

文字もじMOJIの世界

4. 創世記におけるソフトウェアと文字

田丸 健三郎*

文字といえば多くの方がフォント、文字コードを連想するのではないだろうか。1990年頃は文字を取り巻く状況は今とは大きく異なっていた。

私は1992年以降マイクロソフトに所属している。米国ワシントン州シアトル市と言えばご存知の方も多いと思うが、マイクロソフトの本社はそのシアトルではなくワシントン湖対岸のレッドモンド市に存在する。私は所属していた研究開発部門でメッセージングシステム、ディレクトリサービスの開発立ち上げからリリースまでを主に担当していた。

この時に担当していたメッセージングシステムがExchange Serverの最初のバージョン（プロジェクト名：Touchdown）であり、現在はマイクロソフトの主要クラウドサービスMicrosoft Office 365の統合コミュニケーションサービスのメール機能として提供されているものである。当時の電子メールは今日のように身近なものではなく、また使用できる文字にもさまざまな制約が存在していた。メール本文はPlain Textに限られUnicodeを扱えるOSもアプリケーションもほとんど存在していなかった。この時に我々が目指したのは世界の人々をメールにより繋ぐことである。

Exchange Server最初のバージョンを開発するにあたり下記の方針を掲げていた。

◇言語にとらわれないグローバルコミュニケーションの実現

◇文字による、より多くの情報伝達の実現

当時はUS-ASCII、ISO-8859-1などの異なる文字集合を同時に扱えるメールシステムは存在せず、それを可能にするシステムの実現を目指した。今日「グローバルコミュニケーション」という言葉からは急速な高度化を成し遂げている機械翻訳を連想するが、Exchange Serverの開発初期は機械翻訳どころか依然として文字コード、文字集合がメールをはじめとする文字情報の交換において大きな課題であった。

この異なる文字集合の混在を可能にする手段としてマイクロソフトが選択したのがメール本文にRich Text Format、ディレクトリ情報では文字コード情報付き文字列である。

マイクロソフトはそれまでもTrueTypeなどWindows上のグリフ、文字のレンダリングについては積極的に取り組んでいたが、上位レイヤーにおける文字情報の交換においては具体的な解決策を見いだせずにいた。Exchange Serverの開発において特に電子

メールにおける表現力の拡張を目指した。

◇Rich Text Formatにおける言語タグの拡張

言語タグを拡張することによりテキストにおける使用言語の特定を可能にする。当時は動的な言語認識にはまだまだ技術的な制約があり、入力時のキーボードレイアウト情報と入力された文字との紐づけを行うことによる動的な言語、文字コード選択の実現を試みた。

◇Rich Textを扱えるコントロールの提供

今でこそRTFやHTMLを扱えて当たり前だが、当時はCPUの速度、メモリサイズなどからRich Textを一般的なコンピューターで描画、編集できるようにすることは非常に困難であった。急速に進歩するパソコンの近い将来を想定し、RichEditコントロール（Windows上の入力フィールドでRich Textを扱えるようにしたもの）の開発を行い、Windows 95、Windows NT 3.5に標準搭載した。

当時はUnicodeのようにさまざまな文字を一元的に扱える文字集合が普及しておらず、また多くのシステムが異なる文字集合、かつ8ビットという限られたアドレス空間を符号化に使用していた。

このことが多くの困難を生んだ。日本語だけでもEUC-J、シフトJIS、ISO-2022-JPを代表とする異なる文字コードを同一システム内で扱えるようにしなくてはならなかった。上記の2つ、および文字コードと文字列をペアとして

*TAMARU, Kenzaburo

日本マイクロソフト株式会社

業務執行役員 ナショナルテクノロジーオフィサー

〒108-0075 東京都港区港南 2-16-3 品川グランドセントラルタワー

扱うことによる多言語対応の仕組みは当時意欲的な試みであった。

RichEdit コントロール、コードページと文字列をペアとして扱うライブラリ、コンポーネントの開発はほぼ私一人が担当し、しかも私が空いた時間で作業することを強いられた。提案した本人が責任を持って実現することになったのである。一方でこの頃は常に新しいことに取り組むことができ、自身が実装したものを世界中のユーザーに使用いただけることに大きな手応えを感じていた。

容易に想像いただける通り RichEdit コントロールの開発は Rich Text Format に従って正しくディスプレイ、プリンタへの描画がユーザーから見た主要な実装となる。GDI (Graphics Device Interface)、ドライバ (ディスプレイ、プリンタ) がまだまだ安定していない時期において非常に困難な作業であったことを今でも鮮明に記憶している。RichEdit コントロールの開発、更新にあたってはさまざまなことを学んだ。

400K バイト問題

RichEdit コントロールの最初のバージョンは Windows 95 に搭載された。文字コードと Windows の関係において Windows 95 は最も重要なバージョンと言える。1993 年に入ってから Windows 95 における UCS サポートが大きな議論となっていた。今日では想像し難いが UCS に対応することで 400K バイト以上メモリを圧迫することも

問題であった (当時のパソコンの平均的なメモリサイズは 2~4M バイト)。

議論は取束せず 1993 年夏にビル・ゲイツの判断を仰ぐ形で決着を図った。この時の彼の判断は Windows 95 で UCS に対応するというもの。この判断はその後の Windows における多言語対応において非常に大きなマイルストーンとなったことは間違いない。

同じ時期、日本語対応において大きな判断をする。それは「マイクロソフト 標準キャラクタセット」の策定である。この頃までに広く普及していたシフト JIS には、各メーカーが独自に拡張した私用バージョンが複数存在し、これが文字情報交換における相互運用を阻害していた。例えば、あるメーカーの環境で作成した文書ファイルを他のメーカーで読み込むと異なる文字になってしまう問題が生じていたのである。

この相互運用問題の解決を図るために、マイクロソフトは最初で最後の独自文字コード規格である「マイクロソフト 標準キャラクタセット」を策定し Windows 3.1 日本語版に実装した。この時、マイクロソフトはこのキャラクタセットの仕様を凍結し、一切の改変、拡張を行わないことを宣言している。マイクロソフト独自のシフト JIS などと揶揄されることも多いが、元は日本の電機メーカーが JIS X 0208 の集合を連携することなく個々に拡張し、相互運用を損なったことに起因している。このような学びがあるにも関わらず、

未だに外字問題がなくなるということが残念でならない。「マイクロソフト 標準キャラクタセット」については『Unicode IVS/IVD 入門』(日経 BP) で解説している。興味のある方は参照ください。

補助漢字を UCS で拡張

マイクロソフトは Windows 98 を発売するまでは日本語対応をシフト JIS 中心に行ってきた。しかし JIS X 0212、および JIS X 0221 の日本語レパートリーの記号をサポートする必要性から、これまでシフト JIS を軸とした日本語対応から UCS へと方針を変えた。

同時にユーザーインターフェースで使用するフォントの変更 (MS UI ゴシック)、そしてパソコン創世記からの遺産である半角カタカナをユーザーインターフェースから一掃した。

当時ディスプレイの解像度が低く、この決断は容易ではなかった。なぜなら半角カタカナは限られた画面に多くの文字を押し込む手段の一つであったからだ。当然デザイン、バランスの面から半角カタカナはさまざまな制約を受けており、これらの文字を使用することは避けたいものだ。前回に田原氏 (凸版印刷) も本コラムで書かれていたように、使う側が文字、フォントへの意識を持つことは重要と思う。

JIS2004 と Windows Vista

最近ようやく「JIS2004 問題」という表現を耳にすることがなくなってきた。マイクロソフトは可能

な限り最新の標準規格に自社製品、
自社サービスを準拠させることを
ポリシーとしている。

2007年に発表したWindows
Vistaはこのポリシーに沿って国
語審議会の答申に基づく例示字体
の変更と追加漢字への変更に対応
した。追加漢字はBMPへの割り
当てがなかったため、サロゲート・
ペア（UTF-16符号化方式におい
て32ビット）を使用している。

実は2000年冬の米国カリフォル
ニア州で開催された会議におい
てBMP（基本多言語面）領域の
節約といった意見に対して日本側
からの強い反論もなく、303字が
BMP以外の領域の符号位置を割
り当てられてしまった。この時の

決定はその後のソフトウェア開発
のコストを大きく押し上げたこと
は間違いない。

書き手と読み手が同じ文字を見て いる世界の実現にむけて

JIS2004への対応、「マイクロソ
フト標準キャラクタセット」の
根底にあるのは文字情報の相互運
用性の向上に他ならない。確実な
文字情報の交換は、情報システム
において基本となる重要な要素だ。
未だに外字、異体字など中心とし
た文字情報の交換に課題を抱えて
いる先進国は日本だけだろう。

実際のところJIS90、JIS2004
など異なる環境間における文字情
報の交換では、必ずしも書き手と

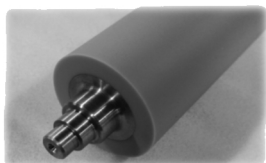
読み手が同じ字形を見ている保証
はない。書き手が字形を意識して
いたとしても、読み手が同じ字形
を見ているとは限らないのが現状
である。

文字情報交換におけるこのよう
な種々の問題解決を目指して設立
されたのが文字情報技術促進協議
会である。

すでに述べたように「マイクロ
ソフト標準キャラクタセット」
を策定する原因となった各メーカ
ーの独自拡張についても依然とし
て外字問題として大きな課題とな
っている。同協議会ではこれらの
課題解決にも積極的に取り組んで
いる。興味のある方はぜひ参加く
ださい。（つづく）

テクノロールは世界へ向けた製品開発を続け、 お客様へ最高品質の製品をお届け致します！

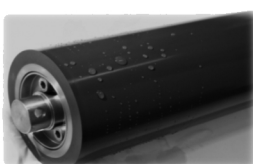
UV印刷用特殊樹脂ロール
トラストシリーズ
TRUST ZETA（油性兼用）
世界12か国のライセンス
最新の高感度UV印刷に対応
最もポピュラーにして高品質
なUV印刷を実現



ドイツ・ウエストランド
油性印刷用ゴムロール
ヴェロシリーズ
優れたインキ転移性
インキマイレージ削減
給水ロールはNON-IPA可



版面ヒッキー除去ロール
ヒッキースーパーW
版面ヒッキー除去
ゴーストの軽減
メンテナンスフリー
各種オフセット印刷に対応



新製品！給水ロール
ヴェロダンプ-LT
次世代オフセット給水
ロールの新定番
非常に安定した水上り
性能が長期間ご使用頂
けます

「印刷志の会」毎度ご好評頂いております



テクノロール株式会社



WESTLAND

本社・工場
関東支店・埼玉工場

〒594-1144 大阪府和泉市テクノステージ3-4-5 Tel 0725-53-3933(代) Fax 0725-53-3922
〒335-0032 埼玉県戸田市美女木東2-3-6 Tel 048-449-7333(代) Fax 048-449-7444
URL : <http://www.technoroll.co.jp> E-mail : info@technoroll.co.jp