

**個人用Webページ管理機構における索引の自動生成機構の設計と実現**

4 T-9

河部 絢史 並木 美太郎

(東京農工大学 工学部)

**1. はじめに**

近年、WWWは急激に普及し、WWW上で様々な情報が公開、掲載されるようになった。そこで多くの人々がWWW上で必要な情報を採すようになった。そのためWWW上のドキュメントを個人用に保存し利用するための個人用の管理機構が必要になってきている。

保存されたWebページを検索するにあたって、その検索には様々な手段が考えられサーチエンジンやディレクトリごとに分野分けして保存するなどの方法が実現されているが、Webページが同一のもので、検索方法が同じならば、当然のことながら検索結果としてユーザに提供される検索結果は同じものが提供される。しかし、その検索結果として表示されたWebページの一覧の中には、ユーザが参照した際に不要と感じるWebページも含まれることがある。しかし、Webページの内容を自ら変更するのはユーザにとって容易な作業ではない。そこで各Webページに対応する索引を生成し、その索引を調査して検索結果を表示する。直接Webページを検索してその結果を表示するのに対して、索引を変更するだけで検索結果を変えることができ、個人用にカスタマイズされた検索結果を常に提供できる。

本報告では個人用のWebページ管理機構における索引の自動生成機構の設計と実現について述べる。

**2. 索引の自動生成機構の目的**

## (1) 索引からWebページを参照

ハードディスクに保存されたWebページに対して索引を自動で生成、保存し、それらの一覧

をユーザに提供する。ユーザはその一覧から見たいWebページを選択できる

## (2) 索引をキーワード検索で自動生成する

Webページの一つ一つに索引をユーザが作っていくのは非常に大きな手間のかかる作業である。そこで索引を自動で生成する。

## (3) 索引の補足、変更を可能にする

索引による検索には参照する都度の検索に対して、新たな情報を次々に組み込む、もしくは変更することで後の参照が容易になるという利点がある。そこで最初に自動で作られた索引ファイルに対して、随時、情報を書き加えていく。

## (4) 複数の項目による絞込みを可能にする

索引の複数の項目を提供し、それらの項目を利用した、andもしくはor検索を行い、検索結果を絞り込めるようにする。

## (5) 索引項目の拡張を可能にする

索引の項目をユーザが後に増やすことを可能にする。これにより、後からユーザが異なる手段で索引を生成した場合に対応できるようにする。

**3. Webページ管理機構の全体構成**

個人用Webページ管理機構の全体構成を図1に示す。次の各機能ごとにWebページ管理機構は複数の部分に分かれている。

## (1) WWW上のドキュメントの検索、保存を行う

## (2) 保存したWebページに索引を生成する

## (3) 索引を操作し、Webページを提供する

## (4) 各部分のデータのやり取りなどを行う

索引の自動生成機構とは、このうち(2)ならびに(3)を指す。

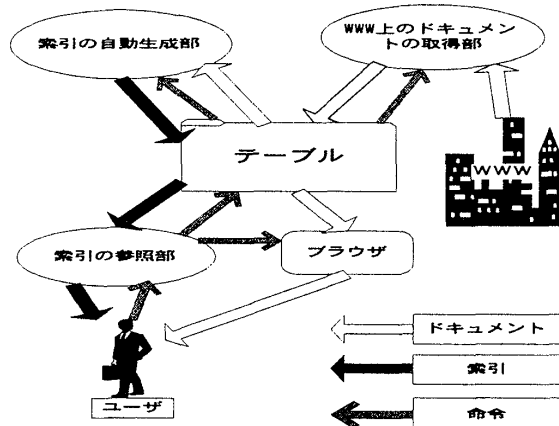


図 1 Web ページ管理機構の全体構成

## 4. 設計

### 4.1 索引の表記方法

索引は複数の項目を取りこむにあたって参照の際の検索手段を統一しないと、索引の項目が増やせても、それらを検索するのに新たな検索システムを作らなくてはならなくなる。そこで各索引の項目の表記方法として、索引の中にどうしても必要なもの以外は、数値によって記録し、その数値の上下によって判断可能なものにかぎる。

### 4.2 索引の構成と項目

索引の内部に書かれる項目は次の通りである。なお、これらとは別に、Web ページのタイトル、URL、パスを記録している。

#### (1) 一次索引

一次索引は保存された Web ページに対して最初に自動的に生成される索引である。

##### ① 取得日時

Web ページを WWW 上から取得した日時を記録する。

##### ② キーワード検索の結果

ユーザはあらかじめ、自分の興味のある分野ごとをカテゴリとして複数設定しておき、そのカテゴリの中に出現すると予想されるキーワードを登録しておく。それをもとにキーワード検索を行い各カテゴリごとの得点を算出する。その

得点を、各カテゴリを一つの項目として保存する。

#### ③ 特定ドキュメント量

Web ページから画像や音声などのドキュメントにリンクを貼っているものは多く、そのようなものを求めるユーザも多い。そこでリンクを調べ、ユーザが指定した、

#### (2) 二次索引

二次索引は一度ユーザが Web ページを参照した後で、索引ファイルに組み込まれる項目のまとまりである。

##### ① 最後に参照した日時

最後に参照した日時を索引に加えることにより、例えば「昨日見た Web ページをもう一度みたい」と言った場合の参照が容易になる。

##### ② 参照した際の感想

ユーザが Web ページを参照した後で、その Web ページ内の情報の有用度を決め、それを索引の中に組み込んでおく。

## 5. 終わりに

本報告では個人用 Web ページ管理機構における索引の自動生成機構の設計と実現について述べた。

本システムを Microsoft Windows 98 上で実現したことにより、索引を介することで個人用にカスタマイズ可能な検索を可能にした。

今後の課題として、Web ページ管理機構の他の WWW 上からのドキュメントの取得など機能を行う部分の設計と実現を行い、システム全体の完成を目指す。また、索引項目をユーザが増やしやすいするためのクラスライブラリの提供もおこなう。World wide web

## 参考文献

土橋喜：ハイパーメディアによるデータベースディレクトリと分類体系のシステム化, Journal of College and University Libraries, No. 45, pp. 8-15 1994.