

JFM ファイルフォーマット

ASCII Corporation & Japanese T_EX Development Community

2018 年 8 月 18 日

1 JFM file format

JFM ファイルのフォーマットは、基本的には TFM ファイルのフォーマットに準拠しており、TFM を拡張した形になっている。ここでは、主にその拡張部分について説明を行い、その他の部分に関しては、T_EX the program 等の TFM の説明を参照してもらいたい。

この文書はもともと株式会社アスキーが開発した pT_EX に付属していたものである。2018 年に日本語 T_EX 開発コミュニティは JFM フォーマットを拡張したため、その内容について追記してある。

1.1 JFM ファイルの構成

JFM ファイル全体の構成は、Table 1(5 ページ) に示すとおりである。ここで TFM と異なるのは次の点である。

1. *char_type* のテーブルが付け加えられたこと。
2. *exten* の代わりに *glue* のテーブルが設けられたこと。
3. 2 に関連して、*lig_kern* から *glue_kern* テーブルへ変更されたこと。
4. これらに伴い、先頭のファイル内の各部分を規定するパラメータ表が変更されている。また、オリジナルの TFM との区別のために *id* を付加している。

最初の 7 ワードは半ワード (= 2 バイト) ずつに区切られ、JFM ファイルを構成する 14 個の要素のサイズが取められている。これらの値は、すべて 2^{15} よりも小さい非負の値で、次の条件を満たしていなければならない：

$$bc = 0$$

$$lf = 7 + lh + (ec - bc + 1) + nw + nh + nd + ni + nl + nk + ng + np$$

1.2 char_type

今回の \TeX の日本語化においては、同一の文字幅、高さ前後に挿入されるグルー等、その文字が持つ属性全てが同じものを、1つの '*char_type*' として英字フォントの1キャラクタと同様にして扱うようにしている。そこで、漢字の文字コードとこの *char_type* との対応付けをこのテーブルを使って行う。

このテーブルの各エントリーは1ワード (= 4バイト) で構成され、上位3バイトに漢字コード、下位1バイトにキャラクタタイプを持つ¹。漢字コードは、それが16進数24bit (3バイト) で `0xABcdef` と表される時、*char_type* テーブルには `cd ef AB` として格納される²。テーブル内にはコードの値の順番に収められていなければならない。またこのテーブルの先頭には、デフォルトのインデックスとして漢字コード及びキャラクタタイプの項が0のものが、必ず1つ存在しなければならない。このテーブルに登録されていない文字は、キャラクタタイプが0として扱う。つまり、このデフォルト以外の文字幅、カーン等の属性を持つキャラクタのコードとタイプが2番目以降のエントリーとして存在しなければならない。

1.3 char_info

char_type をインデックスとしてこのテーブルを参照することにより、各 *char_type* の属性を検索する。各テーブルへのインデックス等の情報を次の順番でパッキングして1ワードに収めてある。

width_index (8bits) *width_table* へのインデックス

height_index (4bits) *height_table* へのインデックス

depth_index (4bits) *depth_table* へのインデックス

italic_index (6bits) *italic_table* へのインデックス

tag (2bits)

tag = 0 *remainder* の項は無効であり使用しないことを示す。

tag = 1 *remainder* の項が *glue_kern* への有効なインデックスであることを示す。

tag = 2, 3 JFM では使用していない。

remainder (8bits)

JFM では *bc* は必ずゼロ³なので、1つの JFM に含まれる *char_info* は全部で *ec* + 1 ワードになる。

¹これが日本語 \TeX 開発コミュニティによって2018年1月に導入された新仕様である；オリジナルのアスキーによる仕様では、「上位半ワードに漢字コード、下位半ワードにキャラクタタイプを持つ」というものであった。p \TeX では内部処理に JIS が用いられ、JFM で扱う漢字コードは2バイトが上限だったため十分であったが、up \TeX で BMP 超えの3バイトの文字を JFM で扱うことを目的に仕様拡張した。

²オリジナルの仕様では、下位半ワードにキャラクタタイプを格納していたが、キャラクタタイプの上限は255なので、実はその上位バイトは常に00であった。このことを利用し、日本語 \TeX 開発コミュニティの新仕様では「この00が実は漢字コードの上位だった」ということにして、3バイト (U+10000以上) の漢字コードで不足する1バイトを確保した。これにより、新仕様はオリジナルの仕様の上位互換であることが保証されている。

³前節にある通り、漢字コード及びキャラクタタイプの項が0のものが必ず1つ存在するため。

1.4 glue_kern

特定のキャラクタタイプの組み合わせ時に挿入すべき *glue* 又は *kern* を簡単なプログラム言語によって指定する。各命令は、以下の 4 バイトで構成される。

第 1 バイト (*skip_byte*)

- 128 より大きいとき
現在のワードが *char_info* から示された最初のワードである場合は、実際の *glue_kern* プログラムが $glue_kern[256 * op_byte + remainder]$ から収められている (すなわち、再配置されている) ことを示す⁴。最初のワードでない場合 (すなわち、既に再配置先あるいはプログラムのステップを開始した後のワードである場合) は、その場でプログラムを終了する。
- 128 のとき
このワードを実行してプログラムを終了する。
- 128 より小さいとき
このワードを実行した後、次のステップまでスキップするワード数を示す⁵。

第 2 バイト (*char_type*)

- 次の文字のキャラクタタイプが、このバイトで示すキャラクタタイプ⁶と同じ場合、第 3 バイトの処理を実行し、プログラム終了。
- そうでなければ次のステップへ。

第 3 バイト (*op_byte*)

この値によってグルーを扱うかカーンを扱うかを規定する。

- 127 以下の場合 $glue[remainder \times 3]$ のグルーを挿入。
- 128 以上の場合 $kern[remainder]$ のカーンを挿入。

第 4 バイト (*remainder*)

第 3 バイトにより規定される *glue* または *kern* へのインデックスを示す。

1.5 glue テーブル

自然長、伸び長、縮み長の 3 ワードで 1 つのグルーを構成する (したがって、*ng* は必ず 3 の倍数となる)。各値は、 $designsize \times 2^{-20}$ を単位として表す。

⁴「再配置」は、サイズが 256 を超える大きな *glue_kern* テーブルを格納するための方策であり、欧文 TFM の *lig_kern* テーブルにおけるそれと同様である。2018 年 2 月に日本語 TeX 開発コミュニティによって、pTeX 及び pPLtoTF で新たにサポートされた。

⁵「スキップ」(SKIP) は、元々アスキーの公式ページ <http://ascii.asciimw.jp/pb/ptex/tfm/jfm.html> に文書化されていたが、実際には (p)PLtoTF の *glue_kern* テーブル内で SKIP 命令を使用することができず、pTeX もやはり SKIP 命令をサポートしていなかった。2018 年 2 月の日本語 TeX 開発コミュニティの改修により、新たにサポートが開始された。

⁶ここにキャラクタタイプが格納されるため、キャラクタタイプの上限は 255 なのである。

第 1 ワード width

第 2 ワード stretch

第 3 ワード shrink

1.6 param テーブル

param[1] 文字の傾き (italic slant)。

param[2][3][4] 漢字間に挿入するグルーのデフォルト値。

param[5] pTeX で zh で参照される寸法。

param[6] pTeX で zw で参照される寸法。

param[7][8][9] 漢字と英字間に挿入するグルーのデフォルト値。

<i>id</i>	<i>nt</i>
<i>lf</i>	<i>lh</i>
<i>bc</i>	<i>ec</i>
<i>nw</i>	<i>nh</i>
<i>nd</i>	<i>ni</i>
<i>nl</i>	<i>nk</i>
<i>ng</i>	<i>np</i>
header	
char_type	
char_info	
width	
height	
depth	
italic	
glue_kern	
kern	
glue	
param	

id = JFM.ID number. (= 11 for yoko, 9 for tate)
nt = number of words in the character type table.
lf = length of the entire file, in words.
lh = length of the header data, in words.
bc = smallest character type in the font. (= 0 for JFM)
ec = largest character type in the font.
nw = number of words in the width table.
nh = number of words in the height table.
nd = number of words in the depth table.
ni = number of words in the italic correction table.
nl = number of words in the glue/kern table.
nk = number of words in the kern table.
ng = number of words in the glue table.
np = number of font parameter words.

表 1: JFM ファイルの構成