The \texttt{zhlipsum} Package: Chinese Dummy Text

Xiangdong Zeng

2018/09/08 \hspace{1em} v1.1.1∗

1 Introduction

The \texttt{zhlipsum} package is used for typesetting dummy text (i.e. "Lorem ipsum") as \texttt{lipsum}, \texttt{kantlipsum}, \texttt{blindtext} etc., but for Chinese language. Dummy text will be pretty useful, for example, when testing fonts or page styles.

\texttt{zhlipsum} supports UTF-8, GBK and Big5 encodings. Packages \texttt{expl3}, \texttt{xparse} and \texttt{l3keys2e} in the \LaTeX Project are required. To typeset Chinese properly, \texttt{zhlipsum} should be used with CJK package or C\TeX bundle.

2 User’s guide

\texttt{encoding = (utf8|gbk|big5)}

Package option for selecting encoding. Default value is \texttt{utf8}. For Unicode engines as \TeXLaTeX, Lua\TeX and upLaTeX, gbk / big5 encodings are invalid and utf8 will be used forcibly.

If you have loaded C\TeX bundle, then the encoding will be selected automatically according to C\TeX. Note that in C\TeX bundle, the corresponding options are \texttt{UTF8} and \texttt{GBK}, while the options in \texttt{zhlipsum} are all in \texttt{lowercase}.

\texttt{\zhlipsum[(paragraph)][(options)]}
\texttt{\zhlipsum*[[(paragraph)][(options)]]

Produce dummy text. Both arguments \texttt{(paragraph)} and \texttt{(options)} are optional. Note that spaces are not allowed between the arguments.

By default, the \texttt{\zhlipsum} command will insert \texttt{\par} before, after and between dummy text paragraphs, while \texttt{\zhlipsum*} will not give any extra processing. To change the default behavior, you can use the before, after and inter options described below.

The first optional argument \texttt{(paragraph)} should be a comma list. It can be specified as the following:

<table>
<thead>
<tr>
<th>Paragraph list example</th>
</tr>
</thead>
<tbody>
<tr>
<td>\zhlipsum[2-4]</td>
</tr>
<tr>
<td>\zhlipsum[4,12,3-8]</td>
</tr>
<tr>
<td>\zhlipsum[-10,40-]</td>
</tr>
<tr>
<td>\zhlipsum[]</td>
</tr>
</tbody>
</table>


3 Programming interface

The second optional argument (options) should be a key-value list. Supported options are the listed below.

- **name** = ⟨name⟩

  Select the name of the dummy text. There are 6 pre-defined dummy texts described in table 1. The default text is simp when encoding=utf8 or gbk, but trad when encoding=big5.

<table>
<thead>
<tr>
<th>Name</th>
<th>Paragraph numbers</th>
<th>Simplified / traditional</th>
<th>Description</th>
<th>Encodings’ support</th>
</tr>
</thead>
<tbody>
<tr>
<td>simp</td>
<td>50</td>
<td>S</td>
<td>Random dummy text</td>
<td>•</td>
</tr>
<tr>
<td>trad</td>
<td>50</td>
<td>T</td>
<td>Random dummy text</td>
<td>•</td>
</tr>
<tr>
<td>nanhanjing</td>
<td>43</td>
<td>T</td>
<td>Shanhaiping, Nanhanjing</td>
<td>•</td>
</tr>
<tr>
<td>xiangyu</td>
<td>45</td>
<td>T</td>
<td>Shiji: Xiang Yu Benji by Sima Qian</td>
<td>• • •</td>
</tr>
<tr>
<td>zhufu</td>
<td>110</td>
<td>S</td>
<td>Zhufu by Lu Xun</td>
<td>• • •</td>
</tr>
<tr>
<td>aspirin</td>
<td>66</td>
<td>S</td>
<td>Wikipedia: Aspirin</td>
<td>• • •</td>
</tr>
</tbody>
</table>

You can use \newzhlipsum command to define new dummy text as well.

- **before** = ⟨content⟩
- **after** = ⟨content⟩
- **inter** = ⟨content⟩

Insert contents before, after or between dummy text paragraphs. Note that the \par command inserted when using \zhlipsum will be overridden by the settings here.

\newzhlipsum{name={⟨name⟩}}{{⟨paragraphs list⟩}}

Declare new dummy text. The ⟨name⟩ is case sensitive and the ⟨paragraphs list⟩ is a comma list. An example is shown below:

**Example 2**

\newzhlipsum{jingyesi}{{% 
  \{床前明月光,\}, \{疑是地上霜。\}, \{举头望明月,\}, \{低头思故乡。\} }}

\zhlipsum*[-][name=jingyesi] % Print all the four sentences without \par'

3 Programming interface

Usually, the commands provided in section 2 are sufficient for users. For programmers professional users, however, the programming interface is also necessary and provided here. L\LaTeX3 syntax should be opened when using them.

\g_zhlipsum_seq

A sequence of dummy text names.

\zhlipsum_use:nn

Produce some dummy text paragraphs.

#1: Name
#2: Comma list of aragraph numbers.
4 Compatibility information

The following option exists in the beta version of \texttt{zhlipsum} package, but has become deprecated after version 1.0.0. It is reserved only for compatibility and may be removed in the future.

\begin{itemize}
\item Deprecated option. Now it’s the same as \texttt{name}.
\end{itemize}

5 Known issues

Dummy text nanshanjing and xiangyu have some rarely used characters. To display them correctly, you can use the \texttt{xeCJK} package and set SimSun-ExtB or Hanazono Mincho as the fallback font. Refer to the \texttt{xeCJK}'s user guide for specific methods (only for UTF-8 encoding and \texttt{XƎLₐₜX} engine).

GBK and Big5 encodings do not escape the ASCII range in the second byte, so the second byte of some Chinese characters may have the same encoding as special characters in ASCII like \{, \}, \ etc., which will lead to compilation failure. The .def files in \texttt{zhlipsum} are created with special techniques. Please do not modify them.

If there is no special requirement, UTF-8 encoding and Unicode engines as \texttt{XƎLₐₜX} and \texttt{LuaLₐₜX} are always recommended.

In special cases, if you must use GBK or Big5 encoding and need to declare new dummy text, the following method can be taken in order to avoid the problem temporarily.

\begin{example}
\begin{verbatim}
% File encoding should be Big5.
% \usepackage[encoding=big5]{zhlipsum}

% Using \newzhlipsum{big5}{許蓋功, 蓋功許, 功許蓋}' directly will
% lead to an error.
% Use <, >, + to replace {, } and \, and set the original {, } and \n% to be 'other' category (i.e. \texttt{catcode=12}).
\begingroup
\catcode`\<=1
\catcode`\>=2
\catcode`\+=0
\catcode`\{=12
\catcode`\}=12
\catcode`\\=12
+newzhlipsum<big5><許蓋功, 蓋功許, 功許蓋>
+endgroup
\zhlipsum[name=big5]
\end{verbatim}
\end{example}