

## WWW ナビゲーション環境の試作(1)

3P-1

## ～ディレクトリサービスの構成～

高野 元

久保 信也

島村 栄

NEC C&amp;C 研究所

## 1 はじめに

World-Wide Web(WWW) はインターネット上のハイパーメディアシステムとして広く利用されるに至っている。しかし、データモデルやナビゲーションI/F、アーキテクチャ等、ハイパーメディアシステムとして不十分な点も多い。

本稿では、ハイパーメディアシステムとしてのWWWの構成を再検討する。これによって提供されるWWW拡張機能をディレクトリサービスとよび、このアーキテクチャについて述べる。また、ディレクトリサービスの一機能として試作したキーワード検索機能についても述べる。

## 2 WWW と Dexter モデル

まず、ハイパーメディアシステムとしてのWWWの構成について検討する。ハイパーメディアの参照モデルである Dexter モデル [1] は、

**Runtime layer:** 表示やユーザインターフェースを規定

**Storage layer:** ノードリンクからなるネットワーク構造を格納するデータベース

**Within-component layer:** 各ノードのコンテンツおよびその構造を規定

からなるが、WWWに適用した場合にはWWWクライアントがRuntime layer、WWWサーバがWithin-component layerに相当するといえる。このことから、WWWをハイパーメディアシステムとしてみた場合にはStorage layerの機能が大きく欠けている。

したがってWWWの機能を拡張していくためには、WWWのハイパーメディア空間がネットワーク上に存在していることを考慮すれば、WWWブラウザと複数WWWサーバの間にStorage layerに相当するサーバが必要になる。

## 3 ディレクトリサービスの構成

これまでのハイパーメディア研究を考慮すると、WWWディレクトリサービスに取り込んで行くべき機能に対して下記のような検討が必要である。

- WWWクライアントによる、単にリンクをたどるナビゲーションだけではなく、
  - ノードリンク地図をグラフィカルに表示
  - フィルタリングやブラウジングI/FといったナビゲーションI/Fが必要である。また、これらのI/Fが使用するデータを提供するサーバが必要である。
- 単なるノードリンクモデルだけではなく、
  - インデックスを用いたフィルタリング等、情報検索技術
  - 集合操作リンク、仮想リンクといった高機能リンクを考慮したデータモデルが必要である。
- 上記の機能を実現するためには、リンク付けやインデキシングなどのオーサリングが不可欠である。WWWは、その構造が頻繁に変化しかつ大規模であるため、できるかぎりオーサリングが自動化されなければならない。

上記の検討にしたがって、WWWディレクトリサービスは次のように構成した(図1参照)。

**ナビゲーションI/F:** WWWドキュメントを表示するWWWクライアントと、これと連携するノードリンク地図やフィルタリング/ブラウジングツールから構成する。これらのツールはWWWクライアントから現在のURLを得て、ディレクトリサーバから関連情報を得る。

**ディレクトリサーバ:** Dexterモデルのstorage layerに相当し、ネットワーク上のノードリンク構造を総括的に管理する。また、キーワードインデックスや意味クラス等、WWWドキュメントの二次情報を管理し利用するデータモデルに基づいて、高度なナビゲーション機能を提供する。

また、こうしたデータモデルは柔軟な検索機能がベースとなっているため、その実現にはDBMSを利用する。

**WWWロボット:** 上記ディレクトリサーバが利用するWWWドキュメントの二次情報を自動的に収集する。探索対象を発見してWWWドキュメントを獲得するノードリンク探索ツールと、二次情報を生成するインデキシングツールから構成される。

"An Experimental Implementation of a WWW Navigation Environment - A Configuration of the Directory Service -," Hajime Takano, Nobuya Kubo, and Hisashi Shimamura, C&C Research Labs., NEC Corporation

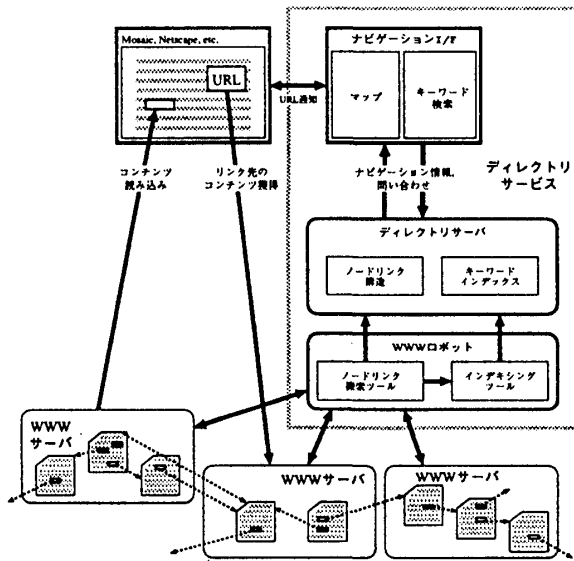


図1 ディレクトリサービスの構成

4 プロトタイプ

前節の議論に従い、二次情報としてキーワードインデックスを用いたナビゲーション機能を対象にディレクトリサービスを試作した。

4.1 キーワード検索 I/F

基本的なキーワード検索機能として、検索したいキーワードを入力するとその検索結果のURLとその抄録の一覧を提示する。また、本文/タイトル/抄録/URLのいずれかに含まれるもの、といった検索範囲指定も可能である。

さらに、現在のドキュメントに含まれるキーワードを提示することによって、キーワード入力の手間を省き、検索したいキーワードの指針を提示する [2] というナビゲーション I/F も採用している (図2参照)。

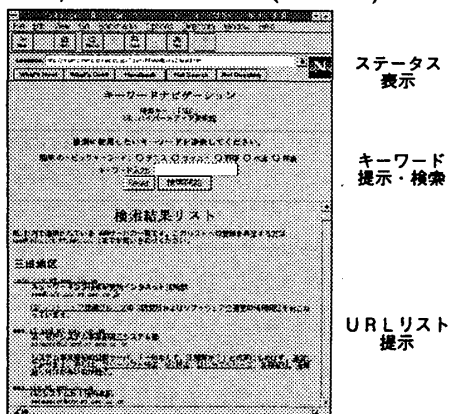


図2 キーワード検索 I/F

4.2 ディレクトリサーバ

URLとその中に含まれるキーワードインデックスを格納し、ユーザ(キーワード検索 I/F)からの検索要求に答える。

検索の柔軟性を重視し、検索エンジンとしてRDBMS

(Oracle Ver.7.1)上に構築しており、CGIプログラムを通じてキーワード検索 I/F と連携している。格納する属性値は、URL 情報として URL 名/タイトル/抄録などを、キーワード情報(インバーテッドファイルに相当)としてキーワード/URL/キーワードの重みなどを格納している。

これらの属性値はSQLを介して検索可能であり、検索範囲の指定や、URLごとの重要キーワードのリストアップ、といったCGIプログラムの作製の便宜を図っている。

4.3 WWW ロボット

まずノードリンク探索ツールが、指定されたドメインに存在するURLをそのリンク関係にしたがって探索し、そのコンテンツを獲得する(文献[3]参照)。

こうして獲得したドキュメントから、キーワードや抄録などのインデックスを抽出し、これをディレクトリサーバに格納する。

キーワード抽出の手順は、(1)形態素解析、(2)品詞や不要語辞書による単語の選別、(3)統計処理による単語の重み付けからなり、<TITLE><H1><H2>等のタグに囲まれたキーワードの重要度を高くしている。

5 おわりに

WWWの機能拡張として必要なディレクトリサービスの構成について検討した。また、これに基づいてキーワード検索機能を持ったディレクトリサービスを試作した。これによって、本稿で提案したディレクトリサービスの構成の妥当性を検証することができた。

既に The Internet 上には、Yahoo, Lycos, Infoseek などをはじめとして、数多くのキーワード検索機能からなるディレクトリサービスがある。これに対して我々は、WWWのハイパーメディア機能拡張としてディレクトリサービスを位置付けており、キーワード検索機能はその第一ステップである。プロトタイプでは既に、ノードリンク探索ツールの探索情報としてノードリンク情報を格納しているため、これに基づいてノードリンク構造を視覚化する予定である。また、さまざまなハイパーメディアデータモデルの適用によってナビゲーション機能を順次高機能化していくことを計画している。

参考文献

[1] Frank Halasz and Mayer Schwartz. The Dexter Hypertext Reference Model. *Communication of the ACM*, Vol. 37, No. 2, pp. 30-39, February 1994.  
 [2] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In *Proceedings of SIGIR'92*, pp. 318-329, June 1992.  
 [3] 久保信也, 高野元. WWW ナビゲーション環境の試作(2)~ノード・リンク探索システムの実装と評価~. 情報処理学会第52回全国大会. 情報処理学会, 1996.