

glc-win.sed 変換内容

松浦高志

2006 年 12 月 7 日

拙作の glc-win.sed スクリプトを用いると、Windows の機種依存文字^{*1} を、主に Adobe-Japan1-5 のグリフに変換することができる^{*2}。変換は sed コマンドを用いて次のように行う。UNIX で処理することを考え、改行コードは UNIX 標準の LF にしてある。pL^AT_EX で処理する際には utf/otf パッケージを用いればよい。

```
$ sed -f glc-win.sed input.txt >output.txt
```

このスクリプトによって変換される内容は次の通りである。かっこ付数字・点付数字・かっこ付アルファベットなどは、使う箇所にあわせて半角アルファベットと半角または全角のかっこ類の組合せに変更することが望ましい。

単位記号 通常のアルファベット（と上付数字）に変換する。

mm cm km mg kg cc m²

ローマ数字 ローマ数字に関しては Adobe-Japan1-5 のグリフではなく、半角ローマ文字に変換する。

I II III IV V VI VII VIII IX X

i ii iii iv v vi vii viii ix x

丸付数字

各種記号 KK (上) (中) (全) (右) (株) (有) (代) 明治 昭和 平成

^{*1} 「Windows の機種依存文字」とは、文字コードは Shift-JIS (SJIS) であるが、SJIS で規定されていない空きコードに Microsoft が独自に規定した文字のことである。Macintosh や UNIX では表示されなかったり、全く別の文字として表示される。

^{*2} 比較的新しい Adobe Reader や Mac OS X の「プレビュー」などの PDF 閲覧ソフトなら問題なく表示される。