

ブラウザ側にフォントを必要としない 多言語 HTML 文書の提供システム

前田 亮[†], ミリアン ダルトア[‡], 太田 純[‡], 阪口 哲男[‡], 杉本 重雄[‡], 田畑 孝一[‡]

[†]奈良先端科学技術大学院大学 情報科学研究科

[‡]図書館情報大学

e-mail: aki-mae@is.aist-nara.ac.jp, {myriam, jun, saka, sugimoto, tabata}@ulis.ac.jp

1. はじめに

WWW 上で多言語からなる HTML 文書を提供しようとする際に問題となるのは、HTML や使用される文字コードの規格自体が多言語に対応していないだけでなく、表示に必要なフォントがクライアント側に必ずしも存在しないことが挙げられる。例えば日本語をベースとする計算機環境では、標準では英語と日本語のフォントしか持たないことが多い。その他の言語の文書を表示するには必要なフォントをインストールしなければならない。WWW の一般的な利用者にとってこれは大きな負担であると考えられる。また、将来的に Unicode などの多言語に対応した符号化方法が普及したとしても、各クライアントにそれに含まれるすべての文字のフォントが用意されるとは考え難い。

そこで我々は、利用者によるフォントのインストールを必要とせずに多言語からなる HTML 文書の提供を可能とするシステムを開発した¹⁾²⁾。これは利用者から要求された URL で示される文書を後述の MHTML 文書形式に変換し、それを利用者の WWW ブラウザ上の MHTML ビューアで表示・ナビゲーションを行うものである。このシステムはゲートウェイサービスとしてインターネット上で公開している*。本稿では、これをもとに開発した多言語からなる HTML 文書の提供を容易に可能とするシステムについて、その概要と本システムを利用した多言語文書の提供例について述べる。

2. MHTML 文書形式

本システムの実現のため、Multilingual-HTML (MHTML) 文書形式を導入した。MHTML 文書とは、通常の HTML 文書にその文書で使用されている文字

のみのフォントのビットマップ情報を付加したものである (図 1)。

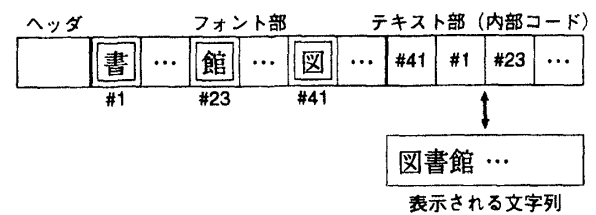


図 1 MHTML 文書形式

フォント部には、文書に含まれる文字のみのフォントグリフ (ビットマップ) が格納される。テキスト部は、HTML のタグの部分のみ ASCII コードが用いられ、それ以外はフォント部に格納された順に振られた内部コードによってエンコードされる。

この MHTML 文書を表示するために、ブラウザ側で MHTML 文書の解釈と表示を行うのが MHTML パネルクラスである。

3. 提供システムの実装

提供システムは、HTML 文書から MHTML 文書への変換および MHTML パネルクラスの提供を行う MHTML サーバと、源となる HTML 文書を保持する通常の WWW サーバによって構成される (図 2)。

3.1 MHTML サーバ

MHTML サーバは、源となる HTML 文書を WWW サーバから取得し、これに必要最小限のフォントを付加した MHTML 文書に変換してブラウザ側に提供する機能と、MHTML 文書を表示するための MHTML パネルクラスをブラウザ側に送る機能を持つ。

3.2 MHTML パネルクラス

MHTML パネルクラスは Java アプレットとして実現しており、ブラウザ上に任意の大きさの領域を確保

Multilingual HTML Document Delivery System without Fonts on the Browser's Terminal

Akira Maeda[†], Myriam Dartois[‡], Jun Ohta[‡], Tetsuo Sakaguchi[‡], Shigeo Sugimoto[‡] and Koichi Tabata[‡]

[†]Graduate School of Information Science, Nara Institute of Science and Technology

[‡]University of Library and Information Science

*URL: <http://mhtml.ulis.ac.jp/>

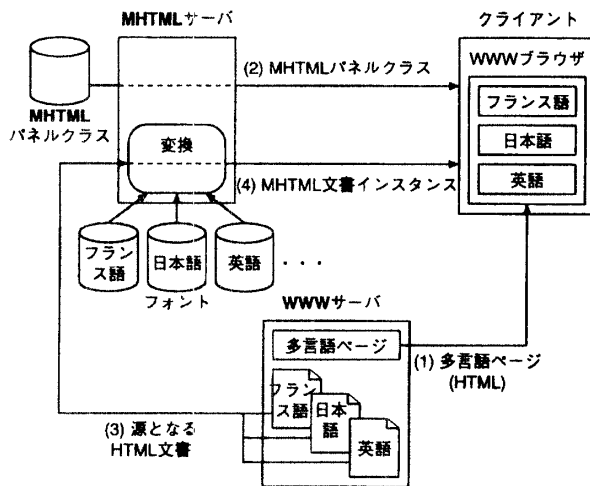


図2 提供システムの概要

し、その上に MHTML 文書を表示する。このパネルは一ページ中に複数含むことも可能である。

ブラウザ側ではまず MHTML パネルクラスをロードするための HTML ページを読み込む。この HTML ページにはアプレットのパラメタとして以下の情報が含まれており、これらが MHTML パネルクラスを通して MHTML サーバに渡される。

パラメタ名	値
url	源となる HTML 文書の URL
lang	源となる HTML 文書の言語（符号化方法）の識別子

MHTML サーバでは url で示される文書を WWW サーバから取得し、lang の情報を用いて MHTML への変換を行い、ブラウザ上の MHTML パネルクラスに返す。これにより、ブラウザ上の MHTML パネルクラス上に文書が表示される。

3.3 テキスト入力クラス

フォームなどで必要な文字の入力は、テキスト入力クラスによって実現している。これは、日本語のフォントや入力メソッドを持たない環境からローマ字を入力し、MHTML サーバとの通信によってローマ字漢字変換を実現するものである。この場合は変換候補の文字のフォントのみならず、JIS コードを付加した拡張 MHTML 文書として MHTML サーバから得る。

4. 多言語文書の提供例—日本昔話の提供システム—

本システムを利用した多言語 HTML 文書の提供例として、日本昔話の日本語・英語・フランス語の3ヶ国

語による提供を行っている³⁾。本文の表示に MHTML を用いることで、世界中どこからでもフォントなどのインストールを一切必要とせずに3ヶ国語で日本昔話を読むことが可能になっている。

3ヶ国語による本文を一画面に同時表示した例を図3に示す。

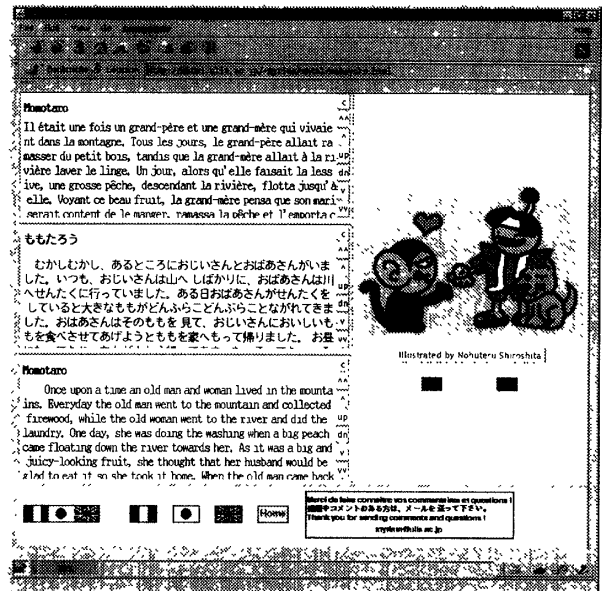


図3 多言語文書の提供例

5. おわりに

本システムの今後の課題としては、日本語以外の言語における入力機能の実装、MHTML 文書の印刷機能の実現などが考えられる。

謝辞： 発表の機会を与えていただいた奈良先端科学技術大学院大学の植村俊亮教授に感謝致します。

参考文献

- 1) Sakaguchi, T., Maeda, A. et al.: A Browsing Tool for Multi-lingual Documents for Users without Multi-lingual Fonts, in *Proc. of 1st ACM International Conference on Digital Libraries*, pp.63-71 (1996).
- 2) 前田 亮 他：WWW 文書のための多言語ブラウザとそのゲートウェイサービス, 情報処理学会研究報告 (96-FI-44), Vol. 96, No. 116, pp.1-7 (1996).
- 3) Dartois, M., Maeda, A. et al.: Building a Multilingual Electronic Text Collection of Folk Tales as a Set of Encapsulated Document Objects: an Approach for Casual Users to Browse Multilingual Documents on the Fly, in *Proc. of 1st European Conference on Research and Advanced Technology for Digital Libraries* (1997年9月発表予定).

[†]URL: <http://www.DL.ulis.ac.jp/oldtales/>