テキストデータの作り方について①

一口にテキストと言っても、プレーンテキスト、リッチテキスト、HTML、XMLなど保存形式は様々です。DTPの原稿として使用する場合、どのようなテキストデータが望ましいのでしょうか。

■文字コードについて

テキストデータをやりとりする際、文字コードについて意識することはあまり多くないと思いますが、文字コードに絡んだトラブルは決して少なくありません。そもそも文字コードとはどんな用途を持ったもので、どのような種類があるのでしょうか。

コンピュータ上で使われる文字には、主に1バイト（半角文字）、2バイト（漢字など）の2種類があり、それぞれ文字コードと呼ばれる文字の番号によって管理されています。

一般に広く使われているのはシフトJISで、全員がこれを使っていれば特にトラブルは起きないのですが、シフトJISは外国語（ハングルや簡体字）が扱えず、漢字の種類にも限りがあるので、表示出来ない文字が多くあります。WindowsとMacintoshは通常、シフトJISを使っていますが、各々が文字コードを拡張し、独自の文字を追加したため、他機種間では、文字化けが起きるようになってしまいました。これを機種依存文字と言います。

機種依存文字の代表的な文字化けには次のようなものがあります。

機種依存文字の例

Windows	Macintosh
① ② ③	㈰ ㈪ ㈫
Ⅰ Ⅱ Ⅲ	㈱ ㈱ ㈱

文字の多様さや互換性といったこれらの問題を解決するために生まれたのがUnicodeです。外国語と日本語を統一して表現でき、特殊な漢字も扱うことができます。

しかし、Unicodeに対応していない環境・アプリでは文字化けしてしまう可能性があるなど、問題もあります。WindowsではXP以降、MacはOS XからUnicodeに対応していますが、テスト入稿の際はそれがUnicodeかどうかを明示しておくことで文字化けや文字飛びを少しでも避けることができます。

Mac OS X（Unicodeテキスト）

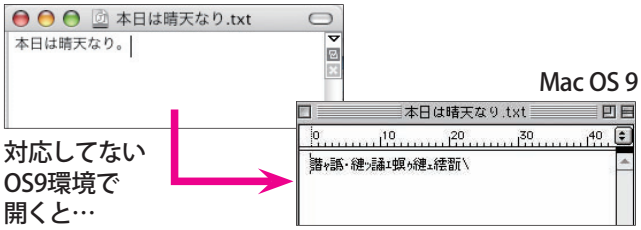

表1　一般的な文字コードの種類

文字コード名	規格団体	種類	特徴	用途
JISコード	JIS	英数字：1バイト 漢字（X0208）：2バイト 切り替え符号：3バイト	ISOコードを元に、カタカナを追加	電子メール
シフトJISコード	マイクロソフト	英数字、半角カナ：1バイト 漢字（X0208）：2バイト	MSが作成した文字コード。現在ではパソコンで使われる漢字コードとして最も普及している	Windows、Mac OS
EUC	AT&T	英数字：1バイト 半角カナ：2バイト 漢字（X0208）：2バイト 漢字（X0212）：3バイト	主にUNIXで使用されている文字コード	UNIX
Unicode	ISO	基本的にすべて：2バイト	世界の主な言語を１つの文字コードで表す	統一文字コード

■Excelでの入稿

表組みであれば、Excelがベストです。簡単なものであれば、Wordの表組みでも問題ありませんが、複雑な表になってくるとやはりExcelでの入稿がよいでしょう。Excelはデータの加工がしやすいため、作業の効率化を考える上で大きな可能性を秘めています。

例えば、ウの項目(D列)をオの項目(F列)右側へ移動したい、オの項目は太字にしたいのでタグを入れるための記号を入れたい、といった加工がとても簡単に行えます。

ただし、表内の文字にルビをふる必要がある場合は、注意が必要です。Excelでもルビ（Excelでは「ふりがな」機能）をふることはできますが、セル単位で「ひらがな」「全角カタカナ」「半角カタカナ」のいずれかの設定しかできないため、同一のセル内にひらがなとカタカナのルビを混在させることができません。表組み内でルビを使う場合はWordの方がよいでしょう。

例　「ひらがな」　Excelの使い方（えくせる つか かた）
　　「全角カタカナ」　Excelの使い方（エクセル ツカ カタ）
　　「半角カタカナ」　Excelの使い方（ｴｸｾﾙ ﾂｶ ｶﾀ）

項目の移動

	A	B	C	D	E	F
1	◆A					
2	名前	ア	イ	ウ	エ	オ
3	aaa	2	24	4	18	24
4	bbb	1	666	3	18	20
5	ccc	1	21	2	14	22
6	ddd	4	78	14	40	40
7	eee	3	57	8	26	33
8	fff	4	1180	14	891	35
9	ggg	5	88	17	40	43
10	hhh	8	112	38	78	56
11	iii	7	96	31	58	51
12	jjj	39	783	777	5623	395
13	kkk	39	774	664	510	878
14						
15	◆B					
16	名前	ア	イ	ウ	エ	オ
17	aaa	10	144	67	130	83
18	bbb	11	173	84	160	98
19	ccc	11	180	72	142	81
20	ddd	12	192	81	130	95
21	eee	12	228	96	150	121
22	fff	14	268	160	145	129
23	ggg	15	350	198	190	144
24	hhh	14	304	173	166	123
25	iii	53	1220	1130	755	700
26	jjj	53	1150	1030	680	675
27	kkk	53	1190	1054	692	662
28						

	A	B	C	D	E	F
1	◆A					
2	名前	ア	イ	エ	オ	ウ
3	aaa	2	24	18	24	4
4	bbb	1	666	18	20	3
5	ccc	1	21	14	22	2
6	ddd	4	78	40	40	14
7	eee	3	57	26	33	8
8	fff	4	1180	891	35	14
9	ggg	5	88	40	43	17
10	hhh	8	112	78	56	38
11	iii	7	96	58	51	31
12	jjj	39	783	5623	395	777
13	kkk	39	774	510	878	664
14						
15	◆B					
16	名前	ア	イ	エ	オ	ウ
17	aaa	10	144	130	83	67
18	bbb	11	173	160	98	84
19	ccc	11	180	142	81	72
20	ddd	12	192	130	95	81
21	eee	12	228	150	121	96
22	fff	14	268	145	129	160
23	ggg	15	350	190	144	198
24	hhh	14	304	166	123	173
25	iii	53	1220	755	700	1130
26	jjj	53	1150	680	675	1030
27	kkk	53	1190	692	662	1054
28						

項目の追加

	A	B	C	D	E	F	G	H
1	◆A							
2	名前	ア	イ	エ		オ		ウ
3	aaa	2	24	18	☆	24	★	4
4	bbb	1	666	18	☆	20	★	3
5	ccc	1	21	14	☆	22	★	2
6	ddd	4	78	40	☆	40	★	14
7	eee	3	57	26	☆	33	★	8
8	fff	4	1180	891	☆	35	★	14
9	ggg	5	88	40	☆	43	★	17
10	hhh	8	112	78	☆	56	★	38
11	iii	7	96	58	☆	51	★	31
12	jjj	39	783	5623	☆	395	★	777
13	kkk	39	774	510	☆	878	★	664
14								
15	◆B							
16	名前	ア	イ	エ		オ		ウ
17	aaa	10	144	130	☆	83	★	67
18	bbb	11	173	160	☆	98	★	84
19	ccc	11	180	142	☆	81	★	72
20	ddd	12	192	130	☆	95	★	81
21	eee	12	228	150	☆	121	★	96
22	fff	14	268	145	☆	129	★	160
23	ggg	15	350	190	☆	144	★	198
24	hhh	14	304	166	☆	123	★	173
25	iii	53	1220	755	☆	700	★	1130
26	jjj	53	1150	680	☆	675	★	1030
27	kkk	53	1190	692	☆	662	★	1054
28								
29	◆C							
30	名前	ア	イ	エ		オ		ウ
31	aaa	18	370	300	☆	180	★	248
32	bbb	18	390	320	☆	187	★	255
33	ccc	17	348	340	☆	166	★	210
34	ddd	17	366	360	☆	172	★	233
35	eee	21	475	380	☆	230	★	270
36	fff	21	460	400	☆	211	★	282
37	ggg	22	527	420	☆	280	★	304
38	hhh	22	530	440	☆	199	★	344
39	iii	25	1675	460	☆	278	★	537
40	jjj	25	2515	480	☆	265	★	472
41	kkk	19	423	500	☆	204	★	266
42								

次回（②）では、Word入稿とプレーンテキスト入稿について説明します。

当社では、財団法人 日本漢字能力検定協会公認の、モトヤ教科書体筆順フォントを所有しています。

筆順フォントとは、小学生向けの漢字ドリルで使用される漢字、仮名の書き方の順番を表現できるように、各文字を複数の部品に分割したデータを含むフォントのことです。

モトヤ教科書体筆順フォントには「一画ずつ組み上がっていく」ヒツジュン版と「一画ずつ独立した」エレメント版（下のイメージ参照）の2種類あります。これらを合成することにより多様な組版表現が可能です。

ヒツジュン版イメージ

エレメント版イメージ

原稿作成のために、必要な漢字のヒツジュン版イメージを全画数分、一覧にしてプリントすることも可能です。

対応文字は、ひらがな・カタカナ96字、常用漢字1,945字の計2,041字となります。

常用漢字の詳細はコチラ　http://ja.wikipedia.org/wiki/常用漢字一覧