

# 使用 gbk2uni 为 pdf 书签增加 GBK 汉字支持 \*

李树钧 (Shujun Li)  
hooklee@hooklee.com

Version: 0.22 build in 26 March 2003

## 摘要

cxterm 编写（经张林波教授修改）的 gbk2uni 工具用于为 hyperref 增加 GBK 中文支持，但是原来的 gbk2uni 不支持 hyperref 的 unicode 模式，而且对非 unicode 模式的支持也存在缺陷，故本人对 gbk2uni 做了改进，加入如下功能：

- 自动探测 unicode 模式，并同时支持这两种模式的 out 文件转换；
- 支持 unicode 特殊字符和 GBK 字符的同时显示；
- 支持 CJK 宏包提供的 \CJKchar{"0xx}{0xx} 命令，并修补了 pdf(la)tex 对该命令的自动过滤；
- 改进了对 cctconv 处理后的 tex 文档的支持。
- 增强了对.out 文件中的“8x9y”形式的字符解析。

本文是改进版本的 gbk2uni 工具的一个简要说明和演示。关于该工具的更多实现细节，请参考随 CCT 新版一起发布的源代码 gbk2uni.c。访问 [CTeX](#) 或 [ChinaTeX](#) 获取 gbk2uni 的最新信息。

## §1 gbk2uni 的使用方法

gbk2uni 是一个命令程序，用于将.out 文件中的 GBK 汉字和特殊符号转换为正确的 unicode 编码，以得到 pdf 文件中正确的 GBK 中文书签。该工具随张林波老师的 CCT 一同发布，部分功能需要 cctconv 工具配合，但是可以该工具可以脱离 CCT 环境与 CJK 中文宏包一起使用。该工具的使用方法可以在命令行敲入 gbk2uni 得到：

```
Usage : gbk2uni [options] filename[.out] [options] Options:
-u(-l)   lock .out file to avoid overwritten in the next (pdf)latex run
         (.out file will be unlocked if no '-u' and '-l' options)
-unlock  unlock .out file without parsing .out file
-s       run gbk2uni silently (but errors remain)
-cjk     parse \CJKchar{"0xx}{0xx} command (default)
-nocjk   disable parsing \CJKchar{"0xx}{0xx} command
-npe     disable display of all parsing errors
```

用张林波教授随 CCT 提供的 ctex 编译时，执行流程很简单：ctex -dvipdfmx foo 或者 ctex -pdflatex foo。如果运行一次得不到正确的书签，请再运行一次 pdflatex foo 或者 ctex -pdflatex foo。由于 dvipdfmx 对 hyperref 宏包的 bookmarksopen=true 似乎不支持，推荐使用后一种方法得到 pdf 文件，除非您想在得到的 pdf 文件中支持 GBK 中文搜索和拷贝功能。

当然，您也可以直接使用 latex/pdflatex 编译得到 pdf 文件。用 latex 编译时，执行流程如下：

```
latex foo => gbk2uni foo => latex foo => dvipdfmx foo
```

为得到书签，您需要为 hyperref 宏包声明 dvipdfm 任选项。用 pdflatex 编译时，执行流程如下：

```
pdflatex foo => gbk2uni foo => pdflatex foo
```

为得到书签，您需要为 hyperref 宏包声明 pdftex 任选项。

---

\*该文档使用 ctex -pdflatex test 编译。使用 ctex -no-cctspace -pdflatex test 编译将压缩英文和中文之间的空格。

## §2 gbk2uni 的部分演示

### §2.1 在 pdf 书签中同时使用 unicode 特殊符号与 GBK 汉字

本节演示在 pdf 书签中同时使用 unicode 特殊符号与 GBK 汉字。一般来说，大部分 unicode 特殊符号均可和 GBK 汉字一起显示在书签中，但是发现部分 unicode 特殊字符与 GBK 汉字似有冲突，当 out 文件中的内容完全正确时，GBK 汉字和某些特殊字符不能一同显示，不能显示的字符（或汉字）在 pdf 书签中一律显示为“.”，下面三个小节我们来演示一下这种问题。目前尚不知道如何解决这个问题，或许这是 acobat 对 unicode 书签支持上的一个 bug。

#### §2.1.1 $[\alpha, \beta]$ {原始社会}(来去)

$[\alpha, \beta]$ {原始社会}(来去)

#### §2.1.2 $[\acute{\alpha}, \beta]$ {原始社会}(来去)

$[\acute{\alpha}, \beta]$ {原始社会}(来去)

#### §2.1.3 $[\acute{\alpha}, \beta]$ {Society}(Come&Go)

$[\acute{\alpha}, \beta]$ {Society}(Come&Go)

#### §2.1.4 {原始社会} $[\alpha, \beta]$ (来去)

{原始社会} $[\alpha, \beta]$ (来去)

#### §2.1.5 {原始社会} $[\acute{\alpha}, \beta]$ (来去)

{原始社会} $[\acute{\alpha}, \beta]$ (来去)

#### §2.1.6 {Society} $[\acute{\alpha}, \beta]$ (Come&Go)

{Society} $[\acute{\alpha}, \beta]$ (Come&Go)

#### §2.1.7 {原始社会}(来去) $[\alpha, \beta]$

{原始社会}(来去) $[\alpha, \beta]$

#### §2.1.8 {原始社会}(来去) $[\acute{\alpha}, \beta]$

{原始社会}(来去) $[\acute{\alpha}, \beta]$

#### §2.1.9 {Society}(Come&Go) $[\acute{\alpha}, \beta]$

{Society}(Come&Go) $[\acute{\alpha}, \beta]$

### §2.2 使用 \CJKchar 命令

本节演示 gbk2uni 对 \CJKchar{"0xx"}{"0xx"} 命令的支持。出于技术上的考虑，gbk2uni 要求 \CJKchar 命令的两个参数必须以三位十六进制数的形式给出。

#### §2.2.1 丂

\CJKchar{"081"}{"040"}: 丂

### §2.3 低位字节为 T<sub>E</sub>X 特殊字符 “{\^\_~” 和 0x80 的 GBK 汉字的问题

当 GBK 汉字的低位字节为 T<sub>E</sub>X 特殊字符 “{\^\_~” 时，直接编译 tex 文档会出现错误，当 GBK 汉字的低位字节为 0x80 且 hyperref 使用了 unicode 任选项，pdf(la)tex 会出错。为解决上述问题，请使用随 CCT 发布的 cctconv 解决此问题，cctconv 将这样的汉字转换为 CJK 代码 ^^7f 的形式以克服这个问题，对这样的文档，您需要按照如下流程进行编译。

```
cctconv foo foo2 => pdflatex foo2 => gbk2uni foo2
=> pdflatex foo2 => rename foo2.pdf foo.pdf
```

为了避免手工运行的麻烦，我们强烈推荐您使用张林波老师随 CCT 提供的集成编译外壳程序 ctex，使用它我们只需运行一次或两次 ctex [-no-cctspace] -pdflatex foo 即可得到最终结果。

本节和下面的小节演示 cctconv 与 gbk2uni 协作生成低位字节为 T<sub>E</sub>X 特殊字符的 GBK 中文书签。

### §2.4 𠂇𠂈𠂉𠂊𠂋𠂌𠂍𠂎𠂏𠂐𠂑𠂒𠂓𠂔𠂕

𠂇 (0x8180) 𠂈 𠂉 𠂊 𠂋 𠂌 𠂍 𠂎 𠂏 𠂐 𠂑 𠂒 𠂓 𠂔 𠂕

### §2.5 𠂖𠂗𠂘𠂙𠂚𠂛𠂜𠂝𠂞𠂟𠂠𠂡𠂢𠂣𠂤

𠂖 (0xB080) 𠂗 𠂘 𠂙 𠂚 𠂛 𠂜 𠂝 𠂞 𠂟 𠂠 𠂡 𠂢 𠂣 𠂤

#### §2.5.1 𠂥𠂦𠂧𠂨𠂩𠂪𠂫𠂬𠂭𠂮𠂯𠂰𠂱𠂲𠂳

𠂥 𠂦 𠂧 𠂨 𠂩 𠂪 𠂫 𠂬 𠂭 𠂮 𠂯 𠂰 𠂱 𠂲 𠂳

#### §2.5.2 𠂴𠂵𠂶𠂷𠂸𠂹𠂺𠂻𠂼𠂽𠂾𠂿𠂿𠂿𠂿

𠂴 𠂵 𠂶 𠂷 𠂸 𠂹 𠂺 𠂻 𠂼 𠂽 𠂾 𠂿 𠂿 𠂿 𠂿

#### §2.5.3 𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿

𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿

#### §2.5.4 𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿

𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿

#### §2.5.5 𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿

𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿

### §2.6 更多的 GBK 汉字测试

本节直到本文的最后演示更多 GBK 汉字，以考察 gbk2uni 的稳定性。

#### §2.6.1 𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿

𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿

#### §2.6.2 𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿𠂿

𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿 𠂿

§2.6.3 亶亶亶亶亶亶亶亶亶亶亶亶

亶亶亶亶亶亶亶亶亶亶亶亶

§2.6.4 仕仕仕仕仕仕仕仕仕仕仕仕

仕仕仕仕仕仕仕仕仕仕仕仕

§2.6.5 物物物物物物物物物物物物

物物物物物物物物物物物物

§2.6.6 伋伋伋伋伋伋伋伋伋伋伋伋

伋伋伋伋伋伋伋伋伋伋伋伋

§2.6.7 佇佇佇佇佇佇佇佇佇佇佇佇

佇佇佇佇佇佇佇佇佇佇佇佇

§2.6.8 佖佖佖佖佖佖佖佖佖佖佖佖

佖佖佖佖佖佖佖佖佖佖佖佖

§2.6.9 佻佻佻佻佻佻佻佻佻佻佻佻

佻佻佻佻佻佻佻佻佻佻佻佻

§2.6.10 瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡

瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡瘡

§2.6.11 瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳

瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳瘳

§2.6.12 皦皦皦皦皦皦皦皦皦皦皦皦

皦皦皦皦皦皦皦皦皦皦皦皦

§2.6.13 盍盍盍盍盍盍盍盍盍盍盍盍

盍盍盍盍盍盍盍盍盍盍盍盍