

事実と規則による情報交換用符号系への変換
- 2段階テーブルルックアップ法 -

若鳥陸夫 (日本ユニバック)

1. 背景

機種・オペレーティングシステム・言語処理系を問わない情報交換の基本は情報交換用漢字符号系を用いることにある。たとえば「開いたマイコン間通信」において、日本語の再現性を得る最もオーソドックスな方法は、伝送回線上の符号系を日本工業規格の情報交換用漢字符号系と定め、各端末の文書準備及び表現は内部符号系のままとし送受信時にそれを情報交換用漢字符号系と変換することがある。ところが、パーソナルコンピュータの内部符号系として多用される、シフトジス群 (CP/M 86シフトジス78, CP/M 86シフトジス83, MS-DOSシフトジス78, MS-DOSシフトジス83, シフトジス汎用漢字コード等) は、シフトジス群と情報交換用符号系との相互変換のアルゴリズムに最大6回の加減算等を必要 [図1] とし、高速伝送の場合やマルチタスキングで変換所要時間を切りつめたい要望等のとき、その変換速度がボトルネックとなり易い。また、その変換ルーチンを機械語で書いたり、オペレーティングシステムや言語処理系の組み込み関数とするケースもあるが、それらは一般的な高水準言語とそれらの特殊技法とを併用するのは一貫性を失うし、また併用の道もLISPやPrologなどの高水準言語処理系に開かれていないし、たとえ使うにしても繁雑な手続きを要する。

ここで紹介する方法は、そのようなユーザに簡便で、処理時間の速い変換方法を提供するものである。そのアイデアは漢字14ビットを上・下2バイトに分け、その各バイトの変換計算を予め行なって変換表としておき、変換時には、2段階のテーブルルックアップにより行なわせるものである。この方法によれば、比較的処理速度の遅いBasic言語・Prolog言語あるいはLisp言語などのインタプリタで変換する際や、高速伝送中に実時間で符号変換する時にも必要な高速性を維持させることができる。

また、本変換法によれば、シフトジス汎用漢字コード [文献1] のように多少変換規則が複雑であっても、2段階テーブルルックアップで短時間で処理できる。さらに、ここで実験したCP/M 86シフトジス78は高位バイトの奇数偶数や大小により、変換表のスイッチを要するが、1回のテーブルルックアップで変換できる符号割り付出現を待つものである。

ここでは、その変換法を規則とPrologプログラムで示したが、通信用LSIや通信ルーチン等へ形を変えて実現すると効果的である。

2. 変換表の規則

内部符号系CP/M86シフトジス78と情報交換用漢字符号系（第一水準及び非漢字）JISC6226(83)の間の符号系を変換する規則及び事実は次の16通りである〔文献2, 4〕。

(1) JISからCP/M86シフトジス78への変換

- (イ) 規則 1: 漢字呼び出しエスケープシーケンス(1/11, 2/4, 4/2)がきたらANKモードから漢字モードへ切り換える。
- (ロ) 規則 2: ANK呼び出しエスケープシーケンス(1/11, 2/8, 4/10)がきたら、漢字モードからANKモードとする。
- (ハ) 規則 3: ANKモードなら無変換とする。
- (ニ) 規則 4: 漢字モードで第1バイトが21h~4Fhで、奇数の場合の第1バイトは、 $(\text{第1バイト} + 1) \div 2 + 70\text{h}$ とする。
- (ホ) 規則 5: 漢字モードで第1バイトが21h~4Fhで、偶数の場合その第1バイトは、 $(\text{第1バイト} + 1) \div 2 + 6\text{Fh}$ である。
- (ヘ) 規則 6: 漢字モードで第1バイトが奇数で、かつ60hより小なら、 $(\text{第2バイト}) + 1\text{Fh}$ を第2バイトとする。
- (ト) 規則 7: 漢字モードで偶数で、かつ60hより小なら、 $(\text{第2バイト}) + 20\text{h}$ を第2バイトとする。

(2) CP/M 86 シフトジス 78 から JIS への変換

- (イ) 規則 8 : 第 1 バイトが 81 h ~ 9F h または A0 h ~ FCh なら、漢字とみなす。
- (ロ) 規則 9 : ANK から漢字モードへ切り換わったとき、1 / 11, 2 / 4, 4 / 2 を挿入する。
- (ハ) 規則 10 : 漢字モードから ANK への切り換えは 1 / 11, 2 / 8, 4 / 10 を挿入する。
- (ニ) 規則 11 : 漢字モードで 2020 h なら 2121 h とする。
- (ホ) 規則 12 : 漢字モードで、第 2 バイトが 9F h より小なら、
新第 1 バイト = (第 1 バイト - 70 h) - 1 とする。
- (ヘ) 規則 13 : 漢字モードで、第 2 バイトが 9F h 以上なら、
新第 1 バイト = (第 1 バイト - 70 h) × 2 - 1 とする。
- (ト) 規則 14 : 漢字モードで、第 2 バイトが 80 h より小なら、
新第 2 バイト = 第 2 バイト - 1F h とする。
- (チ) 規則 15 : 漢字モードで、第 2 バイトが 80 h 以上で、かつ 9F h より小ならば、
新第 2 バイト = 第 2 バイト - 20 h とする。
- (リ) 規則 16 : 漢字モードで、第 2 バイトが 80 h 以上で、かつ 9F h 以上ならば、
第 2 バイト = 第 2 バイト - 7F h とする。

図形文字	情報交換用漢字符号		変換計算	内部符号系	
	16進表現	10進表現		10進表現	16進表現
亜	3021 h	4833	$\begin{array}{r} 48 \quad 33 \\ -) \quad 1 \quad +) 31 \\ \hline 47 \quad 64 \\ \div) \quad 2 \quad +) 94 \\ \hline 23 \quad 158 \\ +) 113 \quad +) 1 \\ \hline 136 \quad 159 \end{array}$	$\begin{array}{cc} 136 & 159 \\ \uparrow & \uparrow \\ & \end{array}$	889F h

図 1 変換計算例

3. プログラムの見本

図2にmicro-PROLOGで記述した変換表生成部分を，図3に変換部分のアルゴリズムを示す。この変換部分では3バイトエスケープシーケンスの処理を含め，JISC6226の要件(4章)を満たしている。

4. 所要メモリ量

変換表に要するメモリ量はPascalのように部分範囲型として下限を指定できる場合と，BasicやFortranのように要素0又は1から最大要素数まで配列変数として準備する方法で変化し，さらに文字型か整数型により現実の必要メモリ量に変化するが，ここではPascalで整数の範囲型の配列で構成した場合の例を第1表に示す。

表1 変換表のメモリ量の目安

名 称	要 素	要素数
ジス高奇数	81h ~ 9Fh	159
ジス高偶数	81h ~ 9Fh	159
ジス低	40h ~ 4Fh	252
シフト高	21h ~ 4Fh	79
シフト低奇数	21h ~ 7Eh	126
シフト低偶数	21h ~ 7Eh	126
合計要素数		901

5. 結論と今後の発展

「2段階テーブルルックアップ法」による符号変換により、メモリを1KB程度犠牲にして、高速度を得る変換手段を紹介した。これにより、たとえば通信における符号変換のように変換速度を重視する応用に対しユーザを救う道を提供した。しかし、このシフトジス内部符号系の変換例では高位バイトの奇数・偶数による変換テーブルの選択を含んでおり、もっと高速化するには、1種類のテーブルルックアップで変換完了する性質をもつ内部符号系への検討も必要であろう。シフトジス汎用漢字符号系は、符号の値の大小により変換表を選択しなくても良い内部符号系の一例である。ともあれ、この変換方式はシフトジスという内部符号系を高速変換できないエンドユーザを救うであろうし、今後の内部符号系の採用にあたって、変換表の選択なしで情報交換用漢字符号系へ変換できる内部符号系の出現や通信用LSI内への変換組み込みの糸口となれば幸いである。参考までに示すと、筆者はこの「2段階テーブルルックアップ法」のアルゴリズムを、Bulletin Board Systemや漢字端末プログラムの送受信部に組み込んで実用試験中だが、実行速度の遅い処理系でも、300BPSでの実用時間符号変換に追随している。

謝辞：本論文の作成過程において、討論いただいた弊社森宗正氏、現実のプログラム作成及び論文を清書いただいた石川睦子氏に感謝する。

- 文献 (1) 木下；”標準プログラム言語における日本語処理”，情報処理，Vol. 26 No. 3, PP 232 (1985)
- (2) Digital Research Inc, マイクロソフトウェア・アソシエイツ；”標準漢字CP/M86 Version 1.1, 7/14/82
- (3) 福山，小林，伊集院，木下，落合；”Ada日本語処理機能の実現法”情報学会第30回全国大会論文誌(I)，PP 459 - PP 460 (1985)
- (4) 日本工業標準調査会；JISC 6226 - 1983, 情報交換用漢字符号系”，日本規格協会 (1985)
- (5) 日本ユニバック(株)；UNIVAC Personal Computer, UP10E50 操作解説書

```

((RUN)
  (CREATE "CODETAB.LOG")
  (Initialize)
  (CLOSE "CODETAB.LOG")
((Initialize)
  (JISHightable 129 X)/
  (JISLowtable 64 Y)/
  (ShiftHightable 33 Z)/
  (ShiftLowtable 33 x))
((event X1 Y 0)
  (NOT INT X1)
  (INT X1 y)
  (SUM y 111 Y))
((oddt X1 Y 1)
  (INT X1)
  (SUM X1 112 Y))
((ShiftLowtable 127 X))
((ShiftLowtable X Y)
  (SUM X 31 Z)
  (SUM X 126 y)
  (OR ((LESS X 96) (SUM 0 Z Z1)) ((SUM 1 Z Z1)))
  (WRITE "CODETAB.LOG" (((ShiftL1 X Z1))))
  (WRITE "CODETAB.LOG" (((ShiftL2 X y))))
  (SUM X 1 Y)
  (ShiftLowtable Y z))
((ShiftHightable 80 X))
((ShiftHightable X Y)
  (SUM X 1 Y) (TIMES Y 0.5 X1)
  (OR ((oddt X1 Y1 x)) ((event X1 Y1 x)))
  (WRITE "CODETAB.LOG" (((ShiftH X Y1 x))))
  (ShiftHightable Y z))
((JISLowtable 253 X))
((JISLowtable X Y)
  (OR ((LESS X 128) (SUM X -31 Z))
    ((OR ((LESS X 159) (SUM X -32 Z)) ((SUM X -126 Z))))))
  (WRITE "CODETAB.LOG" (((JISL X Z))))
  (SUM 1 X Y)
  (JISLowtable Y x))
((JISHightable 160 X))
((JISHightable X Y)
  (SUM X -112 Z)
  (SUM Z Z x)
  (SUM -1 x y)
  (WRITE "CODETAB.LOG" (((JISH1 X y))))
  (WRITE "CODETAB.LOG" (((JISH2 X x))))
  (SUM X 1 Y)
  (JISHightable Y z))

```

図 2 変換表生成プログラム

```

((COMMENT -----))
((COMMENT ..... SHIFTJIS TO JIS CONVERTER .....))
((COMMENT ..... 作者: 若鳥陸夫 .....))
((COMMENT -----))
(jconv X Y)
(/* X = 原始ファイル. Y = 目的ファイル)
(OPEN X)
/
(CREATE Y)
/
(NOT 変換 X Y Z x)
/
(CLOSE X)
(CLOSE Y)
(変換 X Y Z x)
(GETB X y)
(OR ((OR ((LESS y 129) ((LESS 159 y))) / (漢字アウト Y y Z x))
((LESS 128 y) (LESS y 160) / (GETB X z) /
(漢字イン Y y z Z x) (漢字符号変換 Y y z)))
/
(変換 X Y x X1))
((漢字符号変換 X 32 32)
(PUTB X 33)
(PUTB X 33)
(P ~ .))
(漢字符号変換 X Y Z)
(OR ((LESS Z 159) (JISH1 Y x)) ((JISH2 Y x)))
(JISL Z y)
(PUTB X x)
(PUTB X y)
(P Y + Z .))
(漢字イン X Y Z x y)
(LESS 128 Y)
(LESS 63 Z)
/
(漢字エスケープシーケンス X x y))
(漢字イン X Y Z x y)
(LESS Y 130)
/
(漢字アウト Y Z x y))
(漢字アウト X Y 0 0)
(PUTB X Y)
(P Y .))
(漢字アウト X Y 1 0)
/
(PUTB X 27)
(PUTB X 40)
(PUTB X 74)
(P 漢字アウト .)
(PUTB X Y)
(P Y .))
(漢字エスケープシーケンス X 1 1))
(漢字エスケープシーケンス X 0 1)
/
(PUTB X 27)
(PUTB X 36)
(PUTB X 66)
(P 漢字イン .))

```

図3 内部符号系から情報交換用漢字符号系変換プログラム