ConTeXt
up-to-date
1999/7

Chinese

中文
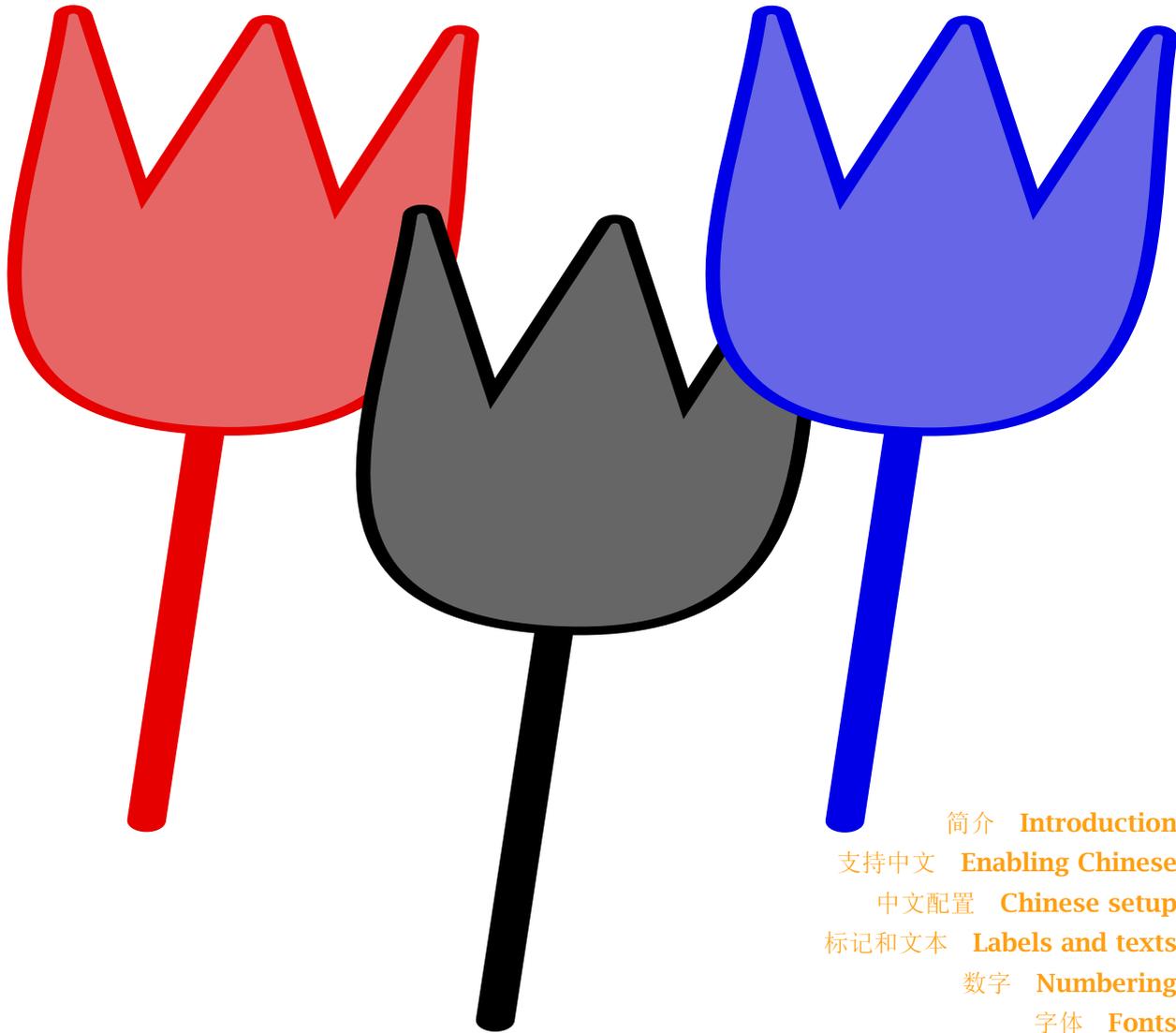
Introduction

简介
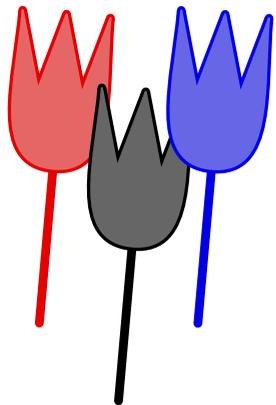
In this up-to-date we will introduce support for the Chinese language. When王磊 —who also kindly translated this document into Chinese— asked me if CONTEXT could handle his language, he had already tried to run existing macro packages on top of CONTEXT, but without result. This meant that support for Chinese had to be implemented from scratch.

Unfortunately the documentation to the existing packages based on LaTeX and PLAIN, were written in Chinese, but Wang Lei provided me with some typeset examples and some font files. It soon became clear that supporting Chinese is mainly a matter of dealing with UNICODE input.

Instead of relying on pre-processors, CONTEXT implements (dual byte) UNICODE support in a way similar to the German " and Polish / prefixing. By making the characters in the range 128 − 255 active, and looking ahead, a rather efficient implementation was possible.

在这一次更新中我们将介绍对中文的支持。当本文的中文译者王磊询问我 CONTEXT 是否能处理汉语时，他已经尝试过运行现有的宏集，但是没有成功。这意味着实现对中文的支持还比须作大量的工作。

不走运的是，目前基于 LaTeX 和 PLAIN 的集的文档是用中文书写的。但是在研究王磊提供给我一些排版例子和一些字体文件后，不久我明白在 CONTEXT 中支持中文主要是处理 UNICODE 输入的问题。

我决定采用类似于德语和波兰语预设定的方式来实现 UNICODE 支持，而不是依赖于预处理。通过将 128 − 255范围内的字符激活，可以预见，在 CONTEXT 中实现十分有效的中文处理是完全可能的。

Enabling Chinese 支持中文

As said, we will use UNICODE input. Because we operate in the upper range of the character table, we can use another language as default. Chinese numbering of chapters and floats are defined as a language specific options, and will be enabled when one switches to Chinese explicitly:

```
\mainlanguage[chinese]
```

Although the underlying mechanism differs from the way CONTEXT normally handles fonts, we still use \setupbodyfont to load, or more correct, to activate, the font:

```
\setupbodyfont[chi]
```

For changing the style (*slanted*, **bold**, etc) we use the same commands as in western languages. Because on the background we use font filename mapping, definitions look like:

```
\defineunicodefont [SimChi] [SimplifiedChinese] [chinese]

\definefontsynonym [SimplifiedChineseRegular] [gbsong]
\definefontsynonym [SimplifiedChineseSlanted] [gbsongsl]
\definefontsynonym [SimplifiedChineseItalic]  [gbsongsl]
```

如上所述，我们将使用 UNICODE 输入。因为我们要处理的汉字位于字符表的高位区，所以可采用其他语言做为缺省语言。使用中文数字的章节和浮动图表被定义为一特别的语言选项，当应用此选项时只须简单的在文档中加入：

尽管底层的处理机制不同于一般，\setupbodyfont 还是被用来载入汉字，更确切的说，激活汉字：

通过采用如下的字体文件名映射，我们得以使用与西方语言相同的命令来改变字形 (*slanted,* **bold**))。

Chinese setup

中文配置
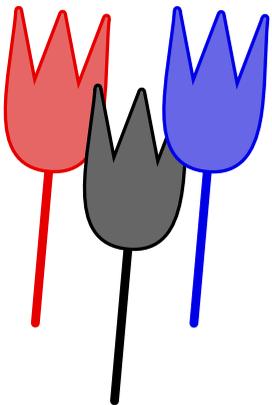
Instead of providing settings as a third argument to \defineunicodefont, we inherit the settings from the more generally defined Chinese ones. Of course one can change these, for instance:

我们从一般定义的中文 \defineunicodefont中继承一些设置，而不是将这些设置做为该命令的第三个参数来提供。当然，你可以改变这些设置。例如：

```
\setupunicodefont
   [chinese]
   [scale=0.9, height=1.2, depth=1]
```

The scale is applied on top of the normal font scaling, while the height and depth deal with a UNICODE specific strut. Think of them as the height and depth of a line.

scale 是相对于一般大小的汉字的缩放值，height 和 depth 设定了 UNICODE 的线条的高度和深度。

Spacing around Chinese is rather special. Redundant spaces can (and should) be surpressed by surrounding the text with the command \purechinese.

汉字的字间距是十分特别的。中文文本中多余的间距可以用命令 \purechinese 来消除掉。

```
Like it or not, but \purechinese{我爱用中文} looks like this!
```

In a similar way, one can prevent unwanted spaces by saying:

同样地，当要防止出现多余的字间距时，你可能会问：

```
We definitely don't want a space inside 明月几时有\index{why}把酒问青天, do we?
```

In this case one can put \stillchinese in front of the \index command.

在这种情况下，你可在命令 \index 前加上 \stillchinese。

The long list of numbers in the previous macro

对于很多在断行时需要特别注意的标点符号，都在预
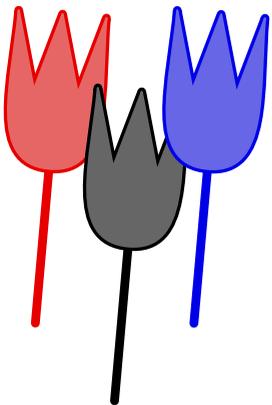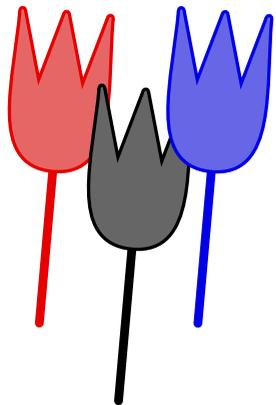
identify the characters where special care is needed for breaking lines. A linebreak is not permitted before:

先定义的宏中列出以便识别。在下述标点符号前不允许断行。

、	161 162	'	161 175	〕	161 189	'	163 167	?	163 191
		"	161 177	】	161 191	)	163 169	]	163 221
		）	161 179	°	161 227	,	163 172	}	163 253
。	161 163	〉	161 181	′	161 228	.	163 174		
·	161 164	》	161 183	″	161 229	:	163 186		
¨	161 167	」	161 185	!	163 161	;	163 187		
…	161 173	』	161 187	"	163 162	>	163 190		

A linebreak is not permitted after the following glyphs:

在下述符号后不允许断行。

'	161 174	﹙	161 178	『	161 186	[	163 219
		〈	161 180	〖	161 188	`	163 224
		《	161 182	【	161 190	{	163 251
"	161 176	「	161 184	﹙	163 168		

Those who want some more insight in the way spacing and linebreaks are taken care of, can enable tracing. When tracing Chinese, one can best enable color and first generate a legend to the symbols that are used.

如果想要更仔细的观察字间距和如何断行，可使用跟踪。当对中文进行跟踪时，你可以很好地使用彩色，并且首先产生一个文中所用到的符号的图示。

```
\tracechinesetrue \showchinesetracelegend
```

Labels and texts　　标记和文本

Chinese head and label texts are already defined in the file `lang-chi.tex`. They are defined using the \uchar command, like in:

中文的标题和标记文本都已在 `lang-chi.tex` 中用命令 \uchar 加以定义。例如：

```
\setupheadtext  [cn] [ content=\uchar{196}{191}\uchar{194}{188}]
\setuplabeltext [cn] [appendix=\uchar{184}{189}\uchar{194}{188}]
```

Conversion of numbers is set up as:　　数字的转换设置为：

```
\setupsection
   [sectionlevel-1]
   [conversion=chinese]

\setupcaptions
   [conversion=chinese]
```

As said, these are defined as language specifics, which means that they can only be overruled by appending them to the list of Chinese language specific settings:

如上所述，这些设置被定义为语言的特性。也就是说，它们只有附加在中文特性设置列表中才能起作用。

```
\startlanguagespecifics[cn]
   ... new settings ...
\stoplanguagespecifics
```

By default, page numbers are typeset in western glyphs. If you want Chinese instead say:

缺省情况下，页码使用与西文一样的设置。如需使用中文则用下面的命令：

```
\setuppagenumbering[conversion=chinese]
```
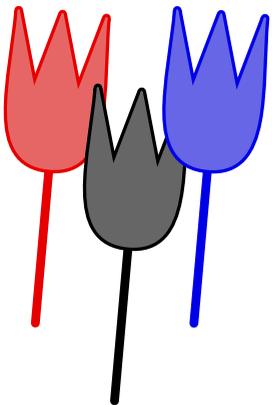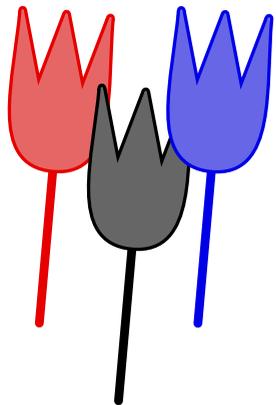
Numbering 数字

The Chinese numbering system is rather straight-forward. First there are the digits:

1	2	3	4	5	6	7	8	9
一	二	三	四	五	六	七	八	九

中文的数字系统是十分直接了当的。首先列出一些基本中文数字：

The powers of ten have their own representation.

10	100	1000	10000	100000000
十	一百	一千	一万	一亿

在中文中对 10的幂有自己的表示方法。

The number 12 is a combination of $1 \times 10 + 2$, or:十二 , while 22 becomes 二十二 . The numbers < 20 are treated a bit different, because there the 1 is omitted.

数字 12 是 $1 \times 10 + 2$ 的组合，或者：十二。而 22 表示成二十二。小于 20 的数字处理起来稍有不同，因为十位数前的 1 在用中文表示时被省略掉了。

1	一		9	九		4	四
11	十 一		99	九 十 九		16	十 六
111	一 百 一 十 一		999	九 百 九 十 九		256	二 百 五 十 六
1111	一千一百一十一		9999	九千九百九十九		65536	六万五千五百三十六

The implementation is not that complicated. The digits $1 - 9$ and numbers 10, 100, 1000, 10000 and 100000000 are hard coded into the file `font-chi.tex`.

数字的处理是很简单的。数字 $1 - 9$ 和 10、100、1000、10000 和 100000000 都在 `font-chi.tex` 中转成了相应的中文数字。

While in arabic calculations addition leads to more digits, in Chinese the number of glyphs

在阿拉伯数字计算中加法会导致更多的数字，而在中文中数字会（暂时）减少。考虑到使用中文的人比
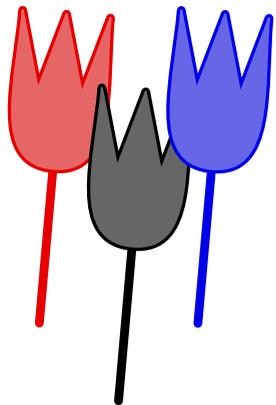
can (temporarily) decrease. Given that the number of people dealing with Chinese is considerably larger than the number of latin (or even english) speaking people, a successor of TₑX definitely must provide a `\chinesenumeral` primitive.

说拉丁语的人多很多，一个成功的 TₑX 一定要提供 `\chinesenumeral`。

2546	二千五百四十六	
9258	九千二百五十八	+
11804	一万一千八百零四	

When Chinese support is loaded, one has access to yet another conversion: `chinese` or `c`. Like any conversion, this one is available for conversion of chapter, page and other numbers, as well as in itemizations

当载入中文支持后，还可以进行另外的转换。如同所有的转换一样，这些转换包括章节、页码和其它一些数字，例如在列表中。

一 .normally, not all numbers in a document are in Chinese, for instance chapter numbers are, but page numbers are not

二 .nowadays, most Chinese texts are typeset from left to right instead of the traditional up–down way

三 .there are two Chinese scripts in use: Simplified Chinese, which is used in the Mainland of China and the other one is Traditional Chinese, which is used in Taiwan and Hongkong

1. 通常情况下，并不是文档中所有的数字都采用中文表示。比如说章节的数目用中文，但页码却不用。
2. 现在中文文章大多都是从左往右号的，而不是从上往下竖排的。
3. 目前世界上使用的汉字有两种，一是中国大陆使用的简体汉字，另一种是台湾和香港使用的繁体汉字。

Fonts

字体

It is no secret that Chinese has many glyphs. There are some public collections, like the one packaged in `mssong.ttf`. By splitting this True Type Font (`ttf`) into many small PostScript Font Binaries (`pfb`), accompanied by native TeX Font Metric files (`tfm`), we have them available in a format suited for TeX.

大家都知道汉字有很多图示符。现在有许多像 `mssong.` 这样的免费字体，通过将这些 `ttf` 分成很多 PostScript 二进制字体 `pfb`，加上相应的 `tfm` 字体，我们就能使 汉字能为 TeX 所使用。

original font	mssong.ttf
derived fonts	gbsong**.pfb
font metrics	gbsong**.tfm

The file `mssong.ttf` contains the so called SongTi Simplified Chinese fonts. In general, the Chinese truetype font files are very large, but the file `mssong.ttf` not that large.

Truetype 字体文件 `mssong.ttf` 包含了简宋体汉 字。一般来说，汉字文件都是很大的，比较而言， `mssong.ttf` 还算是比较小的。

mshei.ttf	1,902,464	...
mssong.ttf	2,569,040	Simplified Chinese
mingliu.ttf	6,272,080	Traditional Chinese

Instead of splitting such big files yourself, you can best try to get hold of the `pfb` and `tfm` files. Once you got these files, make sure you also have (or construct) a `map` file with lines like

你最好直接取得 `pfb`和 `tfm`文件，而不是自己来生成 它们。在取得 `pfb`和 `tfm`文件的同时，不要忘了取得 或自己生成包含如下内容的字体映射文件。

   gbsong81   GB–Song81
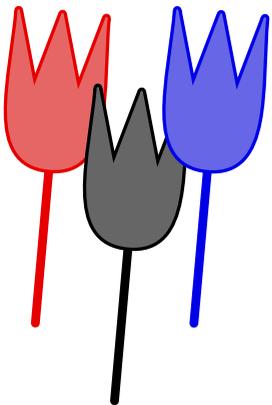
< gbsong81

```
gbsong82    GB−Song82                          < gbsong82
gbsong83    GB−Song83                          < gbsong83

gbsongslfc GB−Songfc ".167 SlantFont" < gbsongfc
gbsongslfd GB−Songfd ".167 SlantFont" < gbsongfd
gbsongslfe GB−Songfe ".167 SlantFont" < gbsongfe
gbsongslff GB−Songff ".167 SlantFont" < gbsongff
```

There are 126 fonts for each alternative style. When using PDFTEX, and given that these definitions are put in a file called chinese.map, don't forget to add the next line to the file pdftex.cfg:

```
map +chinese.map
```

每一种字体共有 126 个相应的 pfb 和 tfm 文件。当使用 PDFTEX 时，将上述字体映射定义放入一个叫作 chinese.map 的文件中，并且在 pdftex.cfg 中加入下面一行：
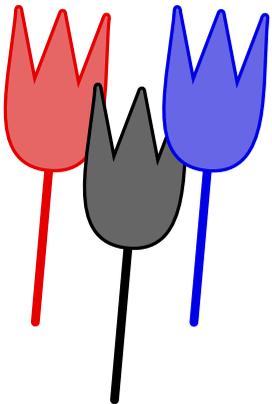
task force members	Matthew Baker
	Tobias Burnus
	Gilbert van den Dobbelsteen
	Hans Hagen
	Taco Hoekwater
dedicated mailing list	**ntg-context@ntg.nl**
contacting authors	**pragma@wxs.nl**
examples, manuals and code	**www.pragma-ade.nl**
	www.ntg.nl/context
CONTEXT mirrors	see www pages
authors	Hans Hagen
	王磊
processing date	November 2, 1999
current version	1999.10.28