

## Sequential Talks on The Multilingual Text Processing (3):

1 Q-3

### The Basic and Advanced Multilingual Text Widgets

Toshio Oya†, Tomoko Kataoka\*, Kazutomo Uezono†, Hidejiro Daikokuyaf,  
Dawa Yidemucaof, Yutaka Kataoka\* and Hiroyoshi Ohara†

\* Media Network Center, Waseda University † School of Science and Engineering, Waseda University

#### 1. Introduction

従来のテキストの2次元座標上の表示及び編集は、処理単位が1文字かつ1字形であり、文字の表記方向及び改行方向が常に定まっていた。しかし、歴史的文字を含む全世界の文字の同時混在処理を目的とする、Internationalized and Multilingual System (IMS) [Talk 1] では、2次元座標上への表示において横書き・縦書きの混在、改行方向の混在、文字表記方向の混在が生じる。さらに処理単位とグリフとの関係が必ずしも1対1とはならない。そこで昨年度において、テキスト処理が Basic/Advanced Text Manipulation に分類・一般化されたのを受け [1]、言語/コード系/フォントに非依存に2次元座標上での表示編集を行なうための必須関数を明らかにし、言語/コード系/フォントに非依存な Multilingual Text Widget の構成について発表を行なった [2]。

この Multilingual Text Widget は、System 1 [3, Talk 1] の WC を処理単位とし、Basic Text Manipulation 及び 2DDLlib を用いて実現された基本的 Text Widget である。しかし歴史的文字を含む全世界の言語の同時混在したテキストの表示編集にはより高度なテキストの配置機能が必須であり、またディスプレイへの表示以外にプリンタへの出力が必須となる。つまり WYSIWYG 機能をもつ Advanced Multilingual Text Widget が必須となる。本稿では、テキストのフォーマットの基本となる、Basic Text Widget におけるテキストの自動配置について述べ、さらに Multilingual WYSIWYG 実現のための必須情報、機能、処理単位について述べる。

#### 2. テキストの自動配置

Basic Text Widget は、WC を処理単位として動作する。従って Basic Text Widget の扱うテキスト中には WC が保持する表示用情報以外、テキストの配置のための情報は存在しない。このため Basic Text Widget におけるテキストの配置は、簡易フォーマットの利用による自動配置以外、行なうことができない。

従来のテキストの表示は、物理的表記方向が一定で改行方向も一定であり、あたかも scroll のようにテキストが一定方向に伸長していた。縦書き・横書きの混在するテキストは、なんらかの制約がなければテキストの伸長方向が4方向になり、テキストが重複して表示される可能性がある。このようなテキストの Window 上への表示は、カーソルの移動、スクロールバーによる画面の移動、また可読性に重大な影響を与える。自動配置の手法はお

まかに、従来通りの Scroll 方式と、テキストを1つ以上の Page として表現する Page 方式の2通りあると考えられる (図1)。

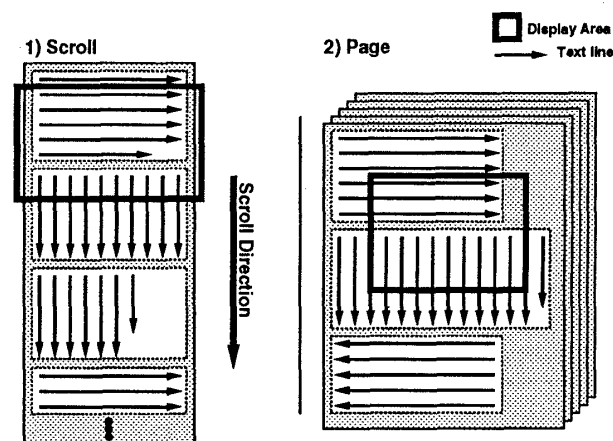


図1. テキストの自動配置

1) Scroll 方式 - 従来のテキスト表示手法を縦書き・横書きのために拡張した手法である。Scroll Direction を定め、また縦書き・横書きの各行の最大長を定めることにより、図1に示すように自動配置を行なう。

2) Page 方式 - ユーザー指定により大きさ定めた長方形の領域を 1 Page と定義し、テキストを1つ以上の Page で表現する。縦書き・横書きの各行の最大長、Page 中のテキストの書き出し位置及びテキスト配置の規則を定めることにより、図1に示すように自動配置を行なう。Page 方式の場合、一度に表示できるテキストは 1 Page 中の表示領域に限定され、スクロールバーにより自由に Page 内の表示位置を変更できる。問題点としてテキスト編集により、編集 Page 以降の全 Page のテキストを再配置する必要が生じる。

上記の自動配置の手法は選択可能でなければならない。さらに自由度の高いシステムの構築のため、ユーザーが用意した自動配置手法をも利用可能とすべきである。つまり、自動配置を実現する簡易フォーマットを複数用意し、これを切替えることにより、ユーザーの希望するテキスト表示を実現する。

#### 3. Multilingual WYSIWYG

Basic Text Widget では、縦書き・横書きの混在を簡易フォーマットを用いることで実現した。しかし IMS の電子図書館、語学研究等で用いるテキストの表示では、より複雑なテキストの配置、ル

ビ振り、文字の装飾などが必須となる。またディスプレイへの表示だけでなく、テキストの編集及びプリンタへの出力機能が必須である。つまり、Multilingual WYSIWYG の実現が Multilingual Text の表示編集には必須である。

ルビ振りなどテキストの配置は、テキストの表示出力から見れば、各グリフの表示座標の特定である。しかし元朝秘史など、実際のテキストでは文字の配置に意味があり、Multilingual WYSIWYG は文字列と文字列との関係をも明確に行なえる能力を持つ必要がある。

#### 4. Processing Unit and Directives for WYSIWYG

従来のテキストフォーマット及び元朝秘史など歴史的文字を含む新聞・書籍などの世界のテキストの調査・分析によると、WYSIWYG の必須操作は、a) 文字に対する操作 b) 文字列に対する操作 c) 行に対する操作 d) 同一物理表記方向のテキストに対する操作 e) 全テキストをページに配置のための操作に分類できる。この分析により a) の操作で要求される情報は、文字単位で操作されるため処理単位毎に保持される。また b) から e) に対する操作はテキストの配置のための操作としてまとめることができ、この操作の要求される情報は、テキスト配置のための指示子 - Directive としてテキスト中に保持される。

WYSIWYG の処理単位は、Basic Text Manipulation を可能とし、処理単位毎にグリフの書体・大きさ・装飾およびハイフネーション処理を含む高度な改行処理などのテキスト処理情報を保持しなくてはならない。この必須情報の一般化によって、WC コードポイントに、WYSIWYG 用の言語/コード系/書体に依存する情報を付加した Text Manipulation Code (TMC) を定義することができる。

また Directive は、WYSIWYG 用 TMC に保持できない、ルビ振りや、縦書き・横書きの混在といった、テキスト配置のための情報をテキスト中に保持する。Multilingual WYSIWYG の Directive は従来のテキストフォーマットの Directive と異なり、その定義が Programable でなければならない。何故ならば、対象となるテキストに用いられるコード系は任意であるため、Directive に用いられるコード系及び Directive を示す文字列が変更可能である方が自由度が高く、また有効である。また、Multilingual WYSIWYG ではテキストの物理表記方向・縦書き・横書きのインタラクティブな変更が可能のため、Directive の定義が Programable であることが必須となる。

#### 5. Implementation Remarks of Advanced Text Widget

上述した WYSIWYG 用 TMC と Programable Directives によって構成される Rich Text を処理する Multilingual フォーマッタと、この Rich Text の編集を行なう Text Widget を用意し、組み合わせることにより、Basic Text Widget より進んだ、Advanced

Multilingual Text Widget を構築できる。

Advanced Text Widget によって作成されたテキストのプリンタ出力は、任意の解像度のドットマトリクスイメージを生成し、任意のページ記述言語/プリンタ制御コードを用いてプリンタに出力させることにより実現する。利用するページ記述言語/プリンタ制御コードの情報は外部に情報記述テーブルを用意し、これをユーザーの指定によって読み込む。この機構を利用することにより、プリンタの機種に依存することなく Multilingual WYSIWYG を実現することができる。

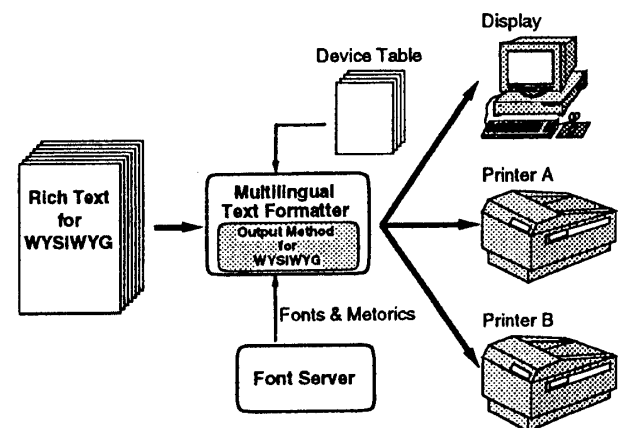


図2. WYSIWYG における出力機構

また Multilingual WYSIWYG に用いられるフォントは、任意の解像度・大きさのグリフが要求されるため、アウトラインフォントが必須である。また横書きだけでなく縦書きのための情報を保持しなくてはならない。既存のアウトラインフォントフォーマットでは、Multilingual WYSIWYG で要求される情報が不足しており、またその文字種が限られているため、現在グリフの表示に必須な情報を含め、アウトラインフォントのフォーマット及び Window System/Application へのフォント情報の供給手法についての研究が行なわれている。

#### 参考文献

- [1] Tomoko, K. et al., The Worldwide Multilingual Computing (4): Essentials for the Multilingual Text Manipulation Processing of the 51th General Meeting of IPSJ, Vol. 3, September 1995, pp 251-252.
- [2] Toshio, O. et al., The Worldwide Multilingual Computing (5): Multilingual Text Manipulation and Text Widget Processing of the 51th General Meeting of IPSJ, Vol. 3, September 1995, pp 253-254.
- [3] Kazutomo, U. et al., The Worldwide Multilingual Computing (3): Functions, Model, Design and Architecture of Multilingual I/O TM/C System Processing of the 51th General Meeting of IPSJ, Vol. 3, September 1995, pp 249-248.