

XML を用いた多言語環境ブラウザ[§]

2J-3

原 淳太郎[†]早稲田大学理工学部[‡]

1. 研究の目的

コンピュータの普及は年を追うごとに広まり、コンピュータ・ネットワークが日常的に使用されるようになった。国際的ネットワーク上でのコンピュータ環境では、特定の言語のみで処理できる範囲を超えてしまうため、文字と言語を無制限に混在して使用することができる、国際化・多言語化が要求されている。

今日、各国ごとに国家規格として制定し使用している情報交換用コードは、同言語地域のみで使用できる地域化された環境であり、複数言語を混在することは非常に難しく、計算可能性と情報交換可能性を著しく阻害する。UNICODE などの国際規格などにより、文字コードに対する規格が定められつつあるが、文字コードは言語情報を持っていないため、多言語化システムが必要である。

以上を踏まえ、本研究は、早稲田大学国際化・多言語化環境^{*}の知識を元にし、複数コードを用いることのできる国際化・多言語化環境の開発を目的としている。

2. 地域化・国際化・多言語化

一般に、ある特定の文字集合に合致した言語処理系で記述されたアプリケーションで、その処理系の文字集合で記述し得ない言語を扱う場合、特定の文字集合への拡張を施し、その文字集合で表現可能な特定の 1 言語に合わせて直接修正している。このように、特定の処理を特定の 1 言語で処理することは「地域化」と呼ばれる。これに対して、特定の処理において任意複数の言語を同時に混在して処理することは、「国際化」と呼ばれる。国際化は、全言語に共有される情報のみを処理する。さらに、特定の言語を特定の情報を使用して処理することを「多言語化」と言う。

3. 縦横混在表記

表記方向の混在する文書において、読みやすい表記というものを考える必要がある。ここで、その種類について幾つか述べておく。

● 表記順表記

表記順序 (場合により表記/改行方向) を併記する。どんなものでもどんな順序でも表記できるが、次に読むべき単語を読者が探索する必要が否めない。同じ文字種のもを同一場所に置くことが出来るため、データ格納に向いている。

● ヒエログリフ式表記

ベースとなる表記/改行の方向を定め、それに従わないものは表記方向はそのまま、改行方向はベースの表記方向にする。完全表記よりみだすが、やはり空白が多くなる。また、多重構造において対応できない場合がある。

● 改行方向統一表記

表記・改行の双方を統一できない場合に、どちらかのみを統一する手法で、両方を統一する場合より読みやすい場合も多い。改行方向統一のばあい、進行方向が一定になるため、一般的な文章に向いている。

● 時間順表記

表記順がわかるよう動的に表示する手法。他の表記と併用する。人によって読解速度は異なるため、表示速度を容易に変更出来る必要がある。また、音読システムを併用することで外国語学習システムとしても利用可能。

● 辞書式表記

表記形式を画一化し、一定のルールに沿って読ませる手法。表記無しでも読解できるような辞書などに有効であり、例えば英語の辞書は英語/発音記号/日本語の 3 種の言語を無表記の手法で表記している。

^{*}小原啓義『全世界の文字の混在処理を可能とするマルチリ
ンガル入出力・文書処理・通信機構の研究』1999

[§] Research about a multilingualized environmental browser

[†] Hara Juntaro

[‡] Faculty of Science and Engineering, Waseda University

○ 総合

実際には1つの表記法であれば全てを網羅できるということはないため、状況に応じて使い分けできるようなシステムが必要である。

a b c d e f g A B C D E F G く り ね ち せ い ね ふ り せ い ね	① A B C D E F G ④ a b c d e f g ⑥ ⑤ ③ ② る ち に い を り ほ ろ わ ん へ は
●表記方向統一表記	●表記順表記
a b c d ほ いろ へ と ろ ち に e f g A B C D	B o o k ブック ほん C a t キャット ねこ
●ヒエログリフ式表記	●辞書式表記

表 1: 縦横混在表記方法

4. 文字コードの混在

すべての文字は図形もしくはその集合である。すなわち文字コードを表示するには、コード情報を一度図形へと変換する必要がある。そこで、ある文字コードで文字を図形に変換し表示した後、文字コードを変更する。変更する以前の文字を画像として処理することにより、混在している複数の文字コードを文字化けすることなく表示することが可能になる。

5. 考察

現在、表記方向の混在方法について定まり、これについて実装をしている。また、多言語 XML ブラウザとして使用できるよう、ファイルの取り込みやなどの環境を整える必要がある。前にも述べたとおり、文字の表示部分の改善や、各言語ごとの設定は言語ごとに分け、それぞれのファイルを取り込むことで初期表示図形の決定・回転等の正しい情報を与えることが可能であることも望ましい。

さらに、現在の表示システムだけでなく、文字列の挿入・削除・比較・検索・置換といった、基本テキスト処理ができるような編集システムを考える必

要がある。編集システムを持つことによって、利用者は、複数の XML ファイルをあたかも1つのファイルであるかのように取り扱うことが可能になる。また、同じ文字を利用する異なる言語の表記を区別した検索や置換が行えるようになる。

6. まとめ

本研究では、言語依存処理と言語非依存処理とを区別したことで定義された国際化・多言語化環境の下で、文字図形について考察し、文字図形を合成した1文字を定義した。これにより基本テキスト処理上での1文字の定義がなされた。

また、文字の分析結果から、基本テキスト処理のために UNICODE の拡張が必要であることがわかった。

さらに、表記方法をいくつか考える必要があり、読み手が読みやすい表記法へ、容易に移行できる環境を整える必要がある。

本研究のシステムは、いまだ試験段階にある。しかし、国際的コンピュータ・ネットワークはますますその広がりを見せつつある。本研究により、XML と複数文字コードを用いて多言語環境を表現することが可能なことは示された。今後はシステムの拡張・改善に重点を置き、国際化・多言語化されたコンピュータ・ネットワーク上で簡単に利用できる国際化・多言語化入出力システムを作成する予定である。

参考文献

- [1] 柴田武『世界の言葉小事典』大修館書店 1993
- [2] 小原啓義『全世界の文字の混在処理を可能とするマルチリンガル入出力・文書処理・通信機構の研究』1999
- [3] 上園一知『全世界の音節文字群への対応による完全な一般化出力機構の実現』1994
- [4] 上園一知『1文字の決定による地域化・国際化・他言語化処理システムとモデル』1996
- [5] XML ワーキングドラフト
<http://www.w3.org/TR/1999/WD.xml>
- [6] UNICODE <http://www.unicode.com>
- [7] 世界の文字 <http://www.nacos.com/moji/> 中西出版社