

梅村 武久*1*2、安田 孝美*1*3

通信・放送機構*1 NTT西日本 マルチメディア技術開発センタ*2 名古屋大学 情報文化学部*3

1. はじめに

学習指導要領の改訂により高等学校のみならず、初等中等学校においても「情報」に関する教育を重視し、コンピュータ等の情報手段の活用を一層推進することが挙げられている[1]。従って、今後、教育シーンでのインターネット利用は必要不可欠なものとなることが予想される。しかし、肥大化傾向にあるインターネットワールド(以下、WWW)で限られた授業時間内に欲しい情報の発見や入手が困難であるといった問題がある。この問題を解決するために、多くのWWW検索サーバが、インタラクティブな検索や分類、関連タームの提示など様々なナビゲーション方式を用いている。しかし、情報の増大により検索サーバが巨大化しリソースや検索ナビゲートの限界が予想されるため、何らかの基準で特化した専門分野別の検索エンジンを知的に使い分ける必要性があることも提言されている[2]。

2. 研究の目的

本研究では上記の背景から、利用者の環境を教育分野に絞り効率的な検索を可能とするWWW検索システムの実現を目指す。具体的には、まず初等中等教育分野を対象とし、検索結果をカテゴリ分類体系に整理して表示することにより、次のアクションを決定するための情報を利用者に提供する実験用のWWW検索システムを構築する。そして、構築した実験システムを実際の教育現場で運用し利用者の検索過程をもとに、(1)教育分野における最適なWWW情報のカテゴリ分類体系を確立し、かつ(2)ユーザフレンドリな検索ナビゲーションへの改良を行うことにより、利用者の検索の効率化を図る。

また、起点からのホスト間距離、情報収集タイミング等の最適値を算出して、(3)検索ロボットを最適に制御す

The Classification/Systematization of WWW Information for Education

Takehisa Umemura(NTTWEST Multimedia Technology & Development Center), Takami Yasuda(Nagoya University)

る方法を検討し、ユーザインタフェースの面からのみでなく最新版のコンテンツや内容のあるコンテンツが検索の対象となるようにすることで検索の効率化を図る。本稿では、実際に構築した実験用のWWW検索システムの紹介をすると共に、上記目的(1)の教育分野におけるカテゴリ分類体系構築の初期実験結果を報告する。

3. WWWサーバ検索システム概要

本システムは、通信・放送機構岡崎公共システム開発リサーチセンタの「ネットワークアーキテクチャに関する研究開発」の一環として構築しているものであり、その概念図を図1に示す。

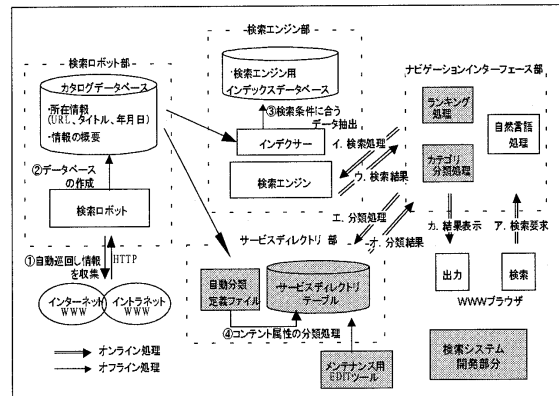


図1 WWW 検索サーバシステム概念図

このシステムは、Sun ワークステーション上のアプリケーションサーバとして構築され、NTTソフトウェア(株)のInfoHitter をベースにカスタマイズを施したものである。主な機能を以下に挙げる。

a. 検索ロボット部

与えられた起点URLからハイパーリンクを順に辿りながら、文書情報を自動収集。収集した単語情報や、情報の所在(URL)等をカタログデータベースに保存する。

b. 検索エンジン部

検索の高速化のため、カタログデータベースの内容を検索エンジン(WAIS)に適した形式に変換し、WAIS を用いて利用者からの要求に応じ検索処理を行う。

c. サービスディレクトリ部

カタログデータベースに含まれる情報を、自動分類定義ファイルに定義されたキーワードを基にカテゴリライズし、分類結果をサービスディレクトリテーブルに出力する。

d. ナビゲーションインタフェース部

利用者の検索要求に適合する情報を検索エンジンより取得すると共に、サービスディレクトリテーブルの分類情報を統合し、検索結果を分類ノード毎に整理して利用者に分かりやすく表示する。

4. 教育分野向けカテゴリ分類体系

4.1 カテゴリ分類体系の作成

本研究では、上述した機能cと機能dを教育分野向けに改良することになる。まず機能cのサービスディレクトリ部を構築するために、既存のWWW検索サービスや指導要領等の教育関連情報、各種書籍を調査し[3][4]、表1に示すカテゴリ分類体系を作成する。更に、検索ロボットに設定する148の初期URLの選出を行う。

表1 カテゴリ分類体系

	大項目	中項目	小項目	種小項目
1	/学校の授業			
2	/学校の授業	小学校		
3	/学校の授業	小学校	国語	
4	/学校の授業	小学校	算数	
5	/学校の授業	小学校	理科	
6	/学校の授業	小学校	社会	
7	/学校の授業	小学校	図工	
153	/役立つ情報	動物・植物		
154	/その他			
計	11	64	76	3

4.2 分類キーワードの抽出

サービスディレクトリ部の自動分類定義ファイルに定義するキーワードを以下の方法により抽出する。

- ①初期URL(148サイト)にて検索ロボットを作動
- ②収集ページ内容を確認後、表1のカテゴリに分類
- ③各カテゴリに属するページの単語(形態素解析により名詞句化)を頻度順に整理
- ④頻度の高い単語から、一般的なものを除き特徴的と思われる単語をキーワードとして抽出

表2に収集した総ページ数と抽出したキーワード数を、図2に自動分類定義ファイルの一部を示す。

抽出したキーワードを自動分類定義ファイルに用いた分類結果としては、“その他”が19ページ、“未分類”が1,653ページ、分類されたページの無いカテゴリは全カテゴリ154中22あり、収集ページに対する分類率

(分類済ページ/収集総ページ)は約90%であった。一方、キーワードによる分類では、類似項目の判定の困難性(小学校/国語と中学校/国語の区別等)や分類が困難なカテゴリ(図鑑や素材集等)の存在などの問題があることがわかった。

表2 収集ページ数と抽出キーワード数

対象サイト数	収集ページ総数	抽出キーワード数
148	17,292	3,917

/学校の授業:	
/学校の授業/小学校:	
/学校の授業/小学校/国語:絵文字,民話,語源,読み方,辞書,ことば,字,片仮名,昔話,語...	
/学校の授業/小学校/算数:算数,たし算,ひき算,算用,角度,ものさし,定規,数字,数,け...	
/学校の授業/小学校/理科:ふしぎ,空気,太陽,花粉,摩擦,月,葉っぱ,力,雲,ムシ,葉緑素...	
/学校の授業/小学校/社会:土器,縄文,曜日,地球儀,古代人,堅穴,恐竜,化石,弥生,ダム,...	
/学校の授業/小学校/図工:はさみ,のり,カッター,クレヨン,ポンド,ホッチキス,セロテ...	
↓	
/役立つ情報/動物・植物:絶滅,希少,高山,草花,水生,葉緑素,知的,集団,動物,脊椎動物...	
/その他:乗り物,ビル,プロペラ,おうち,核的,キャベツ,容積,滝,バルシャ,物理学,自由...	

図2 自動分類定義ファイル

5. まとめと今後の展望

本初期実験では、教育向けカテゴリ分類体系の作成と分類のためのキーワードの抽出を行い、その分類状況や問題点を把握した。今後は、未分類ページの内容分析や分類の正確性を調査すると共に、本システムを用いた教育現場にて利用者の要望や検索ログデータをフィードバックしカテゴリ項目を修正、更に検索対象URLを増加させキーワードの再整理を行い、実運用を目指した教育分野向けカテゴリ分類体系を確立していく。また、キーワード分類による問題も考慮し、URL名による分類やハイパーリンクの共起による分類[5]なども検討する必要がある。

参考文献

- [1]文部省:学習指導要領,
<http://www.monbu.go.jp/news/00000317/index.html>.
- [2]林良彦,小橋喜嗣:WWW上の検索サービスの技術動向,
情報処理学会論文誌,Vol.39 No.9 Sep, pp861~865, (1998).
- [3]越桐國雄:インターネット教育エッセイ,旬報社,(1999.3).
- [4]山極隆,無藤隆:新しい教育課程と21世紀の学校,ぎょうせい,
(1998.12).
- [5]大久保雅且,杉崎正之,田中一男:リンクの共起関係を用いた Web
ページ分類法式の検討,第59回情処全大,pp3-81~82,(1999).