvips/CJK
mkdir -p fonts/tfm/CJK/hei
mkdir -p fonts/type1/CJK/hei
mkdir -p tex/latex/CJK/UTF8
cp ~/font/hei.map fonts/map/dvips/CJK/
cp ~/font/*.tfm fonts/tfm/CJK/hei
cp ~/font/*.pfb fonts/type1/CJK/hei
cp ~/font/c70hei.fd tex/latex/CJK/UTF8
```

6. Update the system.

```
texhash
updmap --enable Map hei.map
```

7. Create the virtual GBK fonts.

(a) run `uni2sfd.pl`

```
cd ~\font
perl uni2sfd.pl hei UGBK.sfd gbkhei gbk
mkdir ~/.texlive2007/texmf-var/fonts/tfm/CJK/gbkhei
mv gbkhei*.tfm ~/.texlive2007/texmf-var/fonts/tfm/CJK/gbkhei
mkdir ~/.texlive2007/texmf-var/fonts/vf
mv gbkhei*.vf ~/.texlive2007/texmf-var/fonts/vf
mkdir ~/.texlive2007/texmf-var/tex/latex/CJK/GBK
```

(b) create a file named `c19hei.fd` in `~/.texlive2007/texmf-var/tex/latex/CJK/GBK` with the following content:

```
% This is c19hei.fd for CJK package.
% created by Edward G.J. Lee
\ProvidesFile{c19hei.fd}
\DeclareFontFamily{C19}{hei}{\hyphenchar \font\m@ne}
\DeclareFontShape{C19}{hei}{m}{n}{<-> CJK * gbkhei}{}
\DeclareFontShape{C19}{hei}{bx}{n}{<-> CJKb * gbkhei}{\CJKbold}
\endinput
```

2.2 Testing the Installation

Create a file (saved using the `utf-8` encoding) as follows:

```
\documentclass{article}
\usepackage{CJKutf8}
\begin{document}
\begin{CJK}{UTF8}{hei}
你好!
\end{CJK}
\end{document}
```

Create a file (saved using the `GBK` encoding) as follows:

```
\documentclass{article}
\usepackage{CJK}
\begin{document}
\begin{CJK}{GBK}{hei}
你好
\end{CJK}
\end{document}
```

If you can successfully compile the two files, the installation is complete and successful.

3 A General Script

The method described in Section 2 can be easily extended for other fonts (`song`, `kai`, etc.)

We give a `bash` script in the Appendix A which will automate this task.

The script can be invoked as:

```
chmod +x ctexlive.sh
./ctexlive.sh font_source_file font_name
```

For example, running

```
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simhei.ttf hei
```

installs the font `hei`. Running

```
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simsun.ttc song
```

installs the font `song`.

We assume that `/media/winsys/WINDOWS/Fonts/` is the directory where the Windows fonts are located.

The recommend settings is to install the following four Windows fonts for TexLive 2007: `song`, `kai`, `fang`, and `hei`. You can install them by running the following commands.

```
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simsun.ttc song
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simkai.ttf kai
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simfang.ttf fang
./ctexlive.sh /media/winsys/WINDOWS/Fonts/simhei.ttf hei
```

4 A Fast Method

The method described in Section 3 is usually very slow (approximately an hour) if you are not using the older `fontforge 2005` version. To make things worse, it is not easy to install old versions of `fontforge` in `ubuntu`. yulewang [2] gave a fast method to accomplish the same task, which took approximately one minute per font installed. We describe the method below.

First, install `perl`, `ttf2pt1`, and `t1-utils` on `ubuntu`. Then download `mkfont.tar.gz` from [4] and decompress it. It contains five files.

```
Unicode.sfd
UGBK.sfd
uni2sfd.pl
unicode-sample.map
mkfont.sh
```

The first three files are from TexLive and we have used them in preceding two sections. Therefore you can also obtain them by

```
cp /usr/local/texlive/2007/texmf-dist/source/latex/CJK/utils/subfonts/* ~/font/
cp /usr/local/texlive/2007/texmf/fonts/sfd/*.sfd ~/font/
```

`unicode-sample.map` and `mkfont.sh` are the new files here.

- `mkfont.sh` is a script file (see Appendix B). [4]
- The file `unicode-sample.map` is shown in Appendix C.

To use the scripts:

```
./mkfont.sh font_source_file filename fontname
```

For example, running

```
./mkfont.sh simsun.ttf simsun song
```

installs the font `song`.

5 Final Remarks

Suppose you have finished the installation of fonts $F_1, F_2, \ldots, F_n$ (usually they are `song`, `kai`, `fang`, and `hei`), you can use them as follows

```
\documentclass{article}
\usepackage{<cjkpackage>}
\begin{document}
\begin{CJK}{<encoding>}{<fontname>}
你好!
\end{CJK}
\end{document}
```

In the above code

- `<cjkpackage>` can be `CJK` or `CJKutf8`. If you choose `CJK`, then `<encoding>` must be `GBK` and the file must be saved using the `GBK` encoding (such as `GB2312`). If you choose `CJKutf8`, then `<encoding>` must be `UTF8` and the file must be saved using the `utf-8` encoding.

[4] The file was originally written by Edward Lee, then modified by resolven, and subsequently yelewang.

- <fontname> can be any of the fonts installed.

A recommended setting (based on CTeX[5]) is

```
<cjkpackage>=CJK
<encoding>=GBK
<fontname>=song
```

In retrospect, we also recommend using the fast method in Section 4. Should you encounter any problem using the fast method, you can always fall back to the general method in Section 3.

References

[1] yulewang. TeXLive 2007 CJK Chinese Howto. http://mailboxpublic.googlepages.com/texlive2007cjkchinesehowto.

[2] yulewang. TeXLive 2007中文配置指南. Post No. 18. http://bbs.ctex.org/viewthread.php?tid=38043&extra=page%3D4&page=2.

[3] xptao. 为Texlive2007安装中文字体的脚本. http://bbs.ctex.org/viewthread.php?tid=39800&extra=page%3D1.

[4] novsep. 有没有人需要自动为Texlive2007安装UTF8中文字体的工具？. http://bbs.ctex.org/viewthread.php?tid=39640&extra=page%3D1.

A ctexlive.sh

```
#!/bin/bash

# $1 source ttf file, such as /media/wind/WINDOWS/Fonts/simkai.ttf
# $2 the name of font, such as song, kai, hei, fang, and etc.

# use format, ./ctexlive /media/wind/WINDOWS/Fonts simsun.ttc song

# for UTF8. \usepackage{CJKutf8} and \begin{CJK}{UTF8}{}

# prepare directory and files
mkdir -p ~/font
cd ~/font
cp /usr/local/texlive/2007/texmf-dist/source/latex/CJK/utils/subfonts/* ~/font/
cp /usr/local/texlive/2007/texmf/fonts/sfd/*.sfd ~/font/
mkdir -p $2
cd $2

# generate font
time fontforge -script ../subfonts.pe $1 $2 ../Unicode.sfd

# create map file
for i in *.tfm
do
cat >> $2.map << EOF
```

[5] http://www.ctex.org

```
${i%.tfm} ${i%.tfm} < ${i%.tfm}.pfb
EOF
done

# create fd file
cat >> c70$2.fd << EOF
% This is c70$2.fd for CJK package.
% created by Edward G.J. Lee
% modify by Yue Wang
\ProvidesFile{c70$2.fd}
\DeclareFontFamily{C70}{$2}{\hyphenchar \font\m@ne}
\DeclareFontShape{C70}{$2}{m}{n}{<-> CJK * $2}{}
\DeclareFontShape{C70}{$2}{bx}{n}{<-> CJKb * $2}{\CJKbold}
\endinput
EOF

# Copy Fonts into TEXMF
cd ~/.texlive2007/texmf-var
mkdir -p fonts/map/dvips/CJK
mkdir -p tex/latex/CJK/UTF8
mkdir -p fonts/tfm/CJK/$2
mkdir -p fonts/type1/CJK/$2
cp ~/font/$2/$2.map fonts/map/dvips/CJK/
cp ~/font/$2/c70$2.fd tex/latex/CJK/UTF8
cp ~/font/$2/*.tfm fonts/tfm/CJK/$2
cp ~/font/$2/*.pfb fonts/type1/CJK/$2

# Update The System
texhash
updmap --enable Map $2.map

# for GBK. \usepackage{CJK} and \begin{CJK}{GBK}{song}
cd ~/font/$2
perl ../uni2sfd.pl $2 ../UGBK.sfd gbk$2 gbk
mkdir -p ~/.texlive2007/texmf-var/fonts/tfm/CJK/gbk$2
mv gbk$2*.tfm ~/.texlive2007/texmf-var/fonts/tfm/CJK/gbk$2
mkdir -p ~/.texlive2007/texmf-var/fonts/vf
mv gbk$2*.vf ~/.texlive2007/texmf-var/fonts/vf
mkdir -p ~/.texlive2007/texmf-var/tex/latex/CJK/GBK

# create fd file for gbk

cd ~/.texlive2007/texmf-var/tex/latex/CJK/GBK

cat >> c19$2.fd << EOF
% This is c19$2.fd for CJK package.
% created by Edward G.J. Lee
\ProvidesFile{c19$2.fd}
\DeclareFontFamily{C19}{$2}{\hyphenchar \font\m@ne}
\DeclareFontShape{C19}{$2}{m}{n}{<-> CJK * gbk$2}{}
\DeclareFontShape{C19}{$2}{bx}{n}{<-> CJKb * gbk$2}{\CJKbold}
```

```
\endinput
EOF
```

B mkfont.sh

```
FULLNAME=$1
TTF2PT1=`which ttf2pt1`
DATE=`date`
TEXMF=~/texmf

if [ $# -eq 3 ]
then
  FNAME=`basename $FULLNAME`
  eval `echo $FNAME | awk -F. '{printf "FHEAD=%s;FTAIL=%s",\$1,\$2}'`
  FHEAD=$2
  FONTNAME=$3
else
  echo "Usage: `basename $0` your.ttf subfont_name font_name"
  exit
fi

check_enc()
{
    NUMLIST=`awk 'BEGIN{ n=1; while(n<256){printf "%02x\n",n; n++}}'`
    MAP=unicode-sample.map
    UTF8ENC=70
    GBKENC=19
    UTF8FD=${TEXMF}/tex/latex/CJK/UTF8
    GBKFD=${TEXMF}/tex/latex/CJK/GBK
}

create_type1()
{
  echo "Now create *.pfb and *.enc files, wait... "
  for i in $NUMLIST
  do
    ttf2pt1 -GAE -pttf -OHUBs -W0 -l plane+pid=3,eid=1,0x$i $FULLNAME ${FHEAD}$i
  done

# avoid dvips t1part module bugs.
  perl -pi -e 's/_/Z/g' *.t1a *.afm

  for ps in *.t1a
  do
    t1asm -b $ps > ${ps%.t1a}.pfb
  done
}

create_map()
{
```

```
MAPFILE=t1-${FHEAD}.map
ENCMAP=ttf-${FHEAD}.map
  echo "Create *.tfm and (dvips)map file, wait..."
  cat > $MAPFILE << EndOfFile
% This is map file for dvips/dvipdfm[x] and LaTeX CJK package.
% Created by Edward G.J. Lee <edt1023@info.sayya.org>
% $DATE
EndOfFile
  cat > $ENCMAP << EndOfFile
% This is map file for PDFLaTeX and LaTeX CJK package to embed TTF.
% Created by Edward G.J. Lee <edt1023@info.sayya.org>
% $DATE
EndOfFile
  for i in $NUMLIST
  do
    PSNAME=‘awk ’/FontName/ {print $2}’ ${FHEAD}$i.afm‘
    afm2tfm ${FHEAD}$i.afm > /dev/null 2>&1
    cat >> $MAPFILE << EndOfFile
${FHEAD}$i $PSNAME <${FHEAD}$i.pfb
EndOfFile
    cat >> $ENCMAP << EndOfFile
${FHEAD}$i <${FHEAD}$i.enc <${FNAME}
EndOfFile
  done
}

create_cidmap()
{
cat >> cid-x.map << EndOfFile
gbk$FHEAD@UGBK@ UniGB-UCS2-H :0:$FNAME
$FHEAD@Unicode@ unicode :0:$FNAME
EndOfFile
}

create_cjkfd()
{
cat > c${UTF8ENC}${FONTNAME}.fd << EndOfFile
\ProvidesFile{c${UTF8ENC}${FONTNAME}.fd}[\filedate\space\fileversion]
\DeclareFontFamily{C${UTF8ENC}}{$FONTNAME}{\hyphenchar \font\m@ne}
\DeclareFontShape{C${UTF8ENC}}{$FONTNAME}{m}{n}{<-> CJK * $FHEAD}{}
\DeclareFontShape{C${UTF8ENC}}{$FONTNAME}{bx}{n}{<-> CJKb * $FHEAD}{\CJKbold}
\endinput
EndOfFile

cat > c${GBKENC}${FONTNAME}.fd << EndOfFile
\ProvidesFile{c${GBKENC}${FONTNAME}.fd}[\filedate\space\fileversion]
\DeclareFontFamily{C${GBKENC}}{$FONTNAME}{\hyphenchar \font\m@ne}
\DeclareFontShape{C${GBKENC}}{$FONTNAME}{m}{n}{<-> CJK * gbk$FHEAD}{}
\DeclareFontShape{C${GBKENC}}{$FONTNAME}{bx}{n}{<-> CJKb * gbk$FHEAD}{\CJKbold}
\endinput
EndOfFile
```

```
}

create_vf()
{
  echo "Create virtual fonts file, wait..."
perl uni2sfd.pl $FHEAD UGBK.sfd gbk$FHEAD gbk
}

# main()
check_enc
create_type1
create_map
create_cidmap
create_cjkfd
create_vf

AFM=${TEXMF}/fonts/afm/$FHEAD
TFM=${TEXMF}/fonts/tfm/$FHEAD
PFB=${TEXMF}/fonts/type1/$FHEAD
ENC=${TEXMF}/fonts/enc/$FHEAD
VF=${TEXMF}/fonts/vf/$FHEAD
TTF=${TEXMF}/fonts/truetype/$FHEAD
MAPDIR=${TEXMF}/fonts/map
rm -f *.t1a
mkdir -p $AFM $TFM $PFB $ENC $VF $TTF
mv -f *.enc $ENC
mv -f *.afm $AFM
mv -f *.tfm $TFM
mv -f *.pfb $PFB
mv -f *.vf $VF
mv -f $FNAME $TTF
mkdir -p $MAPDIR/dvips
mkdir -p $MAPDIR/pdftex
mkdir -p $MAPDIR/dvipdfm
mv -f $MAPFILE $MAPDIR/dvips
mv -f $ENCMAP $MAPDIR/pdftex
cat  cid-x.map >> $MAPDIR/dvipdfm/cid-x.map
rm cid-x.map
mkdir -p $UTF8FD $GBKFD
mv -f c${UTF8ENC}${FONTNAME}.fd $UTF8FD
mv -f c${GBKENC}${FONTNAME}.fd $GBKFD

echo "Running texhash and updmap, pls wait"
texhash >/dev/null 2>&1
updmap --enable Map=t1-${FHEAD}.map >/dev/null 2>&1

echo "Congradulations! you have added ${FHEAD} into your texmf"
```

C unicode-sample.map

```
# this file is a sample Unicode map description.
# It describes which glyphs are to be included in the font
# and at which character position they are to be put.

# If the character position is greater than 255, the glyph is included, but
# does not appear in the encoding table (you must then use font reencoding
# to use this glyph).
# That makes it possible to have more than 256 glyphs in a font.
# Currently the maximum supported number of glyphs is 1024.

# Use this file as the argument to ttf2pt1's -L option.

# 1999-11-24 Thomas.Henlich@mailbox.tu-dresden.de
# 2000-03-01 Sergey Babkin: added 3rd format

# comment lines start with '#' or '%' or '//'

# The default source encoding table in the TTF file is Unicode (pid=3,eid=1).
# However a map may specify another source encoding with the "id <pid> <eid>"
# directive. If this directive is used at the beginning of the map file,
# it applies to the whole file. If it is used after a "plane" directive,
# then it sets the source encoding for this particular destination plane
# (possibly overriding the file-wide id directive). The user can also
# specify the source encoding explicitly at the comman line in the
# argument to the option -L. This used-specified source encoding overrides
# any id directives in the map file.

# examples:

# same as Unicode (default)
id 3 1

# One file may contain multiple actual translation tables. Each particular
# table within a file is named a plane. The primary use of planes is
# for multi-plane Eastern fonts with over 256 glyphs: for them one TTF
# file gets converted into multiple Type1 files, with each resulting file
# containing one plane of the original font. But they may also be used
# in other creative ways. Each plane may be specified in different format
# although this is not recommended for aesthetical reasons. If a map file
# contains any specifications of planes then the plane argument MUST
# be specified to the converter with that map file. If a map file
# contains no specifications of planes then the plane argument MUST NOT
# be specified to the converter with that map file.
#
# The plane maps start from the plane directive and continue to the next
# plane directive or end of file. The plane directive must be located
# at the very beginning of a separate string and contain the word "plane"
# followed by whitespace and the plane name. The whitespace characters
# are not allowed in the plane names. Non-alphanumeric characters are
# discouraged in the plane names as well.
```

```
# examples:

plane 81
=27 U+0027  APOSTROPHE

plane otherplane
 0, 1, 2

% There is one code assignment per line.
// Three formats are recognized:
# 1. optional whitespace, followed by '=', followed by a hex number
# (character position), followed by optional whitespace, followed by
# 'U+', followed by a four-digit hex number (the Unicode of the glyph we want
# here), followed by any number of characters.

// example:

=20 U+0020  SPACE
=48 U+0021  EXCLAMATION MARK
=22 U+0022  QUOTATION MARK
=23 U+0023  NUMBER SIGN
=24 U+0024  DOLLAR SIGN
=25 U+0025  PERCENT SIGN
=26 U+0026  AMPERSAND
=27 U+0027  APOSTROPHE

=E0 U+042E  CYRILLIC CAPITAL LETTER YU
=E1 U+0410  CYRILLIC CAPITAL LETTER A
=E2 U+0411  CYRILLIC CAPITAL LETTER BE
=E3 U+0426  CYRILLIC CAPITAL LETTER TSE
=E4 U+0414  CYRILLIC CAPITAL LETTER DE
=E5 U+0415  CYRILLIC CAPITAL LETTER IE
=E6 U+0424  CYRILLIC CAPITAL LETTER EF
=E7 U+0413  CYRILLIC CAPITAL LETTER GHE

% 2. optional whitespace, followed by '<', followed by one or more
% non-whitespace characters,
% followed by optional whitespace, followed by '/x',  followed by
% a hex number (character position), followed by optional
% whitespace, followed by '<U', followed by a four-digit hex number
% (the Unicode of the glyph we want here), followed by '>' and any number
% of characters.

# example:

<I>                   /x40     <U0049> LATIN CAPITAL LETTER I
<t>                   /x41     <U0074> LATIN SMALL LETTER T
<r>                   /x43     <U0072> LATIN SMALL LETTER R
<o>                   /x44     <U006F> LATIN SMALL LETTER O
<c>                   /x45     <U0063> LATIN SMALL LETTER C
<k>                   /x46     <U006B> LATIN SMALL LETTER K
```

```
<s>                     /x47     <U0073> LATIN SMALL LETTER S

<Eu>                    /xA4     <U20AC> EURO SIGN

# 3. optional whitespace, followed by '!', followed by a hex number
# (character position), followed by optional whitespace, followed by
# 'U+', followed by a four-digit hex number (the Unicode of the glyph we want
# here), followed by the name of the glyph that will be used in the
# output file.

# example:

!20 U+0020 space
!21 U+0021 exclam
!22 U+0022 quotedbl
!23 U+0023 numbersign
!24 U+0024 dollar
!25 U+0025 percent

# 4. compact format: just list of unicodes separated by commas or ranges
# denoted by a dash between unicodes. These unicodes are mapped to
# the output codes starting from 0 and continuously increasing.
# It is possible to reset the current code by using the "at" directive
# which must start at beginning of the line and give the new current
# output code (which will be assigned to the next occuring unicode)
# as decimal, hexadecimal or octal in C notation. The "at directive must
# take a separate line. The spaces around unicodes don't matter.

# example:
# map unicodes 0x40, 0x400, 0x4000 to the output codes 0, 1, 2 and unicodes
# 0xf010 - 0xf020, 0xf030 to the output codes 0x11-0x22

 0, 1, 2
at 0x11
 0xf010- 0xf020, 0xf030

# the first format is used by Roman Czyborra on his fine WWW pages:
# http://czyborra.com/charsets/iso8859.html

# the second format is used in the Linux locale charmaps files:
# /usr/share/i18n/charmaps/*

# we don't need those glyphs in the encoding table
=100     U+0030  DIGIT ZERO
=101     U+0031  DIGIT ONE
=102     U+0032  DIGIT TWO
=103     U+0033  DIGIT THREE
=104     U+0034  DIGIT FOUR
=105     U+0035  DIGIT FIVE
=106     U+0036  DIGIT SIX
=107     U+0037  DIGIT SEVEN
```

```
=108     U+0039  DIGIT NINE
=109     U+0038  DIGIT EIGHT
```