

国際化テキスト入力機構モデルの構築

6Q-9

上園 一知† 片岡 朋子* 片岡 裕‡ 笥 捷彦†

† 早稲田大学理工学部 * 早稲田大学メディアネットワークセンター ‡ 大谷大学文学部

1.はじめに

国際的な通信では、全ての文字を同時に混在使用可能であることが要求される。そこで、全ての文字コード・文字・正書法の分析の結果、テキスト編集が言語非依存であることが明らかとなったため、全ての文字を同時混在使用可能な機構(国際化テキスト処理機構)の構築 [1, 2] が行なわれた。

テキスト処理は、入力・出力・テキスト操作・プロセス間通信に分類できるが、入力は、キーボード等外部入力装置からのイベントをコードポイント列に変換する機構である。その実現には出力等が要求されるため、国際化テキスト処理機構を利用したアプリケーションとして実現される [3]。

入力機構の機能は、1) イベント/コード変換、2) 変換補助のためのユーザインタフェースに分類可能である。そこで本稿では、入力対象・インタフェースの分析を行ない、一般化された入力機構のモデル構築について述べる。

2.入力対象の分析

ユーザはキーボード等を使用して、文字を入力する。しかし、文字の使用環境は言語により異なるため、入力方法は言語の影響を受けている。

そこでまず入力機構に必要な機能を分類するために、入力対象を分析する。入力された文字は、テキスト操作、出力、プロセス間通信に使用される。これらの処理単位はそれぞれ 1) 文字、2) 図形(フォント)、3) 文字コード (mb) であり、これらを入力対象とする(図 1)。

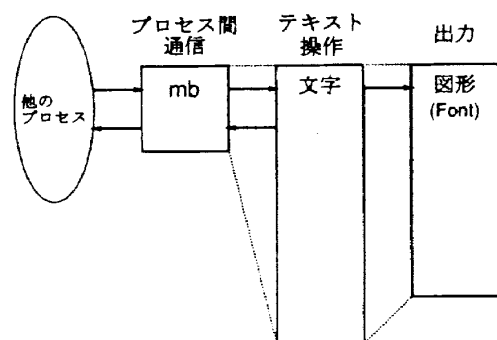


図 1. 処理単位の変換関係

A Model of The Internationalized Text Input Method
Kazutomo Uezono†, Tomoko Kataoka*, Yutaka Kataoka‡ and
Katsuhiko Kakehi†

† School of Science and Engineering, Waseda University

* Media Network Center, Waseda University

‡ Department of Literature, Otani University

テキスト処理は、処理単位の変換により行われる。このうちテキスト処理で編集されるのは文字であるため、3つの処理単位中最も多く情報を持つ。文字の情報は、文字自身を示すインデックスと属性情報に分類されるので、入力は「入力されたイベントから(文字単位)+(文字属性情報)への変換を行なう機構」と定義される。

3.文字単位への変換機能の分析

言語はspeaking languageとwritten languageが存在し、前者は音、後者は文字を伝達媒体とする。これより入力方法は、1) 音に対応した記号の入力と2) 文字を表す記号の入力に分類できる。また、文字は指示する対象により、a) 表音文字とb) 表意文字に分類される。よって文字の入力時に、

- ・ イベント→表音文字 (1a, 2a)
- ・ イベント→表意文字 (2b)
- ・ イベント→表音文字→表意文字 (1b)

の変換の可能性が存在する。

イベントから表音文字への変換は、ラテンアルファベットのように1イベントで1文字に変換される場合と、ハングルのように複数イベントで1文字に変換される場合があるので、

- ・ イベント : 文字 = 1 : 1
- ・ イベント : 文字 = M : 1 (Mは複数の要素)

となる。これに対し、表意文字は、非常に文字数が多いため、変換方法として

- ・ 文字を構成する入力記号列から表意文字へ変換
- ・ 文字の音を示す表音文字列から表意文字へ変換

が存在する。前者はイベント : 文字 = M : 1となる。後者は、表音文字列を入力し、その音を持つ表意文字集合から選択する。

以上より、イベントから文字単位への変換方法を分類すれば、

- 1) イベント : 文字 = 1 : 1 の変換
- 2) イベント : 文字 = M : 1 の変換
- 3) 表音文字列から表意文字への言語依存変換

となる。1)、2)の写像はそれぞれ単一の写像関数であり、一般化が可能である。3)は言語に依存するので、言語毎に変換部が必要である。全ての変換部に入力として表音文字列を渡し、その変換結果を管理する変換部管理機構を用意すれば、言語毎に変換部を切替える必要はなく、変換部管理機構が変換結果をフィルタリングして返せばよいので、変換機構は言語依存とはならない。よって、一般化した変換機構は実現可能である。

4.文字属性情報の分析

文字の分析より、文字は1) - 5)の情報を持つ。

- 1) 文字自身を指示するインデックス
- 2) 文字操作情報(禁則処理情報等)
- 3) 文字コード情報(文字コード名等)
- 4) 最終表示図形の指定(リガチャ、変字体等)
- 5) 配置情報(表示方向等)

1) - 3) は文字自身を持つ固有情報であり、文字単位に変換した時点で決定される。4) と 5) は入力時にユーザが明示的に文字に与える情報である。

また、ハイフネーションなどの言語依存テキスト処理を行なうため、6)と7)の情報を持つ。

6) 言語名

7) 言語処理情報(ハイフネーション情報等)

6) と 7) は、同じ文字でも言語により値が異なるだけであり、文字の属性情報として、1) - 5) と同様に扱うことが可能である。

従って単一の入力機構により、全ての文字に対して属性情報を付加することが可能である。

5.ユーザインタフェースの分析

入力機構は、ユーザの円滑な入力の補助や文字属性情報の指定のため、ユーザインタフェースを提供する。ユーザインタフェースの分析より、次の1) - 3)が必要である。

1) 入力記号列の表示・編集

表示・編集の範囲は一行となる。そこで、One Line Editorを用意し、全ての入力記号列・変換後文字列を表示し、編集可能とする。

2) 文字属性情報の選択候補の提示

文字図形等の選択候補をユーザに提示する。

3) 入力環境情報の提供・変更

a) キーマップ等入力機構内部状態の提示・変更

一つのキーボードで複数言語入力をサポートするために、キーマップを動的に変更可能とし、キーマップ情報を提供する。また入力機構が保持しているエンコーディングや表示方向等の情報を提示・変更可能とする。

b) 言語依存変換部の内部状態の提示・変更

変換部で使用している辞書名等の情報を提示・変更可能とする。

1)、2)、3)のa)は入力機構が汎用機能として提供可能である。3)のb)は各変換部が提供するが、変換部管理機構の機能とすれば、変換部に依存しない。よって1) - 3)は単一の入力機構で実現可能である。

6.国際化入力機構モデルの構築

変換機構は、国際化テキスト処理機構を利用するアプリケーションとして実現される。コード変換は、コード変換機構(Meta Converter System)を、国際化出力は、国際化出力機構を利用する。

国際化入力機構は、1) アプリケーションとのインタフェース部(Input Method Library)、2) 入力機構のカーネル部(Input Method Manager; IMM)、3) 言語依存変換部(Language Specific Converters)の3つのモジュールで構成する(図2)。1) はOS等に依存したライブラリとして実現する。2) では、イベントの文字列への変換、変換文字列の編集を行なう。言語依存変換部を制御することで、カーネルは非言語依存となり入力機構が一般化可能となる。GUIの制御も行なう。3) は IMM より表音文字列を受け取り、表意文字に変換した結果をIMM に返す。

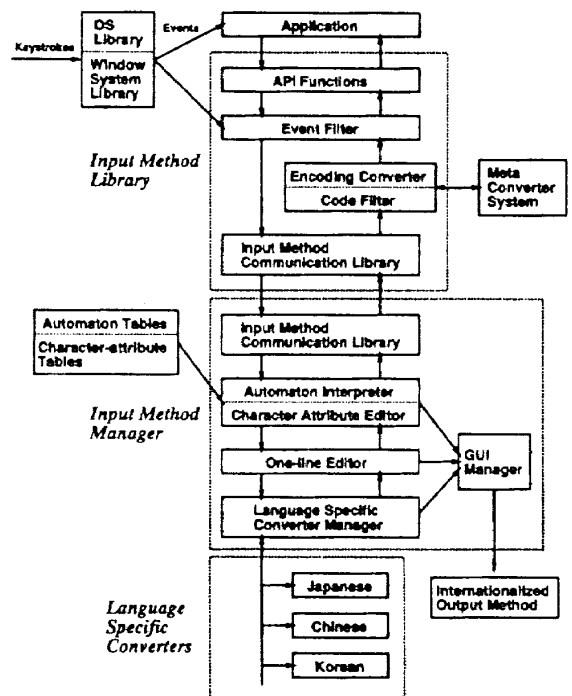


図2. 国際化入力機構のモデル

7.まとめ

国際化入力機構の実現性を示し、モデルを構築した。この国際化入力機構を利用することで、多文字・多言語エディタ等のアプリケーション自身に入力機能を組み込むことなく、比較的容易に複数言語入力機能を実現することができる。

なお本研究は、早稲田大学特定課題研究助成費(98A-910)を受けている。

参考文献

- [1] Kataoka, Y., et al., A Model for Input and Output of Multilingual text in a Windowing Environment, ACM TOIS, Vol. 10, No. 4, pp 428-451, 1992.
- [2] Uezono, K., et al., The Internationalized Environment Enabling Proper Text Processing, 第55回情報処理学会全国大会予稿集, 分冊4, pp 47-48, 1997.
- [3] Daikokuya, H., et al., The Worldwide Multilingual Computing (6): Multilingual Text Interprocess Communication and Input Mechanism, 第51回情報処理学会全国大会予稿集, 分冊3, pp 255-256, 1995.