

Linux操作系统中文字型的转换与内嵌 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/145/2021\\_2022\\_Linux\\_E6\\_93\\_8D\\_E4\\_BD\\_c103\\_145075.htm](https://www.100test.com/kao_ti2020/145/2021_2022_Linux_E6_93_8D_E4_BD_c103_145075.htm) 不论是 Postscript 文件或 PDF 文件，都可以直接内嵌所需的字型，可以只提示所需的字型名称与每个字在字型文件中的编码索引等。后者的档案大小当然比前者来得小，但前提是文件内所使用的字型名与字型规格要有一个通用的标准，如此该份文件才能在各种环境下阅读与打印。故一般只包含 ASCII 或 ISO8859-1 文字的 Postscript 文件是不内嵌字型的，但是一份中文文件在目前常见的情况下是必须内嵌字型的，其原因正是目前还没有通用的中文字型与规格。在早期（也许 PDF 格式尚未问世的时候），用做内嵌的中文字型来源是一种名为 HBF 的点阵字型，当时有许多程序可以将 HBF 字型转换成 Postscript 可以使用的格式，然而这类的点阵字型分辨率有限，故内嵌在 Postscript 文件中往往不够美观，特别是当我们要做有限度的文字缩放时更是惨不忍睹。后来，由于 TrueType 字型逐渐普及，同时可以处理 TrueType 字型的引擎 -- freetype 函式库的问世，许多基于该函式库的 TrueType 字型转换程序一一开发出来，使得它成为 Postscript 与 PDF 内嵌字型的最重要来源。内嵌了 TrueType 转换字型的文件不但比起过去要美观许多，且由于 TrueType 本身就具有可缩放的特性，故我们可以轻易产生高分辨率的转换字型，使得文件在有限度的缩放下仍不失真。同时由于文鼎公司对自由软件世界的支持，捐赠了两套 TrueType 中文字型供大家自由使用，让我们在中文文件输出问题上得以彻底解决。最佳的解决方案仍然是希望以不需内嵌中文字型的方

式，其优点除了可以让输出的文件档变小以外，同时还可以保证在任何尺度上的缩放而不失真。由于内嵌在文件中的字型在很多情况下是将 TrueType 字型转成点阵字型之后，才进行内嵌 (例如在 CJK-LaTeX 所编译的文件中内嵌了由 TrueType 转成 PK 的点阵字)。尽管由转换过来的字型有相当高的分辨率，但再高的分辨率也是有限的，故在做大尺度的缩放时仍然会失真。要做到真正不内嵌中文字型于文件中，有许多工作要做，其中包括：

1. 要有一套统一的字型名称与规格：这可能是最令人头疼的部分，因为这里卡到了授权问题。由于 Postscript 与 PDF 文件格式是 Adobe 公司开发的，故其所用的字型名称都会灌上 Adobe 之名，而这些名称是否能在自由软件世界中，以自由软件的游戏规则来运行，将会是个问题，其中尤以 PDF 最为严重。同时，在定这些名称与规格时，我们不能关起门来自己定，必须同时与许多团体谈，可能包括 Adobe 公司、TeX/LaTeX 开发团队、Ghostscript 开发团队、甚至其他与排版、字型相关的计划与商业公司等。
2. 要有一个字型的来源：这一点在现阶段已有初步成果，目前我们已有将 TrueType 字型转换成 Postscript 所需的 Type1 字型的工具程序，例如 ttf2pt1 或 chpfb,ttf2pfb 等。除此之外，另外还有一套称为 t1lib 的函式库，可以用来进一步处理 Type1 字型，例如将字缩放、旋转，或将它转换成点阵的格式以用于其它的用途 ... 等等。
3. 如果我们有了 Type1 的字型，我们还可以进一步地将它合成 Type0 的 CID-Keyed 字型，可用于 Postscript 与 PDF 的文件中。
4. 我们的文件转换引擎：Ghostscript 要能处理多字节编码、并利用这些字型：这一点目前也正积极开发中，例如下一代的 Ghostscript 就有处理多字节编码的能力，除

了迎接未来的中文 Type1 字型以外，它还内含了一个可直接使用 TrueType 字型的模块，可以做到不需要字型的内嵌，就可以使用 TrueType 字型来显示文件，其原理是将 TrueType 字型仿真成 Type0 CID-Keyed 字型来用。但这可能会面临可移植性的问题，万一将此文件拿到其它没有安装此 TrueType 字型的系统下就无法读了。因此，它目前也同时在开发直接内嵌 TrueType 字型的技术，做为因应此问题的配套措施。中文 Postscript 与 PDF 文件的产生目前较通用的纯文字文件转 Postscript 格式的程序为由 python 语言所写成的 bg5ps 程序，它所产生的 Postscript 文件采内嵌中文字型的方式，而中文字型的来源即为 TrueType 字型。CJK-LaTeX 是一套功能相当强大的幕后排版系统，它使用类似程序语言的方式，来格式化并编排文件，并配合 Postscript 的输出，可以轻易达到高品质文件的输出要求。特别是它容易处理数学符号的特性，使得它特别适于学术文件、书籍、出版的排版工作。要让 LaTeX 能处理中文，必须在 LaTeX 系统上额外加装一套 CJK-LaTeX 的宏。顾名思义，此宏可以付与 LaTeX 处理中文、日文、韩文等编码系统的能力，而它也渐渐成为 LaTeX 的标准之一。而它最后在产生 Postscript 文件时，是采用内嵌中文字型的方式，其字型来源也是 TrueType。它利用 freetype 函式库的衍生工具 ttf2tfm、ttf2pk 将 TrueType 字型转换成 TFM 与 PK 点阵字型，前者用于前段的排版工作，后者则用于最后的字型内嵌。LaTeX 的外衣 -- LyX LyX 是一个以 LaTeX 为基底的一个图形化排版系统，让我们不必须记一大堆 LaTeX 的排版指令，可以直接可视化地在窗口内编排我们的文件，而最后的排版输出才交由 LaTeX 处理。它的中文支持目前正慢慢发展中

，包括支持 XIM 协议，使得我们可以用 XIM server 输入法程序直接在它的窗口内打中文。而它的 CJK-LaTeX 指令的部分在经过少许的修正后也能正确地搭配使用。100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)