

zhlipsum: 中文乱数假文(Lorem ipsum)

曾祥东

2017/09/16 v0.4*

如彭奇和瓦特曼的公共事业所证实的那样有一个胡子雪雪白的上帝超越时间超越空间确实存在他在神圣的冷漠神圣的疯狂神圣的哑哑的高处深深地爱着我们除了少数的例外不知什么原因但时间将会揭示他像神圣的密兰达一样和人们一起忍受着痛苦这班人不知什么原因但时间将会揭示生活在痛苦中生活在烈火中这烈火这火焰如果继续燃烧毫无疑问将使穹苍着火也就是说将地狱炸上天去天是那么蓝那么澄澈那么平静这种平静尽管时断时续总比没有好得多但是别这么快还要进一步考虑到泰斯丢和丘那德的人体测定学院的未完成的研究结果早已断定毫无疑问换句话说除了依附于人类的疑问之外别无其他疑问根据泰斯丢和丘那德的未完成的劳动的结果早已作出如下的论断但是别这么快不知什么原因根据彭奇和瓦特曼的公共事业的结果已毫无疑问地断定鉴于波波夫和贝尔契不知什么原因未完成的劳动以及泰斯丢和丘那德的未完成的劳动已经就业已被许多人所否认的论点作出论断认为泰斯丢和丘那德所假设的人认为实际存在的人认为人类总而言之统而言之尽管有进步的营养学和通大便药却在衰弱萎缩萎缩萎缩而且与此同时尤其是不知什么原因尽管体育运动在各方面都有很大进展如网球足球田径车赛游泳飞行划船骑马滑翔溜冰各式各样的网球各种各样致人死命的飞行运动各式各样的秋天夏天冬天冬天网球各种各样的曲棍球盘尼西林和代用品总之我接下去讲与此同时不知什么原因要萎缩要减少尽管有网球我接下去讲飞行滑翔九穴和十八穴的高尔夫球各种各样的网球总之不知什么原因在番克汉贝汉福尔汉克莱普汉换句话说与此同时尤其是不知什么原因但时间将会揭示要减少减少我接下去讲福尔汉克莱普汉总之自从塞缪尔·约翰逊去世以后到现在每个人的全部损失共计每人一时四两只是大概约略粗粗计算到小数点分量很足保持整数赤裸裸的光穿着袜子在康纳马拉总之不知什么原因不管怎样无论如何事实俱在尤其是考虑到更加远为严肃的看来更加严肃的鉴于斯丹威格和彼特曼的徒劳看来更加严肃的鉴于鉴于鉴于斯丹威格和彼特曼徒劳在平原在山地海洋在烈火沸腾的河里天空是一样的随后是大地换句话说天空随后是大地在一片寒冷一片漆黑中天空大地石头的住所一片寒冷中哎哟哟在我们的主诞生六百年左右天空大地海洋大地石头的住所汪洋中一片寒冷中在海上在陆地空中我接下去讲不知什么原因尽管有网球事实俱在但时间将会揭示我接下去讲哎哟哟总之一句话石头的住所谁能怀疑我接下去讲但是别这么快我接下去讲头颅要萎缩衰弱减少与此同时尤其是不知什么原因尽管有网球胡子火焰球队石头那么蓝那么平静哎哟哟头颅头颅头颅在康纳马拉尽管有网球未完成的徒然的劳动更加严肃的石头的住所总之我接下去讲哎哟哟徒劳的未完成的头颅头颅在康纳马拉尽管有网球头颅哎哟石头丘那德(混战,最后的狂喊)网球……石头……那么平静……丘那德……未完成的……

——萨缪尔·贝克福特《等待戈多》

*<https://github.com/Stone-Zeng/zhlipsum>.

第 1 节 使用说明

`encoding`

New: 2017-09-16

`encoding = <UTF8|GBK>`

用于指定编码的宏包选项，可以在调用宏包的时候设定，默认为 UTF8。对于 X_YLaTeX、LuaLaTeX 和 upLaTeX 等 Unicode 引擎，GBK 编码无效，宏包将强制使用 UTF8 编码。

`\zhlipsum`

`\zhlipsum[<段落>]`

插入假文文本，每段之间会分段。

只给定一个参数时，将输出对应序号的段落，如 `\zhlipsum[17]` 将生成第 17 段。给定两个参数并用“-”连接时，则会输出该范围内的段落。如 `\zhlipsum[5-34]` 将生成第 5 至 34 段假文。没有指定可选参数时，默认输出第 1 至 3 段假文。

目前本宏包只提供了 50 段假文文本，超出范围的数字一律按 50 处理。如 `\zhlipsum[51]` 将给出第 50 段，而 `\zhlipsum[48-55]` 将只会输出第 48、49 和第 50 段。

当前仅支持简体中文的假文，以“劳仑衣普桑”开头。

`\zhlipsum*`

`\zhlipsum*[<段落>]`

插入假文文本，每段之间不分段。用法同 `\zhlipsum`。

第 2 节 已知问题

用 `\zhlipsum*` 构建较长的单段假文文本时，可能会很慢。可能是因为 T_EX 的断行算法需耗费较多时间来处理长段落。

第 3 节 实现细节

```
1 <{*package>
2 <{@@=zhlipsum>
3 \RequirePackage { xparse, l3keys2e }
```

3.1 内部变量和函数

`\l__zhlipsum_tmpa_str` 临时变量。

```
4 \str_new:N \l__zhlipsum_tmpa_str
```

`\g__zhlipsum_encoding_tl` 存放编码信息。

```
5 \tl_new:N \g__zhlipsum_encoding_tl
```

`\c__zhlipsum_max_par_number_int` 假文的最大段落数。目前设为 50。

```
6 \int_const:Nn \c__zhlipsum_max_par_number_int { 50 }
```

`__zhlipsum_if_unicode_engine:TF` 判断是否为 Unicode 引擎。来自于 `zhnumber` 宏包。

```
7 \bool_lazy_any:nTF
8 {
9   { \sys_if_engine_xetex_p: }
10  { \sys_if_engine luatex_p: }
11  { \sys_if_engine_uptex_p: }
```

```

12 }
13 { \cs_new_eq:NN \__zhlipsum_if_unicode_engine:TF \use_i:nn }
14 { \cs_new_eq:NN \__zhlipsum_if_unicode_engine:TF \use_ii:nn }

```

__zhlipsum_input_def_file:N 读入定义文件。

```

15 \cs_new:Npn \__zhlipsum_input_def_file:N #1
16 { \exp_args:Nx \file_input:n { zhlipsum-zh-cn- #1 .def } }

```

3.2 宏包选项

encoding 设置编码。

```

17 \keys_define:nn { zhlipsum / option }
18 {
19   encoding .choice:,
20   encoding / UTF8 .code:n =
21     { \tl_set:Nn \g__zhlipsum_encoding_tl { utf8 } },
22   encoding / GBK .code:n =
23     {
24       \__zhlipsum_if_unicode_engine:TF
25       {
26         \tl_set:Nn \g__zhlipsum_encoding_tl { utf8 }
27         \msg_warning:nn { zhlipsum } { unicode-engine }
28       }
29       { \tl_set:Nn \g__zhlipsum_encoding_tl { gbk } }
30     },
31   encoding / unknown .code:n =
32     { \msg_error:nnn { zhlipsum } { encoding-invalid } {#1} },
33   encoding .value_required:n = true,
34   encoding .initial:n = { UTF8 },

```

处理未知选项。

```

35   unknown .code:n =
36     { \msg_error:nn { zhlipsum } { unknown-option } }
37 }

```

提示信息。

```

38 \msg_new:nnn { zhlipsum } { unicode-engine }
39 {
40   You~ are~ now~ using~ Unicode~ engine~ \c_sys_engine_str. \\
41   Encoding~ "GBK"~ is~ invalid.~ Changed~ into~ UTF8.
42 }
43 \msg_new:nnn { zhlipsum } { encoding-invalid }
44 {
45   Encoding~ "#1"~ is~ invalid. \\
46   Available~ encodings~ are~ "UTF8"~ and~ "GBK".
47 }
48 \msg_new:nnn { zhlipsum } { unknown-option }
49 { Package~ option~ "\l_keys_key_tl"~ is~ unknown. }

```

将宏包选项传给 zhlipsum/option。

```

50 \ProcessKeysOptions { zhlipsum / option }

```

根据编码读入假文文本定义文件。

```

51 \__zhlipsum_input_def_file:N \g__zhlipsum_encoding_tl

```

3.3 核心函数

`\zhlipsum` 输出假文, 可选参数表示段落数, 默认为 1-3。命令前后均插入了 `\par`, 以便强制分段。带星号的命令在假文主体间不分段。

```

52 \NewDocumentCommand \zhlipsum
53 { s >{ \SplitArgument { 1 } { - } } 0{ 1 - 3 } }
54 {
55   \__zhlipsum_process_par_number:nn #2
56   \par
57   \IfBooleanTF {#1}
58   {
59     \zhlipsum_print_multi_no_par:VVV
60     \l__zhlipsum_begin_int \l__zhlipsum_step_int \l__zhlipsum_end_int
61     \par
62   }
63   {
64     \zhlipsum_print_multi:VVV
65     \l__zhlipsum_begin_int \l__zhlipsum_step_int \l__zhlipsum_end_int
66   }
67 }

```

`\l__zhlipsum_begin_int` 循环的开始、结束和步长。

```

\l__zhlipsum_end_int
\l__zhlipsum_step_int
68 \int_new:N \l__zhlipsum_begin_int
69 \int_new:N \l__zhlipsum_end_int
70 \int_new:N \l__zhlipsum_step_int

```

`__zhlipsum_process_par_number:nn` 处理段落数参数。

```

71 \cs_new:Npn \__zhlipsum_process_par_number:nn #1#2
72 {

```

获得循环的开始位置和结束位置。

```

73   \__zhlipsum_process_par_number_aux:Nn \l__zhlipsum_begin_int {#1}
74   \IfNoValueTF {#2}
75   { \int_set_eq:NN \l__zhlipsum_end_int \l__zhlipsum_begin_int }
76   { \__zhlipsum_process_par_number_aux:Nn \l__zhlipsum_end_int {#2} }

```

开始位置在结束位置之前时, 步长为 1, 否则为 -1。

```

77   \int_compare:nTF
78   { \l__zhlipsum_begin_int <= \l__zhlipsum_end_int }
79   { \int_set_eq:NN \l__zhlipsum_step_int \c_one }
80   { \int_set:Nn \l__zhlipsum_step_int { -1 } }
81 }

```

`__zhlipsum_process_par_number_aux:Nn` 与最大段落数进行比较, 防止下标越界。

```

82 \cs_new:Npn \__zhlipsum_process_par_number_aux:Nn #1#2
83 {
84   \int_compare:nTF { #2 <= \c__zhlipsum_max_par_number_int }
85   { \int_set:Nn #1 {#2} }
86   { \int_set_eq:NN #1 \c__zhlipsum_max_par_number_int }
87 }

```

`\zhlipsum_print_multi:nnn` 输出带段落的假文。

```

\zhlipsum_print_multi:VVV
88 \cs_new:Npn \zhlipsum_print_multi:nnn #1#2#3
89 {
90   \int_step_inline:nnnn {#1} {#2} {#3}
91   { \__zhlipsum_print_single:n {##1} \par }
92 }
93 \cs_generate_variant:Nn \zhlipsum_print_multi:nnn { VVV }

```

\zhlipsum_print_multi_no_par:nnn
 \zhlipsum_print_multi_no_par:VVV

输出不带段落的假文。

```
94 \cs_new:Npn \zhlipsum_print_multi_no_par:nnn #1#2#3
95 {
96   \int_step_inline:nnnn {#1} {#2} {#3}
97   { \__zhlipsum_print_single:n {##1} }
98 }
99 \cs_generate_variant:Nn \zhlipsum_print_multi_no_par:nnn { VVV }
```

__zhlipsum_print_single:n

输出第 n 段假文。方法是将段落数 n 先转换成罗马数字，再组成完整命令。目前仅支持简体中文。

```
100 \cs_new:Npn \__zhlipsum_print_single:n #1
101 {
102   \str_set:Nx \l__zhlipsum_tmpa_str { \int_to_roman:n {#1} }
103   \str_put_left:Nn \l__zhlipsum_tmpa_str { c__zhlipsum_text_zh_cn_ }
104   \str_put_right:Nn \l__zhlipsum_tmpa_str { _tl }
105   \use:c { \str_use:N \l__zhlipsum_tmpa_str }
106 }
107 </package>
```

3.4 假文文本——简体中文

```
108 <*zh-cn>
109 <@=@zhlipsum_text_zh_cn>
```

由 <http://www.cancms.com/content/dummytext> 网站生成,略有改动。

\c__zhlipsum_text_zh_cn_i_tl

第1段。

```
110 \tl_const:Nn \c__zhlipsum_text_zh_cn_i_tl
111 {
112   劳仑衣普桑，认至将指点效则机，最你更枝。想极整月正进好志次回
113   总般，段然取向使张规军证回，世市总李率英茄持伴。用阶千样响领
114   交出，器程办管据家元写，名其直金团。化达书据始价算每百青，金
115   低给天济办作照明，取路豆学丽适市确。如提单各样备再成农各政，
116   设头律走克美技说没，体交才路此在杠。响育油命转处他住有，一须
117   通给对非交矿今该，花象更面据压来。与花断第然调，很处已队音，
118   程承明卹。常系单要外史按机速引也书，个此少管品务美直管战，子
119   大标蠢主盯写族般本。农现离门亲事以响规，局观先示从开示，动和
120   导便命复机李，办队呆等需杯。见何细线名必子适取米制近，内信时
121   型系节新候节好当我，队农否志杏空适花。又我具料划每地，对算由
122   那基高放，育天孝。派则指细流金义月无采列，走压看计和眼提问
123   接，作半极水红素支花。果都济素各半走，意红接器长标，等杏近乱
124   共。层题提万任号，信来查段格，农张雨。省着素科程建特色被什，
125   所界走置派农难取眼，并细杆至志本。
126 }
```

\c__zhlipsum_text_zh_cn_ii_tl

第2段。

```
127 \tl_const:Nn \c__zhlipsum_text_zh_cn_ii_tl
128 {
129   水厂共当而面三张，白家决空给意层般，单重总歼者新。每建马先口
130   住月大，究平克满现易手，省否何安苏京。两今此叫证程事元七调联
131   派业你，全它精据间属医拒严力步青。厂江内立拉清义边指，况半严
132   回和得话，状整度易芬列。再根心应得信飞往清增，至例联集采家同
133   严热，地手蠢持查受立询。统定发几满斯究后参边增消与内关，解系
134   之展习历李还也村酸。制周心值示前她志长步反，和果使标电再主它
135   这，即务解旱八战根交。是中文之象万影报头，与劳工许格主部确，
136   受经更奇小极准。形程记持件志各质天因时，据据极清总命所风式，
137   气太束书家秀低坟也。期之才引战对已公派及济，间究办儿转情革统
138   将，周类弦具调除声坑。两了济素料切要压，光采用级数本形，管县
139   任其坚。切易表候完铁今断土马他，领先往样拉口重把处千，把证建
140   后苍交码院眼。较片的集节片合构进，入化发形机已斯我候，解肃飞
```

```
141 口严。技时长次土员况属写，器始维期质离色，个至村单原否易。重
142 铁看年程第则于去，且它后基格并下，每收感石形步而。
143 }
```

\c_zhlipsum_text_zh_cn_iii_tl 第 3 段。

```
144 \tl_const:Nn \c_zhlipsum_text_zh_cn_iii_tl
145 {
146 她己道接收面学上全始，形万然许压己金史好，力住记赤则引秧。较
147 处高方据近学级素专，者往构支明系状委起查，增子束孤不般前。验
148 相斗真它增备听片思三，听花连次志平品书消情，清市五积群面县。
149 开价现准此省持给，争式身在南决就集般，地力秧众团计。日车治样
150 政技便角想持中，厂期平及半干速区白土，观合村究研称始这少。海
151 验商眼件容果经风中，质江革再的采心年专，光制单万手斗光就，响
152 报却蹦杯材。内同数速果报做，属马市参至，入极将管医。但强质候
153 交上能只拉，据特光农无五计据，来步孤平葡院。江养水图再难边
154 气，做林因列行消特段，就解届罐盛。定她识决听人自打验，快思准
155 月断细面便，事定什呀传。边力心层下等共命每，厂五交型车想所
156 利，直下报亲积速。元前很地传气领权节，求反立全各市状，新上录
157 所走值上。明统多表过变物每区广，会王问西听观生真林，二决定织
158 助议苏。格节基全却及飞口悉，难之规利争白观，证查李却调代动。
159 斗形放数委同领，内从但五身。当了美话也步京边但容代认，放非济
160 边建按划近些派民越，更具建火法住收保步连。
161 }
```

以下还有 47 段，此处从略。

```
162 <@@=>
163 </zh-cn>
```

版本历史

v0.1	(2017/04/08)	<code>_zhlipsum_process_par_number:nn</code> : 改为内部函数。	4
General: 开始编写宏包。	1		
v0.2	(2017/04/14)	v0.4	(2017/09/16)
General: 仿照 <code>kantlipsum</code> 宏包, 实现任意的段落选取。	1	General: 将安装、测试文件集成进源文件。	1
使用名字空间。	1	新增 <code>encoding</code> 选项。	1
v0.3	(2017/07/28)	优化宏包实现。	1
<code>_zhlipsum_print_single:n</code> : 改为内部函数。	5		

代码索引

意大利体的数字表示描述对应索引项的页码; 带下划线的数字表示定义对应索引项的代码行号; 罗马字体的数字表示使用对应索引项的代码行号。

B	<code>\l_keys_key_tl</code>	49
bool commands:		
<code>\bool_lazy_any:nTF</code>	7	
C		
cs commands:		
<code>\cs_generate_variant:Nn</code>	93, 99	
<code>\cs_new:Npn</code>	15, 71, 82, 88, 94, 100	
<code>\cs_new_eq:NN</code>	13, 14	
E		
encoding	2, 17	
exp commands:		
<code>\exp_args:Nx</code>	16	
F		
file commands:		
<code>\file_input:n</code>	16	
I		
<code>\IfBooleanTF</code>	57	
<code>\IfNoValueTF</code>	74	
int commands:		
<code>\int_compare:nTF</code>	77, 84	
<code>\int_const:Nn</code>	6	
<code>\int_new:N</code>	68, 69, 70	
<code>\int_set:Nn</code>	80, 85	
<code>\int_set_eq:NN</code>	75, 79, 86	
<code>\int_step_inline:nnnn</code>	90, 96	
<code>\int_to_roman:n</code>	102	
<code>\c_one</code>	79	
K		
keys commands:		
<code>\keys_define:nn</code>	17	
M		
msg commands:		
<code>\msg_error:nn</code>	36	
<code>\msg_error:nnn</code>	32	
<code>\msg_new:nnn</code>	38, 43, 48	
<code>\msg_warning:nn</code>	27	
N		
<code>\NewDocumentCommand</code>	52	
P		
<code>\par</code>	56, 61, 91	
<code>\ProcessKeysOptions</code>	50	
R		
<code>\RequirePackage</code>	3	
S		
<code>\SplitArgument</code>	53	
str commands:		
<code>\str_new:N</code>	4	
<code>\str_put_left:Nn</code>	103	
<code>\str_put_right:Nn</code>	104	
<code>\str_set:Nn</code>	102	
<code>\str_use:N</code>	105	
sys commands:		
<code>\c_sys_engine_str</code>	40	
<code>\sys_if_engine luatex_p:</code>	10	
<code>\sys_if_engine uptex_p:</code>	11	
<code>\sys_if_engine xetex_p:</code>	9	
T		
tl commands:		
<code>\tl_const:Nn</code>	110, 127, 144	

\tl_new:N	5	\g_zhlipsum_encoding_tl	<u>5</u> , 21, 26, 29, 51
\tl_set:Nn	21, 26, 29	\l_zhlipsum_end_int	60, 65, <u>68</u> , 75, 76, 78
U			
use commands:			
\use:N	105	_zhlipsum_if_unicode_engine:TF	<u>7</u> , 24
\use_i:nn	13	_zhlipsum_input_def_file:N	<u>15</u> , 51
\use_ii:nn	14	\c_zhlipsum_max_par_number_int ...	<u>6</u> , 84, 86
Z			
\zhlipsum	2, 2, <u>52</u>	_zhlipsum_print_single:n	91, 97, <u>100</u>
zhlipsum commands:			
\zhlipsum_print_multi:nnn	64, <u>88</u>	_zhlipsum_process_par_number:nn ...	55, <u>71</u>
\zhlipsum_print_multi_no_par:nnn	59, <u>94</u>	_zhlipsum_process_par_number_aux:Nn ..	73, 76, <u>82</u>
zhlipsum internal commands:			
\l_zhlipsum_begin_int ...	60, 65, <u>68</u> , 73, 75, 78	\l_zhlipsum_step_int	60, 65, <u>68</u> , 79, 80
		\c_zhlipsum_text_zh_cn_i_tl	<u>110</u>
		\c_zhlipsum_text_zh_cn_ii_tl	<u>127</u>
		\c_zhlipsum_text_zh_cn_iii_tl	<u>144</u>
		\l_zhlipsum_tmpa_str	<u>4</u> , 102, 103, 104, 105
		\zhlipsum*	2, 2