Typesetting Chinese in XeTeX: `zhspacing` user's manual*

Yin Dian†

2012/03/14

Contents

*Revision: 2012/03/14.

†Lab. of Network Comm. System & Control, Univ. of Sci. & Tech. of China.

File I

zhspacing.sty

1 Introduction

XeTeX is a new TeX engine by Jonathan Kew and SIL International, which combines ε-TeX with pervasive Unicode support and advanced font support[1]. Using XeTeX, it is possible to typeset scripts of any languages whose glyphs are contained in the font in use, without the help of any extra packages, such as `CJK`[2].

However, XeTeX itself does not solve all the problems. Some details in Chinese typesetting have not been dealt with by XeTeX, such as the automatic font switch between Chinese and Western characters, the skip adjustment of fullwidth punctuations, the automatic skip insertion between Chinese and Western characters or math formulas, etc.

To meet the need of easy and high quality typesetting of Chinese documents using XeTeX, the package `zhspacing` is coded, utilizing the new primitive of `\XeTeXinterchartoks` provided in XeTeX version 0.997.

2 Getting Started

`zhspacing` can be used in both plain XeTeX or XeLaTeX. The usage is quite straightforward. For instance, in plain XeTeX,

```
\input zhspacing.sty
\zhspacing
这是中文测 试。中文和English的混排。中
文和$E=mc^2$的混排。
\bye
```

[1] Words copied from the ConTeXt Wiki.

[2] `CJK` is a LaTeX 2e macro package which enables the use of CJK scripts (Chinese/Japanese/Korean) in various encodings, written by Werner Lemberg.

And in XeLaTeX,

```
\documentclass{article}
\usepackage{zhspacing}
\zhspacing
\begin{document}
这是中文测 试。中文和English的混排。中
文和$E=mc^2$的混排。
\end{document}
```

Both of the examples above produce the same result as follows,

这是中文测试。中文和 English 的混排。中文和 $E = mc^2$ 的混排。

As you can see, spaces after Chinese characters are always ignored. Moreover, a noticable skip is inserted between Chinese characters and English characters as well as math formulas. In fact, all of the following inputs can produce mixed langauge output with skip automatically inserted between Chinese and English characters,

中Eng文， 中 Eng文， 中Eng 文 and 中 Eng 文.

And their corresponding output is,

中 Eng 文， 中 Eng 文， 中 Eng 文 and 中 Eng 文.

Look close and you'll find, that the first and second input generates exactly the same output, and so does the third and fourth. However, the skip between `Eng` and 文 in the last two cases is wider than the skip between 中 and `Eng`. That is because the space is produced by the space token after the letter `g`, not the skip automatically inserted by `zhspacing`'s skip mechanism.

The space ignoring and skip auto-inserting ability of `zhspacing` can make writing documents much comfortable. You don't need to worry about the manual insertion of skips such as adding the annoying `~` in `CJK`.

Now come to the topic of punctuation skip adjustment. Proper Chinese typesetting requires consecutive fullwidth punctuations be compressed, and a linebreak before or after a fullwidth punctuation will cut off the blank spaces of this punctuation, making it align to the margin. `zhspacing` does have solved these problems, as well as proper prohibitions(禁则). Here's an example.

他强调，“三个代表”重要思想是在新的历史条件下运用马克思主义的立场、观点和方法的典范，是我们学习马克思主义的立场、观点和方法最现实、最生动的教材。“三个代表”重要思想是与时俱进的理论。

3 Advanced Usage

3.1 Fonts

`zhspacing` uses an extensible way of selecting fonts. The rules can be summarized as follows,

- Western characters, i.e. those that are not CJKV ideographies nor CJKV punctuations, use the same font as default.
- Chinese characters use seperate fonts. Font changes in the document does not affect the font used to display Chinese, unless you are using NFSS and change the font series or shape.
- On displaying basic Chinese ideographies, execute the command `\zhfont`.
- On displaying Chinese punctuations, execute the command `\zhpunctfont`.
- On displaying CJK Ext-A characters, execute the command `\zhcjkextafont`.
- On displaying CJK Ext-B characters, execute the command `\zhcjkextbfont`.
- When switching from non-Chinese characters to Chinese characters, execute `\zhs@savefont`. When switching back, execute `\zhs@restorefont`.

When importing `zhspacing` from plain XeTeX, the default definitions of the above commands are,

```
\font\zhfont="SimSun" at 10pt
\font\zhpunctfont="SimSun" at 10pt
\def\zhcjkextafont{\message{CJK Ext-A}}
\def\zhcjkextbfont{\message{CJK Ext-B}}
\let\zhs@savefont=\begingroup
\let\zhs@restorefont=\endgroup
```

When imported from XeLaTeX, the defaults are,

```
\newfontfamily\zhfont[BoldFont=SimHei]{SimSun}
\newfontfamily\zhpunctfont{SimSun}
\def\zhcjkextafont{\message{CJK Ext-A}}
\def\zhcjkextbfont{\message{CJK Ext-B}}
\def\zhs@savefont{\zhs@savef@nt{old}}
\def\zhs@restorefont{\zhs@restoref@nt{old}}
```

Here `\zhs@savef@nt` and `\zhs@restoref@nt` are internal macros to save and restore the NFSS info of the current font to specified places.

CJK Ext-A/B fonts have not been defined by default in consideration that not every user has installed the particular fonts. I recommend to use `Sun-ExtA` and `Sun-ExtB` as the corresponding fonts. You can define the ext-font macros manually in a similar way to the definition of `\zhfont`.

An example of mixed typesetting of various classes of characters is as follows, cited from The Tale of Kiều[3].

> 𤾓𢆥𥪝𡎝𠊛些 Trăm năm, trong cõi người ta, 𡨸才𡨸
> 命窖罗恄饒 Chữ tài, chữ mệnh, khéo là ghét nhau.
> 𣦆戈文局滄桑 Trải qua một cuộc bể dâu, 仍條𥉞𧡊
> 𦓡疸㤇𢚸 Những điều trông thấy mà đau đớn lòng. [4]

[3]《金云翘传》, or Truyện Kiều, a Vietnamese lục bát poem written in chữ Nôm. Chữ Nôm is a system of ideographies created by the Vietnamese laboring people, most of which are encoded in CJK Ext-A/B area, or not even included in Unicode.

[4]Some of the ideographies are replaced by similar characters due to the limitation of character set. A picture of the original manuscript can be found at this link.

3.2 More on Fonts

Sometimes it may be desirable to use different Chinese fonts for different NFSS font families in LaTeX. This can be done by changing the definition of `\zhs@savefont` and `\zhfont`, without modifying the `zhspacing` source.

Here is the way to do it.[5]

```
\makeatletter
\g@addto@macro\zhs@savefont{%
  \long\edef\zhs@tmpmacro{\f@family}%
  \def\zhs@curr@fam{0}%
  \ifx\zhs@tmpmacro\sfdefault
    \def\zhs@curr@fam{1}%
  \else\ifx\zhs@tmpmacro\ttdefault
    \def\zhs@curr@fam{2}%
  \fi\fi
  \edef\zhs@tmpmacro{\f@family}%
  \ifx\zhs@tmpmacro\sfdefault
    \def\zhs@curr@fam{1}%
  \else\ifx\zhs@tmpmacro\ttdefault
    \def\zhs@curr@fam{2}%
  \fi\fi
}
\newfontfamily\zhrmfont[BoldFont=SimHei,
  ItalicFont=KaiTi]{SimSun}
\newfontfamily\zhsffont{SimHei}
\newfontfamily\zhttfont[BoldFont=SimHei]{KaiTi}
\def\zhfont{\ifcase\zhs@curr@fam\zhrmfont\or\zhsffont
  \or\zhttfont\else\zhrmfont\fi}
```

3.3 Skips

The skip mechanism in `zhspacing` is also flexible. `zhspacing` utilizes the following skip commands. Note that they are defined as macros instead of skip register values, in order that the skip varies according to font size changes.

[5] See zhfont.sty later for convenient ways.

`\skipzh` Skip between adjacent Chinese characters.

`\skipenzh` Skip between a Chinese character and a Western character or a math formula.

`\skipzhopen` Skip before fullwidth opening punctuations, such as “‘”, “(”, “《”, etc.

`\skipzhinteropen` Skip before a fullwidth opening punctuation when preceded by another fullwidth punctuation.

`\skipzhlinestartopen` Skip before a fullwidth opening punctuation when it occurs at the start of a line.

`\skipzhclose` Skip after fullwidth closing punctuations, such as “’ ”, “) ”, “》 ”, etc.

`\skipzhinterclose` Skip after a fullwidth closing punctuation when followed by another fullwidth punctuation.

`\skipzhlineendclose` Skip after a fullwidth closing punctuation when it occurs at the end of a line.

`\skipzhfullstop` Skip after fullwidth fullstop punctuations, such as “； ”, “． ”, “。 ”, etc.

`\skipzhinterfullstop` Skip after a fullwidth fullstop punctuation when followed by another fullwidth punctuation.

`\skipzhlineendfullstop` Skip after a fullwidth fullstop punctuation when it occurs at the end of a line.

`\skipzhhalfstop` Skip after fullwidth halfstop punctuations, such as “、 ”, “， ”, “： ”, etc.

`\skipzhinterhalfstop` Skip after a fullwidth halfstop punctuation when followed by another fullwidth punctuation.

`\skipzhlineendhalfstop` Skip after a fullwidth halfstop punctuation when it occurs at the end of a line.

`\skipnegzhlinestartopen` Negative skip to `\skipzhlinestartopen`.

`\skipnegzhlineendclose` Negative skip to `\skipzhlineendclose`.

`\skipnegzhlineendfullstop` Negative skip to `\skipzhlineendfullstop`.

`\skipnegzhlineendhalfstop` Negative skip to `\skipzhlineendhalfstop`.

All of the skip commands above are defined in the pattern
`\def\skipxxx{\hskip xxxxx}`.

`zhspacing` comes with three pre-defined skip schemes, namely `\simsunskipscheme`, `\emptyskipscheme` and `\haltskipscheme`. The first scheme should be suitable for font `SimSun` and other popular Chinese fonts used in China, which does not support OpenType features of `halt`, and needs negative spaces be inserted before opening punctuations and after closing or judou punctuations. The second scheme simply addes zero length. And the last one should be fit for OpenType Chinese fonts supporting `halt` feature such as Adobe Song Std, where positive spaces should be inserted before or after certain punctuations. You can define your own skip schemes for customization, of course.

3.4 Vertical Chinese

Vertical Chinese can be achieved by adding raw feature `vertical` to the specified font. An example could be,

```
\documentclass[12pt]{article}
\usepackage{graphicx}
\usepackage{zhspacing}
\zhspacing
\begin{document}
\newfontfamily\zhfont[
  RawFeature={script=hani:language=CHN:vertical:+vert}]{SimSun}
\let\zhpunctfont\zhfont
\haltskipscheme
\rotatebox{-90}{我是中国人，我爱自己的祖国。}
\end{document}
```

我是中国人，我爱自己的祖国。

Note that in this example, in order to have proper vertical punctuations, we set `\zhpunctfont` to use the Adobe one supporting `vert`

feature, and change the skip scheme to `\haltskipscheme` to match the `vhal` feature specified.

However, there exists some bug in typesetting vertical Chinese containing punctuations using some fonts such as `KaiTi_GB2312` etc. I've told this bug to jjgod in May, but it seems that he hasn't solved the problem yet. Moreover, the baseline of vertical Chinese is not correct, so mixed typesetting of Chinese and English in vertical mode generates ugly results, and thus should be avoided.

3.5 Compatibility

Theoretically, `zhspacing` should be compatible with all macro packages, except those who change the definition of `\hskip` and `\penalty`, in which case special treatment should be applied. I haven't found any conflict when using common packages such as `hyperref` and `fancyhdr`. However, `ulem` redefineds `\hskip` and `\penalty`, and causes unexpected output[6]. Use `zhulem` provided along with `zhspacing` instead.

`zhspacing` may be not compatible with macro packages which use `\XeTeXinterchartoks` faculty. For example, the `polyglossia` package (some languages).

3.6 Working with other Chinese package

`zhspacing` is not compatible with `xeCJK`. The latter can be only used in LaTeX format. `zhspacing` is now supposed to be mainly used in Plain TeX format; in LaTeX, `xeCJK` is a better choice.

The `ctex` package, made by `ctex.org`, is a widely used Chinese document framework which hides the underlying Chinese processing

[6] I was using `\begingroup` and `\endgroup` as the font saving/restoring commands. However, I found `ulem` redefines `\hskip` and `\penalty` to end the current hbox group and start a new hbox, and thus group mismatch occurs. So I changed the font saving/restoring commands when used in LaTeX, using NFSS info instead of groups. But in the new hbox group, no Chinese font is applied by default so the Chinese characters disappears. Therefore I changed the definitions of `ulem`'s new `\hskip` and `\penalty` in `zhulem`, and advice users to use this modified one.

details to the user, which enables portability between different Chinese processing packages or systems. New version of `ctex` package supports XeTeX through `xeCJK` package, `zhspacing` is not necessary.

To use `zhspacing` in `cxetex`, a slight modification of the file `ctexcjk.clo` should be made. Change the corresponding lines to the followes (% means the original lines).

```
%\XeTeXlinebreaklocale "zh"
%\XeTeXlinebreakskip = 0pt plus 1pt minus 0.1pt
%\setmainfont[BoldFont={cwTeXHeiBold},
%              ItalicFont={cwTeXKaiItalic},
%              Mapping=tex-text]{cwTeXMing}
%\setsansfont[BoldFont={cwTeXHeiBold},Mapping=tex-text]{cwTeXYen}
%\setmonofont{cwTeXFangSongTT}
\usepackage{zhspacing}
\newfontfamily\zhfont[BoldFont=SimHei]{SimSun}
\newfontfamily\zhcjkextafont{Sun-ExtA}
\newfontfamily\zhcjkextbfont{Sun-ExtB}
\zhspacing
```

You can surely change these lines to your own taste.

4 Getting `zhspacing`

You can get latest stable version of `zhspacing` on CTAN.

`zhspacing` was hosted on an open-source project at google-code. The old project homepage is `http://code.google.com/p/zhspacing/`.

Now `zhspacing` is part of `ctex-kit` project (`http://code.google.com/p/ctex-kit/`). You can get the lastest source via SVN.

File II
zhfont.sty

1 Introduction

To simplify the complex font setting to use NFSS in LaTeX, the package `zhfont` is coded. It also adds simple interface to define fonts with fake slant, and also supports fake bold.

2 Package Options

`zhfont` currently has only one option — `fakebold`. This option sets Chinese show fake boldface when the current series is bold.

3 Provided Macros

- `\zhrmfont`
- `\zhsffont`
- `\zhttfont`

  The font commands to be executed according to the current font family. You can redefine them to customize your fonts.
- `\setzhmainfont`
- `\setzhsansfont`
- `\setzhmonofont`

  Simply macro to set `\zh{rm/sf/tt}font` using `\newfontfamily`.
- `\newfontfamilywithslant`

  Create a font family with fake slant. Take one argument as the font name.

  Example: `\newfontfamilywithslant\zhrmfont{SimSun}`
- `\newfontfamilywithslantandbold`

  Create a font family with fake slant, and specify another font for its boldface. Take two arguments as the font name and bold font name.

  Example:
  `\newfontfamilywithslantandbold\zhrmfont{SimSun}{SimHei}`

4 Examples

The following example is generated with `zhfont` loaded with option `fakebold`.

```
\newfontfamilywithslant\zhrmfont{SimSun}
中文\textbf{测试}。\textit{中文\textbf{测试}。}
\newfontfamilywithslant\zhrmfont{FangSong}
中文\textbf{测试}。\textit{中文\textbf{测试}。}
```

中文**测试**。*中文**测试***。 中文**测试**。*中文**测试***。

File III

zhmath.sty

1 Introduction

`zhmath` is a generic package which allows the Chinese characters to be displayed in math formulas. It also allows changing the math font for alphas and numbers.

`zhmath` only sets the basic Chinese characters to be displayable, not CJK Ext-A/B ones, as they are seldom used and discouraged in math. You can use `\hbox` to let `zhspacing` display them, however.

2 Package Options

In LaTeX, `zhmath` has two options — `active` and `noactive`. They set the method to implement Chinese character displaying in math. The default is `noactive`.

When `active` is set, the Chinese characters in math formulas are set to active, which, when executed, display the corresponding characters in `\zhmathfont`. This behavior is much like that of the

package `mathcjk`. The font size in different styles are also set in `\mathcjksizea/b/c/d`.

When `noactive` is set, the mathcodes of Chinese characters are set to use a special math font family to display them. This way is more native than to use active characters.

Both `active` and `noactive` have disadvantages. In `active` mode the Chinese math font sizes are determined by `\mathcjksizea/b/c/d`, so they are somewhat independent with the main text size, which means their size cannot change when the main font size changes, which gives ugly results. However, in `noactive` mode, although Chinese math font size varies with the main font size, Chinese inside `\mathrm` etc cannot be displayed, because the math font family are set to `\mathrm`'s. So choose the option to your own need.

In plain TEX no package options are available.

3 Provided Macros

- `\setzhmathfont`

  Set the math font used to display Chinese. Take one argument as the font command. Preamble only.

- `\zhmathfont`

  The font command which determines the math font for Chinese. However, if you use `noactive` option in LATEX, or you're in plain TEX, they are not executed when Chinese in math occur, but when `\setzhmathfont` is invoked. So you should always call `\setzhmathfont\zhmathfont` after the change.

- `\usecustommathfonts`

  Enable the use of custom math fonts for alphas and numbers. The default is Times New Roman.

- `\setalphanummathfonts`

  Set the math font used to display alphas and numbers. Preamble only. Take one argument as the font family in LATEX, and two argument as the font command for alphas and numbers accordingly in plain TEX.

- `\zhmathalphanumfont`

  The font command which determines the math font for alphas and numbers in LaTeX. Should always call `\setalphanummathfonts` `\zhmathalphanumfont` after it changes.

- `\zhmathletterfont`
- `\zhmathnumberfont`

  The font command which determines the math font for alphas and numbers in plain TeX. After their change `\setalphanummathfonts` with corresponding arguments should be called.

4 Examples

```
{
$中文 in math 测_试$
\usecustommathfonts
\Large $能量E = mc^2$, $s_总 = {1 \over 2} at^2$
}
```

中文*inmath*测试 $\quad$ 能量$E = mc^2$, $s_{总} = \frac{1}{2}at^2$