

会員ネットワーク「ネットで百科」の概要と実現法

1 J-1

澤田 瑞穂 山崎 幹夫

藤井 泰文

西岡 真吾 高野 明彦

金田 泰

日立東北ソフトウェア

日立デジタル平凡社

日立製作所 基礎研究所

同 中央研究所

1. はじめに

WWWの普及に伴い、コンテンツの内容・質に加えてコンテンツの構造や検索結果に対する視覚化(ビジュアルイゼーション)への関心が高まってきている。また、ポータルサイトが注目を集めているように、インターネット上にある様々な知に対する玄関口が求められている。「ネットで百科」は日立デジタル平凡社が提供する「知への入り口」を目指したインターネットを基盤とする会員ネットワークである。同社の技術およびコンテンツの集大成である世界大百科事典 [HDH 98] の次世代の検索機能に加え、WWWを利用したコンテンツの定期出版を可能とするプラットフォームを実現した。操作性とコンテンツの再利用性を両立させるため、検索操作は独自に開発したクライアントを用い、コンテンツの表示にはWWWブラウザを用いる構成とした。本稿では「ネットで百科」の概要とクライアントプログラムを中心とした実現法について説明する。

2. 「ネットで百科」の概要

「ネットで百科」は、クライアントとサーバから構成される。クライアントは2種類の次世代の検索インタフェースを利用者に提供する。この2つの検索インタフェースは日立製作所基礎研究所および同中央研究所が開発した検索機能をそれぞれベースとしている。

・デュアル連想検索…日立製作所 基礎研究所の研究成果であるDualNAVY検索モデルに基づいた、単語の共起関係による世界大百科事典の連想検索を可能とした検索インタフェースである。

・テーマ年表検索…日立製作所 中央研究所の研究成果である「軸づけ検索」機能の適用により、世界大百科事典に含まれる年代からのダイナミックな年表作成を可能としたインタフェースである。

実際の検索はそれぞれの検索機能に対応したサーバにインターネット接続することで実行できる。サーバは検索機能とコンテンツへのアクセス管理の仕組みを提供

し、適切な方法で認証を受けた利用者にもみサービスを行う。

また、これらの検索機能とは別に、デジタル月刊百科という名称により、様々なコンテンツをWWWを使って定期出版する計画も推進中である。

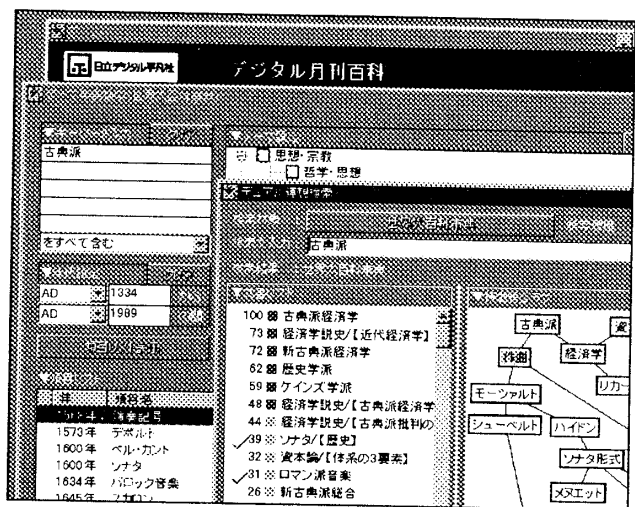


図 1. 画面例(部分)

3. 「ネットで百科」の実現法

「ネットで百科」の実現にあたっては、前項で述べた2つの検索機能のアーキテクチャを大きく変更することなく、クライアントに関しては世界大百科事典の特徴をそのまま生かし、サーバに関しては会員ネットワークの基盤を構築することが要求された。つまり、両検索機能のアーキテクチャをベースとして、最も利用者側の部分と、最もシステム側の部分に関して両者を密接に統合する必要があった。主な統合化内容を以下に示す。

3.1 GUIの統一化

両検索機能は、全く個別に研究されてきたという経緯があり、そのアーキテクチャやGUIにおいては異なりがある。これらを世界大百科事典のGUIと違和感がない形で統一化し、かつWindowsのMDI(Multi Document Interface)形式の1つのアプリケーションとして統合した。

* The Outline and Implementation method of "Pedia by net", by Mizuho Sawada and Mikio Yamazaki (Hitachi Tohoku Software, Ltd.), Yasufumi Fujii (Hitachi Digital Heibonsha Co.), Shingo Nishioka, Takano Akihiko (ARL, Hitachi, Ltd), and Yasusi Kanada (Hitachi Ltd).

3.2 外字処理方式の統一化

CD-ROM版世界大百科事典は、Windows環境で動作することを前提に設計されており、その制約からやや特殊な外字処理方式を採用している。通常であれば、フォントが持つ外字領域を使用するところを、数千にも及ぶ外字を実現するために外字専用のフォント(HDHゴシックフォント)を提供し、通常のフォント(MSゴシックなど)との切り替えを行うことでの混在表示を可能としている。

このような、マルチフォントに対応した汎用的なソフトウェアコンポーネントは存在しないため、「ネットで百科」専用のコンポーネントとして独自に開発した。中でも、様々な入出力要求およびフォーマットへの対応を要するエディットボックスに関しては、「ネットで百科」共通の外字処理方式も合わせてその実装を検討する必要があった。エディットボックスを中心とした入出力のパターンを図2に示す。

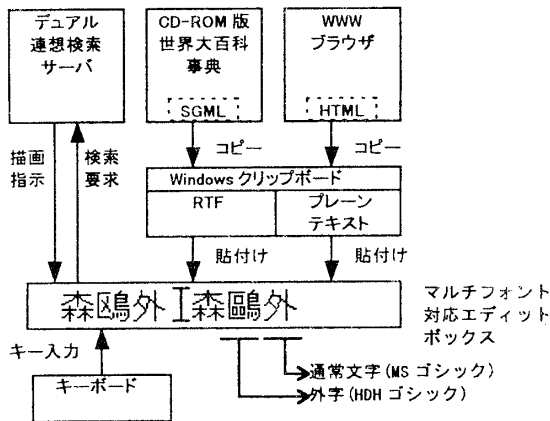


図2. エディットボックスの入出力パターン

この中で、特に問題となるのはWWWブラウザからコピーされたHTML中の外字の扱いである。WWWブラウザで外字を含んだHTML文書を表示することは、FONTタグにより実現できる。しかし、HTML文書をクリップボードに複写した場合、クリップボード内ではフォントに関する情報は失われてしまう。このため、このデータをエディットボックスに貼付けると通常文字として認識されてしまうことになる。そこで、世界大百科事典では使用されていない特定の文字を外字の標識(インジケータ)として特別に扱い、内部的には外字を「外字コード+インジケータ」のペアで扱う方式とした。このインジケータは、外字フォント内では極細の空白文字として定義しているため、WWWブラウザに表示した際でも違和感はない。かつ、WWWブラウザからの貼付けにおいてはこのインジケータも同時に操作されるため、エディットボックスでは外字としての取り扱いが可能となる。

3.3 ユーザ認証およびアクセス制御方式の統一化

ネットで百科におけるクライアントとサーバ間の通信は、大きく(a)認証要求と応答、(b)検索要求と結果、(c)コンテンツ表示要求と要求されたコンテンツ、の3種類に分類できる。ここで、(b)(c)の処理の可否は、(a)で得られた認証情報に基づいて決定される必要がある。デュアル連想検索が当初から本機能を実現していたことから、これをシステム全体で共有するための方式を検討した。基本的には、テーマ年表検索に対する要求の可否を逐一デュアル連想検索サーバ側に照会し、その結果に基づいて判断を行えばよい。WWWサーバが提供するAPIを使用することで、テーマ年表検索サーバに対する全要求を実際の動作の直前にフックし、デュアル連想検索サーバに照会する機能(認証フィルタ)を実装した(図3)。この認証フィルタは、サーバ間をまたがる照会処理による性能の悪化を極小化するため、特定の種類のコンテンツに関してはローカルにキャッシングした認証情報を参照することでアクセスの可否を判断することもできる。

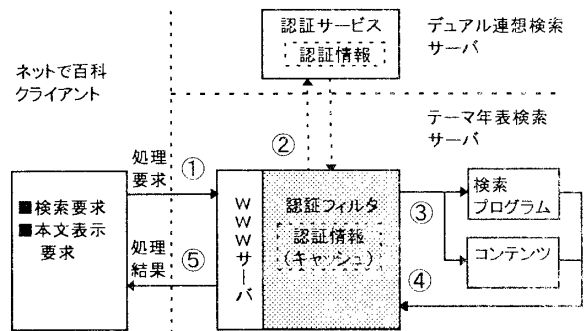


図3. 認証フィルタ

4. 結論

「ネットで百科」により、百科事典の次世代の検索インタフェースをインターネット経由で提供できる環境を実現した。また、参照に認証を要するようなコンテンツの提供を可能とする会員ネットワークの基盤を構築した。今後は「ネットで百科」をプラットフォームとして、他のコンテンツ所有者とのアライアンス等に取り組んでいきたい。

謝辞

サーバの設置、運用等で協力していただいた(株)日立国際ビジネスの三村、神庭両氏に感謝します。

参考文献

[HDH 98] CD-ROM 世界大百科事典 第2版, 日立デジタル平凡社, 1998.