

張陽 梅村恭司 中島浩

豊橋技術科学大学 情報工学系

1. はじめに

複数の言語によって書かれた文書の処理は、Unicode や ISO2022 のような多言語対応のコード体系を用いて行なわれている。しかしこの国際化（あるいは地域化）の方法の対象は、OS もしくは入力方法の制限で、英語とそれ以外の特定の言語（例えば日本語）の 2 言語で書かれた文書にほぼ限定され、3 種以上の言語や、日本語と中国語のような組合せには対応できないことが多い。また、特殊の方法を使い、作られた多言語文書は元の OS 及びアプリケーションに依存しがちである。

そこで本研究では、任意の種類の言語を含む文書処理の枠組として、個々の文字を GIF 形式のフォントファイルに対応付ける多言語 HTML 形式を考案した。この枠組は文字コードと独立であるため、多言語への対応が容易であるだけでなく、文字コードセットや対象言語の拡張も簡単に行なうことができ、OS 及びアプリケーションに依存しない。また本研究では、この枠組を利用した文書処理システムとして、通常の文字コードで書かれた複数のテキストファイルから、多言語 HTML 形式のファイルを生成するエディタを構築した。さらに、多言語テキストを含むプログラムの生成ツールと、文書中の単語の意味を複数の辞書を検索して求めるツールも作成した。

2. 多言語テキストエディタ

多言語テキストエディタは、複数の通常テキストファイル（あるいは各言語の HTML ファイル）をソースとし、多言語 HTML ファイルを生成するためのシステムである。個々の通常テキストファイルは、汎用の変換ツール TxtToHtml により多言語 HTML 形式に自動変換され、元のテキストイメージでマルチウィンドウ表示される。ユーザは cut & paste などの編集機能を用いて、個々のウィンドウから文字列を抽出することにより、ターゲットの多言語 HTML ファイルを容易に生成することができる。

このシステムには他に、多言語 HTML 形式のテキスト操作のための諸機能を実現するツールキットと、BDF 形式のフォントファイルから GIF 形式のフォントライブラリを構築するツールである BdfToGif も含まれている。またシステムは全て Java で記述されており、OS や機種に依存しない高い移植性を持っている。

3. 多言語対応プログラム生成ツール

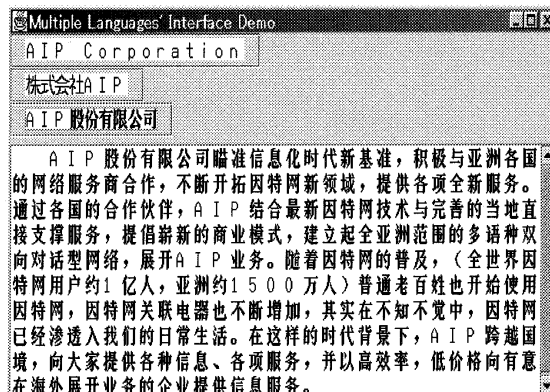
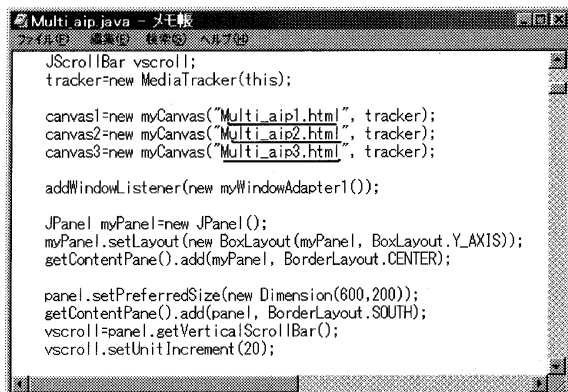
Java の国際化の枠組では、個々の言語ごとのロケール指定と、それに対応するテキストリソースファイルを用意することにより、一つのプログラムの様々な言語環境での利用を実現している。しかしこの枠組は、日本語と中国語といった複数の言語に同時に対応することはできず、また OS でサポートされていない言語への対応も極めて難しい。

一方、我々の枠組を用いれば、多言語テキストエディタで Java ソースを作成することにより、複数の言語のテキストを直接プログラム中に埋め込むことができる。そこで多言語 HTML 形式のソー

スを、Java 処理系が認識可能な形式に変換するツールを作成した。

このツールはまず、ソース中の多言語テキストを抽出し、それぞれのテキストについて対応する多言語 HTML ファイルを生成する。次にテキストを、生成した多言語 HTML ファイルへの参照に置き換え、最後にプログラム全体を ASCII コードに変換して、処理可能な Java ソースを得る。

下の左図はツールで生成した多言語文字を持つ Java プログラムのソースの一部である。下線が付いているところは多言語 HTML ファイルへの参照である。下の右図はこの Java プログラムの実行例であり、「株式会社 AIP」や「AIP 股份有限公司」（日中英混在）といったテキストは、直接ソース中に埋め込んで記述されている。



4. 多言語辞書検索ツール

多言語文書を作成する際には、作成者の母国語以外の言語についての辞書を参照したくなることがしばしばある。たとえば日本人が、日、中、英の三カ国語からなる文書を作成する際には、日中／中日や和英／英和の辞書のオンライン検索は必要不可欠である。そこで、多言語テキストエディタの拡張機能として、指定された単語の意味を、指定された言語で表示するツールのプロトタイプを作成した。

このツールの重要な特徴は、辞書の不備や辞書自体の欠落により直接的な検索ができない場合でも、複数の辞書を順次検索して指定言語での意味を求める機能である。たとえば「インターネット」を意味する中国語「因特网」が中日辞書にない場合でも「因特网→Internet」（中英）と「Internet→インターネット」（英和）の二つの辞書検索によって、日本語での意味表示を行なうことができる。

5. おわりに

本研究では、我々が考案した多言語 HTML 形式の枠組を利用して、テキストエディタ、プログラム生成ツール、および辞書検索ツールを作成した。またこれらのツール群の開発を通じ、多言語 HTML 形式が種々のテキスト処理を容易に実現できる枠組であることを実証することができた。

参考文献

- 「1」 Ken Lunde：日本語情報処理，ソフトバンク株式会社（1995）
- 「2」 有賀妙子・竹岡尚三：Java1.1プログラミング，ソフトバンク株式会社（1997）
- 「3」 錦見美貴子・高橋直人・戸村哲・半田剣一・桑理聖二・向川信一・吉田智子：
マルチリンガル環境の実現，プレントイスホール株式会社（1996）
- 「4」 張暘、梅村恭司：文字コード独立の多言語テキストエディタの実装，情報処理学会第58回
全国大会講演論文集（3），pp.19-20（1999）