

外国製計算機への日本語文の埋め込み

4K-11

大石 東作 内堀 義信 末広 尚士

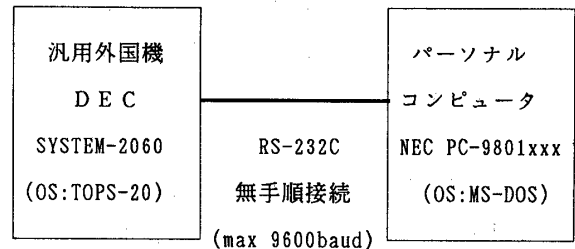
(電子技術総合研究所)

1、はじめに 外国製計算機を使用していて最も不便を感じるのは、日本語文の処理が困難な点である。我々は、日本語が全くサポートされていない外国製計算機と漢字処理が可能なパーソナルコンピュータを組合せることで、日本語文による電子メール・マニュアル・応用プログラムのユーザインタフェース等を利用可能とした。漢字コードとしては、7bit系のJIS C6226コードを用い、漢字シフトシーケンスに独自のものを工夫した。この方法によれば、外国製計算機においてもOSソフトウェアに手を入れることなく、日本語文が利用できる。

2、外国製計算機上での日本語処理 汎用の外国製計算機を使用していると、そのユーザインタフェースの良さや、ネットワーク機能・メール機能の充実には満足させられるが、一方日本語処理がサポートされていないのは、何といっても不便である。プログラム言語まで日本語化しようとは思わないが、日本語電子メール機能の実現は、研究者間の情報交換等にも不可欠になっている。そこで、計算機メーカーからのサポートが期待できない状況で、日本語処理を実現することを試みた。

汎用外国機に日本語処理全体を搭載することは、短時間には不可能と考えられる。そこで、第1図に示すように汎用外国機とパーソナルコンピュータとを組合せ、全体として日本語処理を実現している。漢字処理については、同処理の進歩が著しいパソコン側にほとんどをまかせ、汎用機側での漢字処理に関する手当ては最小限にとどめた。

3、漢字コードと汎用機上での処理 漢字コードに、情報交換用漢字符号系JIS C6226(1983)を用いることは基本的に問題はないが、汎用機上でのバイトの表現法には2種類の体系が考えられる。7bitと8



処理機能: メール、VTxxエミュレータ
ユーティリティ、ネット接続 漢字処理機能

第1図、汎用機とパーソナルコンピュータの処理の分担

bitの体系である。unix系統の日本語処理やパーソナルコンピュータでは、8bitの体系(いわゆるシフトJIS符号やその他)が用いられることが多い。しかし、我々は、以下の2つの理由で7bit体系を採用した。

- ①汎用機上でのメールシステムを含むほとんどのユーティリティが、7bit ASCIIコード系で実現されており、8BIT体系とは整合性が保てない。
- ②8bit体系は、JISやISO規格に合致するものではない。また、今後増加すると考えられる国際ネットワークを経由する日本語情報の交換に対応が困難である。

次に問題となったのは、7bit体系におけるエスケープシーケンスの問題である。情報交換用符号の拡張法(JIS C6228)に素直に準拠できればよいのであるが、これが不可能であった。漢字シフト用の、ESC符号(1BH)が汎用機上での制御用符号であるため、透明性を欠いているからであった。特にESC符号は、汎用機のメールシステムにおいて、入力文のEOF符号としても用いられており、この符号を使用するがぎり、オンラインで日本語メールを発信することは不可能である。EOF符号を変更することは、メールシステムのユーザインタフェースを変更することになり、同システムの従来からのユーザに負担を強いることになる。よってEOF符号は変更しなかった。かわりに漢字シフト用

のESC符号を、他の制御用符号に変更することにした。

汎用機上で、相当数の制御用符号に関して調査した結果、SYN符号(16H)の入出力シ

Embedding Japanese to Foreign Computers

Tosaku OISHI, Yoshinobu UCHIBORI, Takashi SUEHIRO
ELECTROTECHNICAL LABORATORY

システムに対する透明性が比較的に高いことが判明した。そこで、このSYN符号を漢字シフト用のコードとし、TOPS-20漢字コードと称することにした。SYN符号をシフトコードとして採用することは、前述のJISC6228規格を満たさないことになるが、上記の理由でやむをえなかった。この非標準的な漢字シフトコードへの対応は、パーソナルコンピュータ上の端末エミュレータで実施する。また、非標準的な漢字体系は、1種類の汎用機上で利用に限定される限り特に問題を発生しないが、外部世界との情報交換が必要になると、直ちに困難に陥る。そこで、比較的多く使用されている複数の7bit漢字体系をも汎用機上でサポートすることにした。新・旧JISコードとNEC-PCコードである。以上の各漢字コードのシフトシーケンスを表1に示す。ただし、オンラインあるいはそれに近い形態でメールを発信できるのは、依然 TOPS-20漢字コードのみである。

漢字コード	漢字イン	漢字アウト
TOPS-20	SYN,K	SYN,H
NEC-PC	ESC,K	ESC,H
旧JIS	ESC,\$,0	ESC,(,H [ESC,(,J]
新JIS	ESC,\$,B	ESC,(,J

表1、漢字シフト・シーケンス(JIS C6226,7bitコード用)

以上のように工夫したが、汎用機システムの制御用符号に対する透明性がいまだ低く、メールシステムにおける受信メールの表示時に、SYN符号等が特定の印刷文字に変換されてしまい、日本語メールを読むことが不可能であった。そこでアセンブリ言語で記述されたメールシステムプログラムの入出力部に手を加え、制御用符号の透明性を高めることで、汎用機上でサポートする全ての漢字コードによる日本語メールを読むことができた。これでJUNETから転送されてくる旧JIS日本語メールへも対応可能となった。

この他に汎用機上のユーティリティとして、全ての漢字コードをTOPS-20漢字コードに変換して端末エミュレータに送出する KTYPE,それぞれの漢字コードをそのまま送出

する CATK、パソコン上の漢字ファイルを汎用機上のファイルにアップロードする GETK,各漢字コードを相互に変換するTOPJIS,JISTOP,NECTOP 等を作成した。

4、パーソナルコンピュータ上での処理

パーソ

ナルコンピュータ上では、ほとんど全ての日本語処理機能とVT100(52)端末エミュレータ機能を主としてサポートしている。日本語ワードプロセッサや日本語フロントエンドプロセッサには、市販の物を利用した。これらの分野の市場では激しい競争が続けられているため、各自の好みに近い優れた機能の物が選択できた。

端末エミュレータには、汎用機側に対応した特殊な処理が要求されるため、独自のものを開発した。前述の7bit系漢字コードの表示や変換の他に、8bit系漢字コード(shift-JIS,DEC VAX/VMS)への対応も可能である。漢字コードは、エミュレータの起動時と実行時に任意に設定できる。この他の、ファイルのアップ・ダウンロード、端末セッションのロギング、回線属性設定、日本語フロントエンドプロセッサ機能の取込、等通常の端末エミュレータに必要なとされる機能は、全て実現している。なお、TOPS-20上の日本語メールの読み取りは、SORD MX68,Apple Macintosh,Symbolics 36xx でも可能になっている。

5、おわりに

外国製汎用計算機への日本語文の

埋め込みにあたり、パーソナルコンピュータ側の日本語処理機能を最大限に生かすことにより、汎用機上での開発手数をほぼ最小限にとどめることができた。制御用符号のシステムに対する透明性の問題は、当該汎用機種にのみに限定されるものではない。したがって、本論文で述べた日本語文の埋め込み法は、多くの外国製計算機に適用可能な一般的な方法と考えられる。最後に残された課題は汎用機上に、高機能な日本語エディタを開発することである。

末筆ながら研究の機会ををあたえられた棟上ソフトウェア部長、柏木電子計算機部長、白井制御部長、藤村情報システム研究室長、弓場計算機方式研究室長、赤堀システム制御研究室長、討論・協力して下さった皆さんに感謝する。