

---

---

# システム全般に関わる文字コードの取扱い

～外字はどう取り扱う？～

(株)九州地区農協オンラインセンター

---

## 執筆者 Profile



池田 和正

2006年 (株)九州地区農協オンラインセンター入社  
業務部 県機能班

### 論文要旨

システムを扱う上で「文字コード」は必ず意識する必要がある。全国センター（以下、JASTEM という）より還元されるデータを当社システムで処理し JA 及び顧客に還元するまでに「文字コード」を取り扱う上で、どのような考慮が必要か。単一のシステムの中では意識されることのない「外字」についても複数のシステムを取り扱う電算センターならではの苦悩やその対策がある。

その苦悩や対策について電算業務受託・システム構築・日々のシステムの運用などを通じて経験してきたことと、その対策についてシステムだけの問題でなく、顧客への影響まで含めた形で調査を実施し、どのように対応してきたかを論じている。

プロジェクトであまり意識されることのない「文字コード」について、SE としてしっかりと学び、幅広い観点で影響調査を行ってきたことだけでなく、今後発生するであろう、同様の問題についてどのように「対応」すべきかを述べている。

## 論文目次

<b>1 . はじめに</b> .....	《 3 》
1 . 1  当社の概要	
1 . 2  当社のデータの扱い	
1 . 2 . 1  各システムでの文字コード	
1 . 2 . 2  データの流れ	
1 . 2 . 3  双方向でのコード変換	
<b>2 . 各システムでの外字の取扱い</b> .....	《 6 》
2 . 1  ホストでの外字の取扱い	
2 . 2  サーバでの外字の取扱い	
<b>3 . 外字の取扱いの問題と対策</b> .....	《 8 》
3 . 1  外字はアプリケーションにも依存する	
3 . 2  外字でないのに外字?と思われる	
3 . 3  ダイレクトメール作成時の注意点	
3 . 4  その他文字化けについて	
3 . 4 . 1  プログラムの日本語対応の甘さによるもの	
3 . 4 . 2  改行を表す制御コードの違いによるもの	
<b>4 . 今後発生するであろう文字コードでの課題</b> .....	《 12 》
4 . 1  ホスト更改時のシステム統合	
4 . 2  サロゲートペア	
<b>5 . おわりに</b> .....	《 13 》

## 図表一覧

<b>図1</b> データの流れ .....	《 4 》
<b>図2</b> コード変換の流れ .....	《 5 》
<b>図3</b> JEF漢字コードのやりとり .....	《 7 》
<b>図4</b> ダイレクトメール作成の流れ .....	《 10 》
<b>表1</b> 文字コード一覧 .....	《 3 》
<b>表2</b> コード変換アプリケーション .....	《 5 》
<b>表3</b> JASTEM外字の取扱い .....	《 6 》
<b>表4</b> 韓国語フォント削除前後 .....	《 8 》
<b>表5</b> 外字登録無しの場合 .....	《 8 》
<b>表6</b> 2バイト目が0x5Cの文字 .....	《 11 》
<b>表7</b> OSごとの改行コード .....	《 11 》
<b>表8</b> 丸オンホストコード変換結果 .....	《 12 》

## 1. はじめに

### 1.1 当社の概要

九州地区の7県 JA、信連の信用事業に関するシステム会社として、昭和 52 年 10 月 1 日に設立された。主に以下の事業を行っている。

- (1) 九州 7 県 JA・信連の信用事業に関する情報処理システムの開発、保守、運用
- (2) ネットワークの設計・管理・運用・監視
- (3) 計算事務の受託
- (4) 情報提供サービス業務の受託
- (5) ソフトウェアの開発・販売
- (6) 施設の賃貸並びに受託管理

近年、IT 事業の拡大及び電算業務受託などで様々な機種を扱うことが増えている。

### 1.2 当社のデータの扱い

#### 1.2.1 各システムでの文字コード

当社ではホスト並びにサーバを運用管理しており、様々な文字コードでデータをやり取りしている。基となるのは JASTEM から還元されるデータ（以下、JASTEM 還元データという。）である。JASTEM 還元データ及び当社で管理する機器ごとの文字コードを表 1 に示す。

システム	1 バイト文字	2 バイト文字
JASTEM 還元データ	EBCDIC	JIS X0208-1983
ホスト(GS21 400/200)	EBCDIC	JEF 漢字コード
Windows サーバ	ASCII	Shift_JIS
Solaris サーバ	日本語 EUC	日本語 EUC
OA パソコン	ASCII	Shift_JIS

表 1 文字コード一覧

また、JASTEM 還元データは、JIS X0208-1983 コードで未使用な領域(外字領域)に対応し、1,880 文字分のコード領域を使用している（以下、JASTEM 外字という）。

## 1.2.2 データの流れ

データの流れを図1に示す。

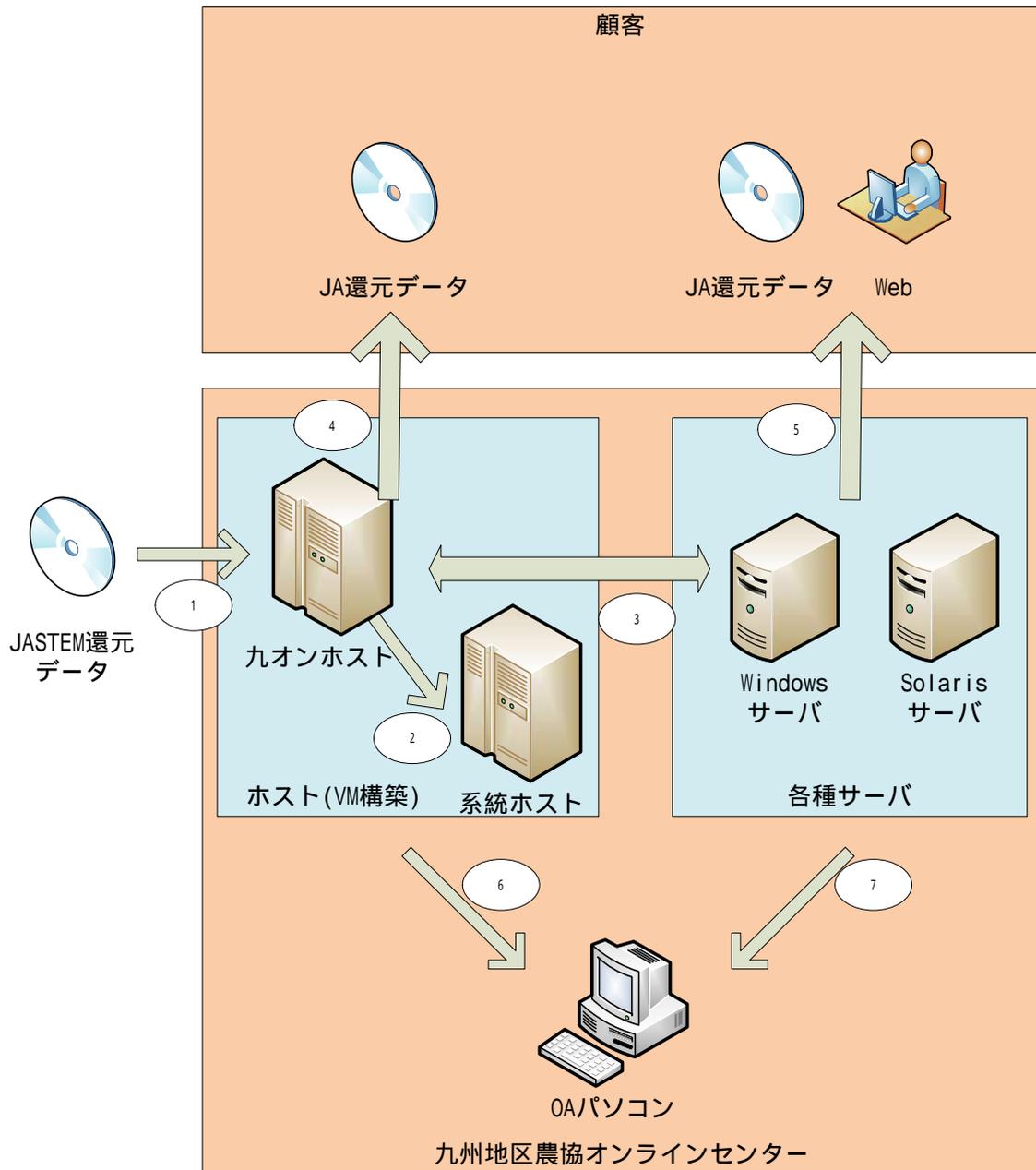


図1 データの流れ

- (1) JASTEMからの還元データをホストへ展開
- (2) 九オンホストよりシステムホストへデータを展開
- (3) 九オンホストより各サーバへデータを展開。各サーバよりホストへデータ取り込み
- (4) JAへの還元データを各ホストで作成
- (5) JAへの還元データをサーバで作成・WEBサーバにて、JA補助端末に展開
- (6) ホストデータをOAパソコンにて取り込み
- (7) サーバデータをOAパソコンにて取り込み

### 1.2.3 双方向でのコード変換

各システム間での文字コード変換を図2に示す。

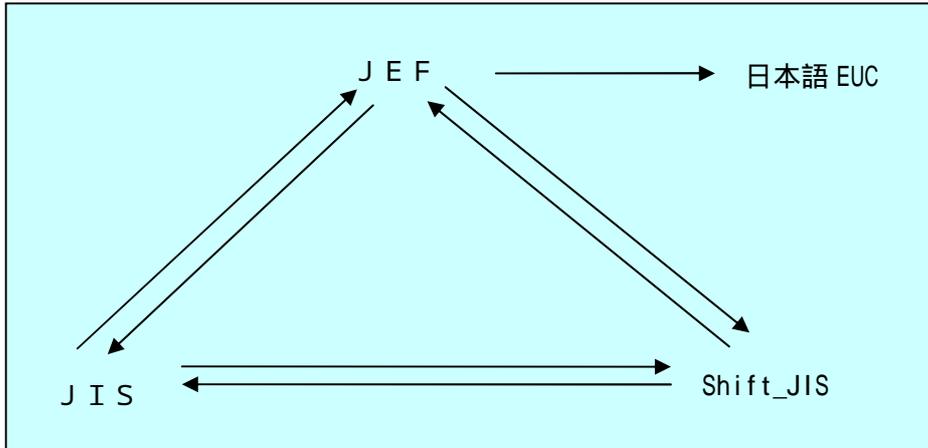


図 2 コード変換の流れ

日本語 EUC より JEF 漢字コードへの変換以外は、それぞれの文字コードにて双方向のコード変換が必要であり、JASTEM 外字を取り扱う上で必ず変換テーブルを持つ必要がある。文字コード変換については、以下のアプリケーションを表2に示す。

	JIS	JEF 漢字コード	Shift_JIS	日本語 EUC
JIS		ベンダーアプリ	ユーザアプリ	
JEF 漢字コード	ベンダーアプリ		ベンダーアプリ	ベンダーアプリ
Shift_JIS		ベンダーアプリ		
日本語 EUC				

表 2 コード変換アプリケーション

## 2. 各システムでの外字の取扱い

### 2.1 ホストでの外字の取扱い

ホストはVM(Virtual Machine)を構築しており、JASTEM 外字の取扱いがそれぞれのシステムで違う。元々の経緯としては系統組合のサーバ及びホストを当社に移行すると共に運用・開発・保守を受託する(以下、電算業務受託という)際に、当社の外字の取扱いを統合出来なかったことに始まる。(以下、VM 構築後の系統組合のホストを系統ホスト、当社のホストを九オンホストという。)

電算業務受託移行作業にて、系統組合のホストで JASTEM 外字の取扱いを調査したところ JEF 漢字コードに変換する際に、類似した拡張漢字に置き換えていることや、ホストに外字登録がされていないことも判明した。

受託を機に当社同様に外字登録を行い、JEF 漢字コードへのコード変換も外字を割り当てることで対応する予定であった。ただし、システムだけの問題でなく、コード変換後のデータは JA などへ毎月還元しているデータであり、影響を調査する必要がある。受託前と受託後での JASTEM 外字に対応する JEF 漢字コードを表3に示す。

JASTEM外字 (JISコード)	受託前システム (JEF漢字コード)	受託後システム (JEF漢字コード)
7F32 黍	A1B2 —	8BB2 黍
7F33 澗	56CE 澗	8BB3 澗

表 3 JASTEM 外字の取扱い(上段：文字コード、下段：文字)

上記のように、受託前システムでは JIS コード「7F32」を「A1B2」に変換しており、「7F33」については、拡張漢字の「56CE」に変換している。これを受託後システムでは、JASTEM 外字を取り扱うように「7F32」を「8BB2」、「7F33」を「8BB3」に変換するように JASTEM 外字を取り扱う方向で提案した。

この際の、受託後システムの課題点は以下のとおりである。

- ・ 文字コードが変わるが問題ないか？
- ・ JAなどで JEF 漢字コードから Shift\_JIS へコード変換を行っている場合の影響は？
- ・ 還元先で JASTEM 外字以外の外字登録を行っている可能性はないか？

結果、JASTEM 外字以外の外字登録を行っている還元先があることや、文字コードが変わることへの影響調査にあまりに時間がかかるため、今回の提案は却下された。

このため、JASTEM 外字(JIS)が同一であったとしても、九オンホストと系統ホストではコード変換後の JEF コードは文字コードが異なる。データのやりとりをする際は、必ず、九オンホスト JEF 漢字コードから系統ホストの JEF 漢字コードに変換する処理を組み込む必要がある。図3にその処理の流れを示す。

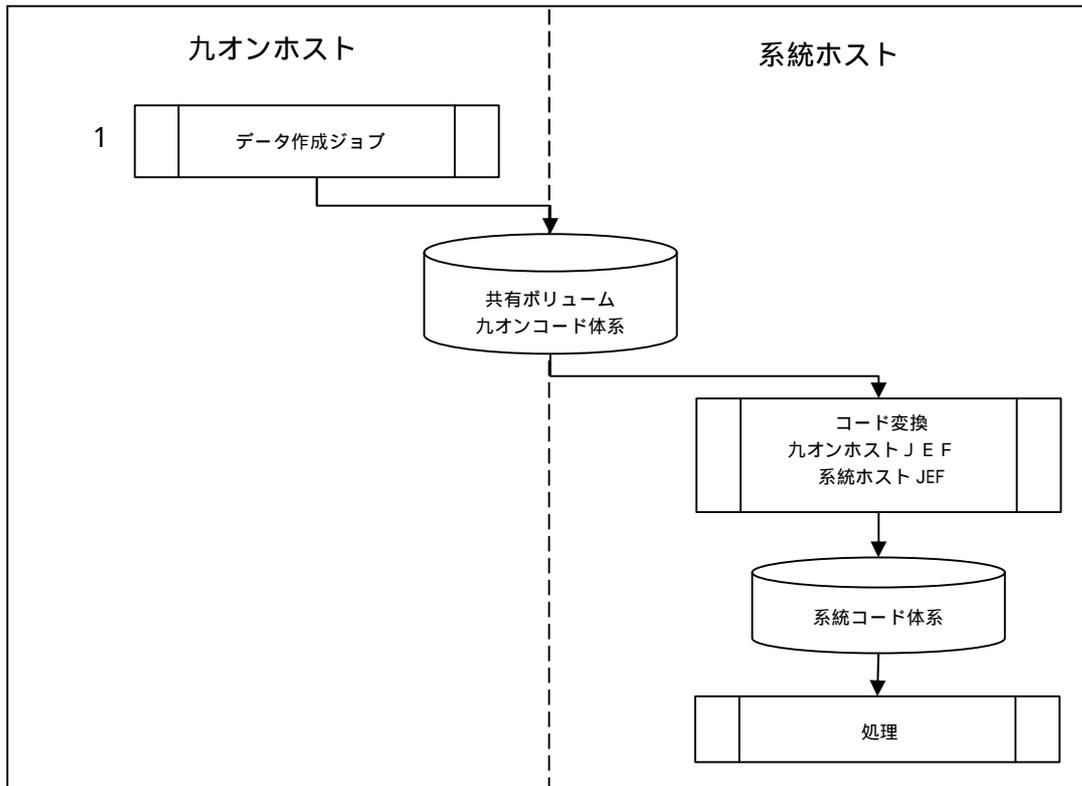


図 3 JEF 漢字コードのやりとり

また、Shift\_JIS についても文字コードが異なるため、同様の考慮が必要となる。

## 2.2 サーバでの外字の取扱い

ホストからサーバへ受け渡すデータについては、通常、ホスト側でコード変換する。ただし、系統ホスト側では元々サーバ側で JEF 漢字コードから Shift\_JIS へのコード変換を行っていた。サーバ側でのコード変換もサーバごとに異なるアプリケーションを利用していた。それぞれのアプリケーションごとに、以下の調査を行い、アプリケーションの統一を図った。

- ・ 変換テーブルは利用しているか？
- ・ 外字は正しく変換されているのか？
- ・ 外字が変換できなかった場合は、どういった文字に置き換えているか？

以上のような点に着目し、受託後もそのデータを利用する顧客及び担当者へ迷惑がかからないように、JEF 漢字コードの全文字コードを Shift\_JIS へ変換する作業を各サーバのアプリケーション毎に行い、問題がないことを確認した。

また、サーバごとに何故、コード変換のアプリケーションが違うのかといった疑問もあったが、システム稼働時の担当者もおらずこの件については、アプリケーション統一に問題がないことを確認し作業を進めた。最終的にコード変換のアプリケーションも各サーバで同一のものを利用し管理しやすくなった。

### 3. 外字の取扱いの問題と対策

#### 3.1 外字はアプリケーションにも依存する

外字は、機種依存文字と呼ばれているが、機種だけでなくアプリケーションにも依存する。以下に、その事例を示す。

- (1) 韓国語フォント (NewGulim) をインストールした Excel 2002、Access 2002、または PowerPoint 2002 で外字を使用できない
- (2) Windows Vista 上の 2007 Office プログラムまたは Office 2003 プログラムで外字 (EUDC) が中国語フォント (MingLiU\_HKSCS) の文字に置き換えられる

どちらも、ユーザー定義の外字ファイルよりも優先されるフォントがあるということである。これは、あくまで仕様に基づく動作らしい。この対応としては、該当のフォントを削除することで対応できる。(1)の韓国語フォント (NewGulim) を削除する前後での Excel 2002 での文字の表示を表 4 に示す。

JASTEM外字 (JISコード)	NewGulimフォント 削除前	NewGulimフォント 削除後
8221	ㄱ	ㄱ
8222	ㄴ	征
8223	ㄷ	ㄷ

表 4 韓国語フォント削除前後

上記のように一見して NewGulim フォント削除前の状態は文字化けが発生しているように思えてしまう。NewGulim フォント削除後の表示をみると文字化けではなく、あくまでフォントの優先順位の問題だけであったことが分かる。

マイクロソフトサポートオンラインには、「Microsoft Office XP アプリケーション (Excel 2002、Access 2002、または PowerPoint 2002) で登録した外字を表示しようとすると、他のハングル文字に置き換えられることがあります。」<sup>1)</sup>と記述されている。ここで、着目したのが、「登録した外字を表示しようとすると」という文言である。では、JASTEM 外字が登録されていない OA パソコンであればどうなるのであろう。その、結果を表 5 に示す。

JASTEM外字 (JISコード)	外字登録無し
8221	・
8222	・
8223	・

表 5 外字登録無しの場合

表 4 ではハングル文字が表示されていたが、表 5 では外字が登録されていないため、それすら表示されないことが分かる。

### 3.2 外字でないのに外字？と思われる

ホスト及びサーバでコード変換したデータを OA パソコンへファイル転送してみると、文字化けしたかのような文字をみることがある。原因は様々であるが、先日、面白い事例があった。社員の OA パソコンで顧客名称を見ていたところ名称の 1 文字が「・（中点）」のデータが存在した。この際に、以下のような原因からこの現象が起きたのではと調査を行った。

- ・ OA パソコンに外字が登録されていないのでは？
- ・ JASTEM 外字のコード変換がうまくいかなかったのでは？
- ・ OA パソコンに転送する際に、何かしら問題があったのでは？

実際にコード変換前のデータを調査したところ元々のデータが「・（中点）」であった。JASTEM システムから還元されるデータは、殆どが JA 窓口からオペレーションされた情報から作成されている。窓口端末では JASTEM 外字は登録されており、この文字についても、窓口で「・（中点）」として登録されたものであった。

こういった事例は稀にあり、オペレーションする JA 職員が該当する文字が存在しないことや、JASTEM 外字から該当の文字を探し出せないということで故意に入力している。当社ではそのデータは正しいものとして、処理することしか出来ないのが現状である。

### 3.3 ダイレクトメール作成時の注意点

「2.1 ホストでの外字の取扱い」で述べたように、システムホストでは JEF 漢字コードの外字領域を使用していない。このため、システムホストで作成した帳票及びデータについては、外字が拡張漢字または、「\_（アンダーバー）」に置き換えられている。

ここで注意が必要なのが、顧客に直接送付するダイレクトメールである。当社では、ダイレクトメールを作成する設備が整っているため、電算業務受託を機に、システムホストでのダイレクトメール作成依頼が度々発生している。

ただし、顧客情報データには、JASTEM 外字がかなりの確率で含まれておりシステムホストのデータを利用すると、ダイレクトメールに印字する「住所」及び「顧客氏名」などが「\_」に置き換えられて発送されてしまう。

ここで考えたのが、システムホストは移行作業時に JASTEM 外字の登録はしているという点である。JIS コードから JEF 漢字コードへのコード変換は JASTEM 外字を利用していないが、外字は扱えるということである。

九オンホストで JIS コードから JEF 漢字コードへ変換されたデータを顧客情報としてシステムホストへコピーし利用すれば JASTEM 外字も印字することに問題がない。図 4 にダイレクトメール作成の流れを示す。

九オンホスト

系統ホスト

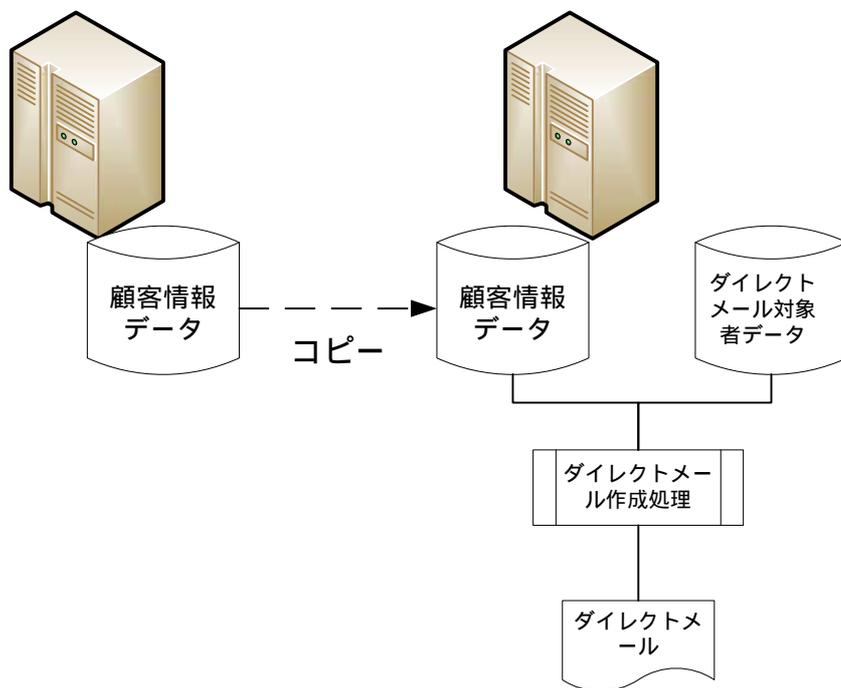


図 4 ダイレクトメール作成の流れ

また、ダイレクトメール作成の依頼者からは、ダイレクトメールを送った顧客の一覧が欲しいと言われることがある。この際も、単に Excel ファイルにおとすだけでは、閲覧する OA パソコンに JASTEM 外字が登録されていないと表示されないため、PDF 化して提出するようにしている。

### 3.4 その他文字化けについて

その他、文字化けについては以下のような原因から発生する事が考えられる。

#### 3.4.1 プログラムの日本語対応の甘さによるもの

Shift\_JIS を内部コードで利用するアプリケーションでは、エスケープシーケンスの取得の仕方に工夫が必要である。ところがそれがなされていないため問題となる場合がある。Shift\_JIS において、2 バイト目が 0x5C (日本の円記号、米国ではバックスラッシュ) となる文字 (俗に「ダメ文字」と言われる) の場合、2 バイト目の 0x5c がエスケープを意味する制御文字として動作することがあり、正しく表示できなくなる場合がある。ダメ文字の一覧を表 6 に示す。

文字	文字コード	文字	文字コード	文字	文字コード	文字	文字コード
	815C	申	905C	媾	9B5C	觸	E65C
ソ	835C	曾	915C	彌	9C5C	體	E75C
	845C	箆	925C	拿	9D5C	鐔	E85C
	875C	貼	935C	朽	9E5C	饅	E95C
噂	895C	能	945C	畝	9F5C	鷓	EA5C
湮	8A5C	表	955C	濬	E05C	僣	ED5C
欺	8B5C	暴	965C	畚	E15C	砒	EE5C
圭	8C5C	予	975C	秉	E25C	纒	FA5C
構	8D5C	禄	985C	綵	E35C	状	FB5C
蚕	8E5C	兔	995C	臀	E45C		
十	8F5C	嗜	9A5C	藹	E55C		

表 6 2 バイト目が 0x5C の文字

多くのプログラミング言語では、この 0x5C をエスケープ文字としている。この問題を回避する伝統的な方法として、ソースコード全体を EUC コードや UTF-8 などに変換してからコンパイルしたり実行したりする方法がある。

#### 3.4.2 改行を表す制御コードの違いによるもの

OS ごとに改行を表す制御コードの指定が異なるため、変換に失敗するとその部分が化けることもある。表 7 にそれを示す。

OS	改行コード
Windows	CRLF(0x0D 0x0A)
Mac OS	CR(0x0D)
Mac OS X	LF(0x0A)
UNIX	LF(0x0A)

表 7 OS ごとの改行コード

上記のように、異なる OS を使用しているコンピュータ間でファイルを転送させた場合、改行コードの違いから、期待しているテキストが表示されない文字化けが起きてしまう。これらの改行コード、文字コードの違いに対する問題は、変換ソフトや複数の改行コードや文字コードに対応したテキストエディタなどの利用で補える。

## 4. 今後発生するであろう文字コードでの課題

### 4.1 ホスト更改時のシステム統合

次期のホスト更改が数年後に控えている。ここでの課題としては、資源を有効活用するためにも VM を構築せずに、九オンホストと系統ホストを統合出来ないかということである。ここでネックとなるのが九オンホストと系統ホストでの JASTEM 外字の取扱いがまた課題となるのが明らかである。

電算業務受託時に、外字の取扱いを統合出来なかった経緯はあるが、その際に影響があると答えた相手先には、今後、システム統合の際には、外字の取扱いを統合する旨は伝えられている。

ホストで外字を扱うように統合出来たとしても、サーバ側での外字対応も必要でありその点についても、調査・分析が必要となる。現在できる事として九オンホストと系統ホストを利用してのコード変換の結果についてそれぞれ纏めている。

九オンホストにて JASTEM 外字(JIS)を JEF 漢字コードと Shift\_JIS に変換した結果を表 8 に示す。

JIS(16進)	JEF(九オン)	SJIS(九オン)	文字
7F21	8BA1	F040	●
7F22	8BA2	F041	●
7F23	8BA3	F042	●
7F24	8BA4	F043	●
7F25	8BA5	F044	●
7F26	8BA6	F045	●

表 8 九オンホストコード変換結果

上記のように JASTEM 外字がどのようにコード変換されるかを資料として残すことで、システム統合の際に九オンホストと系統ホストでのコード変換結果にどの程度、影響があるかを再度、見直す事ができる。今後は、系統ホストでのコード変換結果を纏めていくとともに、九オンホストと系統ホストの対比表も作成していく必要がある。

## 4.2 サロゲートペア

サロゲートペアとは、Unicode の一文字を 2 バイトという基本は維持しつつ、一部の文字を 4 バイトで表現したものである。これは、Unicode に組み込みたいという文字の要望が増えたことから導入されたものである。Windows Vista の JIS2004 から対応されている。JIS2004 とは、現在主流の JIS90 に第三・第四水準漢字と非漢字を追加したものである。文字数は JIS90 に対して、約 4,300 文字増加している。また、JIS90 範囲内の 168 文字の字体が変更されている。今後、文字セット変更が必須となってくるが、それに伴う問題点を以下に示す。

- ・ 追加された文字は、JIS90 対応のクライアント・サーバ・プリンタでは扱えない
- ・ 字体変更された文字は、JIS90 対応のクライアント・サーバ・プリンタで表示が異なる
- ・ Unicode 非対応のソフトウェアは JIS2004 を扱えない

今後、OA パソコンの更新に伴い、現在利用している WindowsServer2003、WindowsXP と WindowsVista、Windows7 が混在することが懸念される。WindowsServer2003 と WindowsXP については JIS90 のため、JIS2004 対応パッチを適用しておく必要がある。

また、これが具体的にどうプログラムに影響するかというと、サロゲートペアを使用する文字は 4 バイト / 1 文字であり、従来の 2 バイト / 1 文字と文字が混在した場合に 2 バイトを意識しての処理がすべて使えなくなってしまうことである。Excel 関数や VBA 関数で文字列長を取得する際や、バイト数を取得する処理がある場合は変更が必要となる。

## 5. おわりに

電算業務受託やシステムの運用を通じて「文字コード」については、色々と学ぶ事が出来ている。本稿では、複数の「文字コード」を扱う上での課題と対策を論じてきたが、まだまだ改善の余地はあるのかも知れない。今後も発生する「文字コード」の課題について混乱を防ぐため、早期の対応を心がけていきたい。最後になるが、同様の課題をもつ SE の方々に本稿が僅かでも参考にして頂ければ幸いである。

## 参考文献

なし

---

\*1 <http://support.microsoft.com/kb/417636/ja>