

海外利用者のための日本語 OPAC

藤田岳久, 前田亮, 阪口哲男, 杉本重雄, 田畑孝一
図書館情報大学

〒305 茨城県つくば市春日 1-2

Tel: 0298-52-0511 Fax: 0298-52-4326

E-mail: {take, maeda, saka, sugimoto, tabata}@ulis.ac.jp

概要

我々の開発した海外利用者のための日本語 OPAC(MOPAC) は、日本語環境を持たない WWW ブラウザから日本語資料の検索を行うための OPAC である。MOPAC は、我々が開発したワークステーションベースの OPAC を基に、我々の提案している多言語 HTML に基づく MHTML エンコーダおよび MHTML ブラウザと EDR 電子化辞書により構成される。MOPAC の利用者は、ローマ字または英単語からの検索が可能であり、英単語は電子化辞書により日本語単語に翻訳されて検索が行われる。今後、電子化辞書を利用して、ローマ字漢字変換機能、類義語検索などの拡張を行う。

キーワード

OPAC, 多言語環境, 電子化辞書, WWW

Japanese OPAC for Overseas Users

Takehisa Fujita, Akira Maeda, Tetsuo Sakaguchi, Shigeo Sugimoto, Koichi Tabata

University of Library and Information Science

1-2 Kasuga, Tsukuba, Ibaraki, 305, JAPAN

Phone: +81-298-52-0511 Fax: +81-298-52-4326

E-mail: {take, maeda, saka, sugimoto, tabata}@ulis.ac.jp

Abstract

Japanese OPAC for overseas users, called MOPAC, is a WWW-based OPAC system which doesn't require the Japanese browsing environment. MOPAC consists of the OPAC system we have developed on workstations, MHTML encoder and browser based on the Multi-lingual HTML we have proposed, and the EDR Electronic Dictionary. MOPAC users can use "roma-ji" words and English words as retrieval terms, and English words are translated into Japanese words before retrieval. We are planning to incorporate "roma-ji to kanji" translation function and related term retrieval function into MOPAC, using the EDR Electronic Dictionary.

Keywords

OPAC, Multi-lingual environment, electronic dictionary, WWW

1. はじめに

インターネットによる OPAC の公開は、米国では以前からごく普通に行われてきており、最近日本でも多数の図書館が OPAC の公開を行うようになってきた。また、WWW による公開も見られるようになってきた [1]。

図書館においては一般的に、収集する資料の使用言語の制限は行っていない。よって、OPAC も様々な言語に対応することが望ましい。しかし、WWW ブラウザで表示できる文字はクライアントマシンの持つフォントに依存しており、ほとんどのマシンが自国語以外には ASCII のフォントしか持っていないという現状では、WWW サーバが様々な言語による OPAC を提供したとしても対応するフォントがない言語は正しく表示されない。また、フォントセットに含まれない文字、例えば「秘」という字を丸で囲んで「まるひ」と読ませるような「合字」はサーバで提供することができない。

我々の開発した「海外利用者のための日本語 OPAC」(以下 MOPAC) は、多言語対応 OPAC へのアプローチの第一歩として、日本語環境を持たない WWW ブラウザからの利用を可能とした OPAC である。MOPAC は、我々の提案している MHTML および MHTML ブラウザ [2][3][4] を利用して、多言語表示を可能としている。ローマ字による日本語資料の検索が可能なので、日本語入力環境を持たないマシンでも利用することができる。さらに、英和辞典を内蔵することにより入力した英語の単語や句を日本語単語や句に翻訳して検索することができ、日本語能力が十分でなく英語を解する外国人利用者でも「意味のわかる言葉」からの日本語資料の検索を行うことができる。

本稿では、MOPAC の構成および利用について述べる。

2. MOPAC の構成

MOPAC は、我々が開発したワークステーションベースの OPAC [5][6] (以下 XOPAC) のデータベースと検索エンジンを核に、WWW サーバ、MHTML エンコーダ、外字フォントサーバ、EDR 電子化辞書 [7][8] によって構成している。また、利用の際は WWW ブラウザ (Mosaic) と MHTML ブラウザを用いる。XOPAC のデータベースと検索エンジンには一切手を加えていない。

図 1 に MOPAC の構成を示す。WWW ブラウザから発せられた検索/翻訳要求は WWW サーバから MOPAC インタフェースに渡される。検索要求ならば XOPAC にアクセスし、翻訳要求ならば電子化辞書にアクセスする。いずれの場合も結果は MHTML エンコーダで変換される。MHTML エンコーダは必要に応じて外字フォントサーバにアクセスする。WWW ブラウザは、返された MHTML 文書を MHTML ビューアによって表示する。

以下、各構成要素について解説する。

• XOPAC

大型計算機上の OPAC をワークステーションに移植したものである。大型計算機からダンプしたデータをフォーマット変換してワークステーションに転送し、SONY MediaFinder によってデータベース化した。漢字、カタカナによる読み、ローマ字 (訓令式、ヘボン式) による読み、アルファベットによる検索が可能である。また、X-Window による専用のグラフィカルユーザインタフェースを持つ (今回は使用していない)。

• MHTML エンコーダと外字フォントサーバ

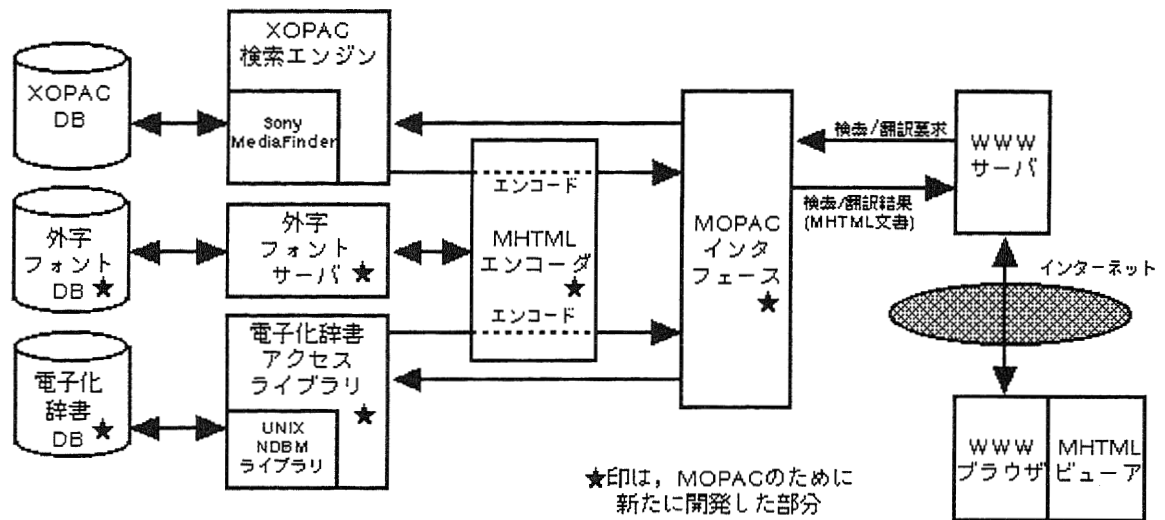


図1 MOPAC の構成

MHTML エンコーダは、データベースから検索された書誌記述を MHTML 文書に変換する。書誌記述が前述の合字のような「外字」を含んでいる場合は、外字フォントサーバからフォントデータを得て MHTML 文書を構成する。

我々は既に汎用の MHTML エンコーダ (MHTML ゲートウェイ [4]) を開発しているが、MOPAC においては変換するデータの構造 (タイトル、著者名などの書誌事項) を意識する必要があるため、また、外字フォントサーバとやりとりをする必要があるため、MOPAC のためのエンコーダを新たに開発した。

- MHTML ビューア

MHTML 文書は、テキストデータとそれを表示するためのフォントデータを含んでおり、読むためには専用のビューアが必要である。今回は、我々が既に開発した MHTML ビューアをそのまま利用した。

- EDR 電子化辞書

(株) 日本電子化辞書研究所がコンピュータによる言語処理のために開発した大規模な機械可読辞書である。日英それぞれの単語辞書、日英・英日対訳辞書、言葉の係り受けの関係を収めた共起辞書、単語辞書中で定義した概念の類義を記述する概念辞書 (シソーラス)、辞書記述の典拠としてのコーパスデータベース (例文集) から成る。今回は、英日対訳辞書を利用した。本辞書はプレーンテキスト形式で提供されているので、UNIX 標準の簡易 DBMS である NDBM を利用してデータベース化した。

3. MOPAC の利用

利用者が WWW ブラウザを用いて MOPAC に検索要求を発すると、MOPAC はデータベースの検索を行い、結果 MHTML 形式で返す。WWW ブラウザは MHTML 文書を受けると MHTML ビューアを起動し表示する。また、利用者が英単語を与えて翻訳を指示すると、MOPAC は辞書を用いて日本語単語をリストアップする。利用者がいずれかの日本語単語を選択するとその単語によって OPAC を検索する。

図 2 に MOPAC の検索語入力ページの様子を示す。ローマ字による検索語を入力し、検索範囲 (Title, Author など) を指定して検索を開始すると、図 3 のように MHTML ビューアによってタイトルと著者名などの簡略結果表示を行う。いずれかを選択すれば、図 4 のように詳細表示を行う。なお、図 4 の詳細表示には、外字フォントサーバに取められた文字 (「知」という字を丸で囲んだ字) の表示が含まれている。

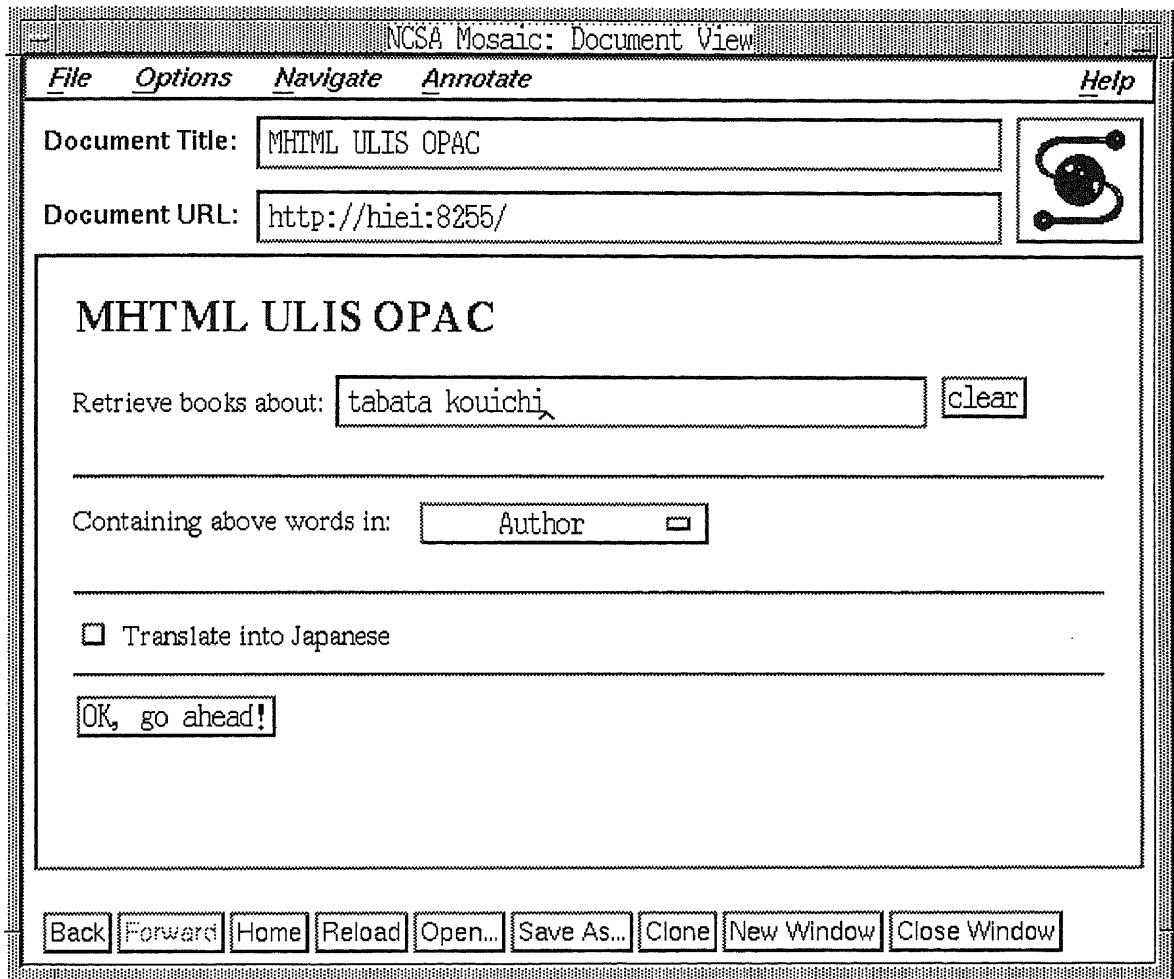


図 2 検索語入力ページ

検索語入力の際に英語単語を入力し、“Translate into Japanese” をクリックして検索を開始すると、図 5 のように日本語単語をリストアップする。いずれかをクリックするとその語によって検索を行い、結果の簡略表示を行う。

4. おわりに

MOPAC はローマ字からの日本語資料の検索を可能としているが、これには「同音異義語による検索精度の低下」という問題が残る。EDR 電子化辞書の日本語単語辞書は各単語の読みを取めているので、「ローマ字漢字変換機能」をサーバに持たせれば (例えば、MOPAC の翻訳機能と同様に、「koukou」と入力し漢字変換を指示すると「高校」「孝行」「航行」などの候補一覧を表示し、利用者がいずれかを選択するとそ

の語によって検索を行う)、この問題は解決できる。また、EDR 電子化辞書の持つ概念辞書を利用すれば、検索語の類義語・上位概念語・下位概念語による検索が可能となり、網羅的な検索を行うことができる。いずれも今後実現の予定である。

参考ドキュメント

- [1] 林賢紀. 日本国内図書館 OPAC リスト Ver 13.0. <http://ss.cc.affrc.go.jp/ric/opac/opac.html>
- [2] 前田亮他. 組み込みフォントを必要としない WWW のための多言語ブラウザ. デジタル図書館. No. 4, p.21-25(1995) (<http://www.dl.ulis.ac.jp/DLjournal/No4/maeda/maeda.html>)
- [3] A. Maeda; et al. A Multilingual Browser for WWW without Preloaded Fonts. Proceedings of 1995 International Symposium on Digital Libraries. p.269-270, 1995.
- [4] T. Sakaguchi; et al. A Browsing Tool for Multi-lingual Documents for Users without Multilingual Fonts. (1996 年 3 月に米国 Maryland 州 Bethesda で行われる Digital Library '96 にて発表予定。 <http://fox.cs.vt.edu/DL96/> を参照)
- [5] 藤田岳久他. 分散環境を利用した CD-ROM オンラインカタログシステム. 情報処理学会研究報告, Vol.93, No.39, p.43-50, 1993. (情報学基礎研究会 FI-29)
- [6] T. Fujita; et al. Transporting an Online Public Access Catalog from Mainframe to Distributed Environment. Proceedings of 47th FID Conference, p.444-448, 1994.
- [7] EDR と EDR 電子化辞書. <http://www.ijnet.or.jp/edr/Intro.html>
- [8] EDR 電子化辞書仕様説明書. <http://www.ijnet.or.jp/edr/TG.html>

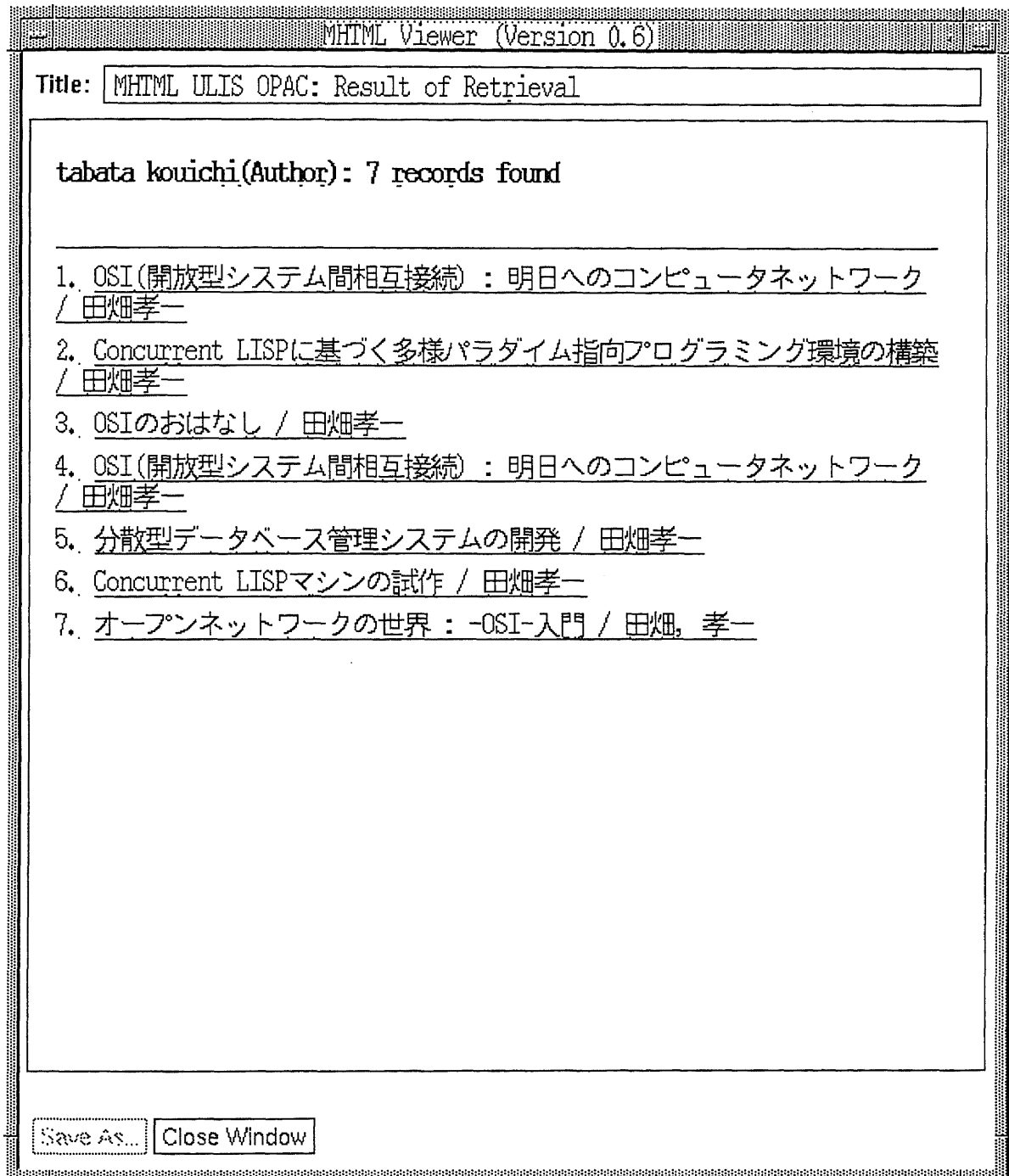


図3 検索結果の簡略表示

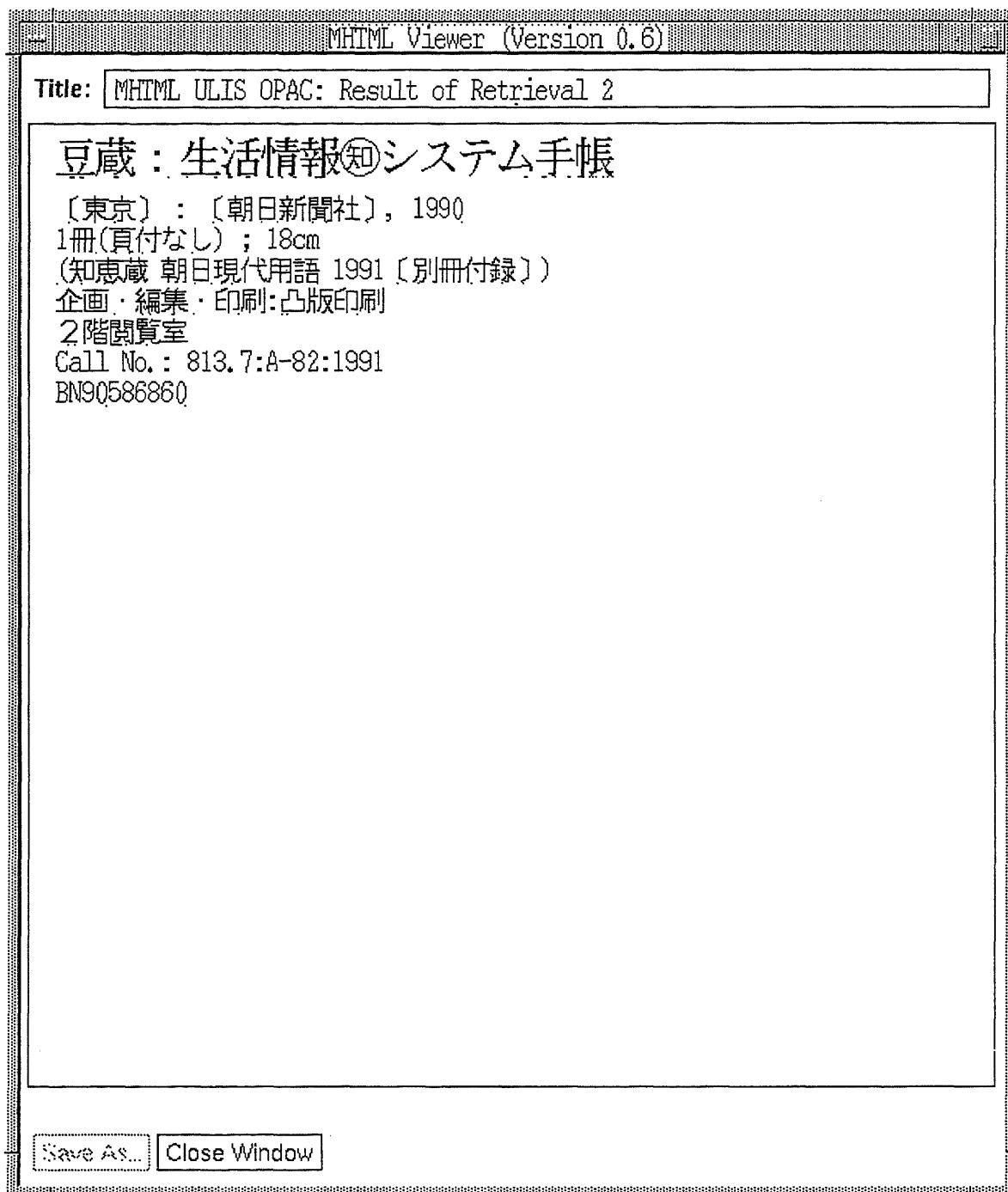


図 4 検索結果の詳細表示

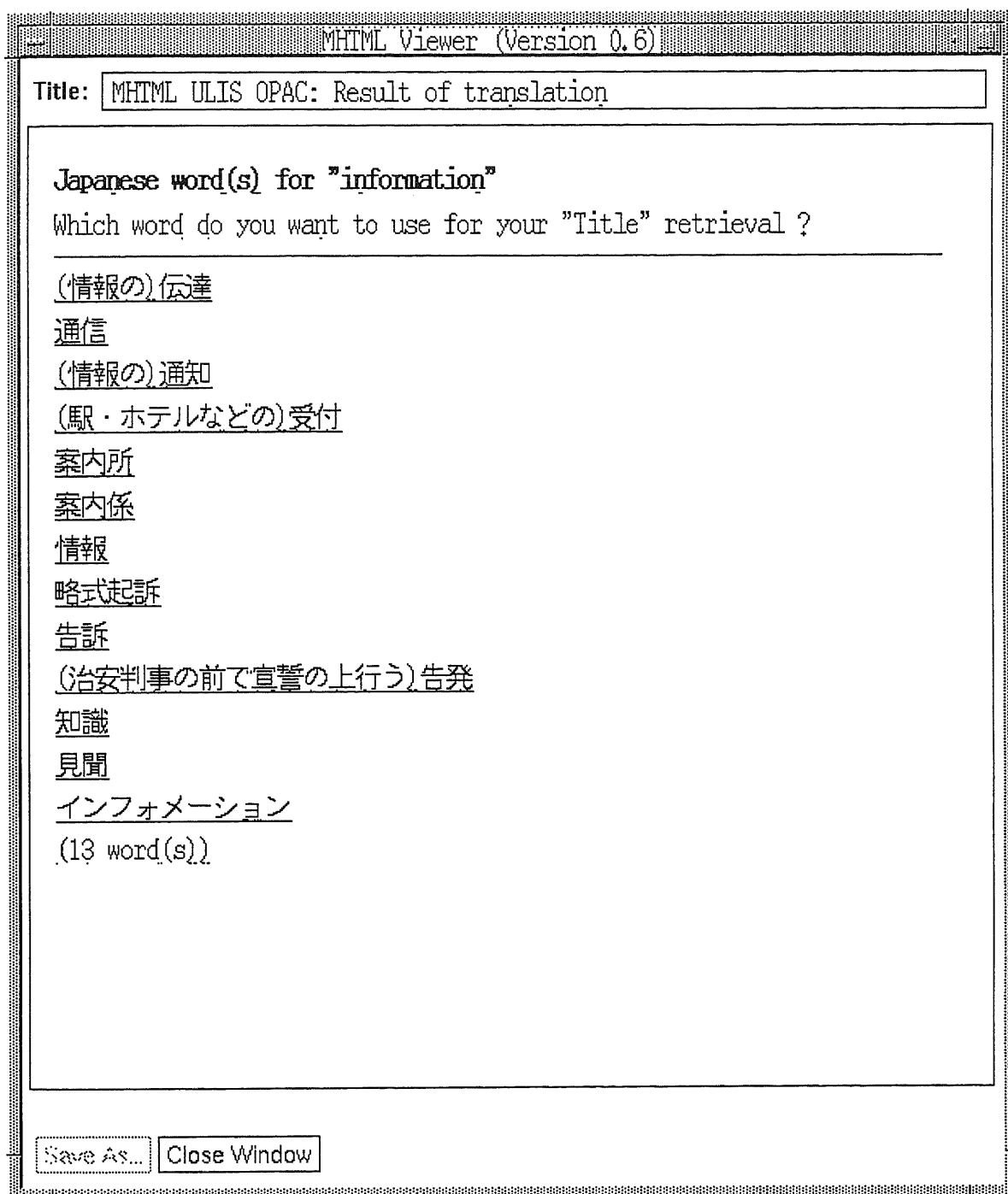


図 5 英語単語の日本語単語への翻訳結果の表示