

検索先の自動選択を行う WWW ベース横断検索システム

6 Z - 3

柳本 豪一 谷 幹也 市山 俊治

NEC ヒューマンメディア研究所

1 はじめに

インターネット上でデータベース検索サービスを公開している検索サーバは数多く存在している。そのような検索サーバの一例として、図書館のオンライン図書目録(OPAC:Online Public Access Catalog)がある。OPACサーバを利用して図書の検索を行う場合、検索対象として利用できるサーバの数が多すぎ、全ての検索対象に対して検索要求を出すネットワーク負荷などの検索コストがかかりすぎてしまう。そこで検索を行う前に絞り込む必要があるが、従来は利用者自身が検索先を指定することにより絞り込みを行っていたため、利用者の負担が大きかった[1],[2]。このため、自動的に利用者の検索要求に即したOPACサーバを絞り込む仕組みを提供すれば、利用者の負担を軽減することができる。

本稿では利用者が入力した検索語をもとに検索先の絞り込みを行うWWWベース横断検索システムについて述べる。

複数の検索サーバを検索すると、得られる検索結果はサーバ間でばらつきがあり、更に入力する検索語によってそのばらつきが変化する。このことから、このばらつきは単に蔵書の多少によるのではなく、蔵書の分野ごとに偏りがあるために生じることが分る。この偏りが、例えば大学であれば学部学科構成や教員の専門分野を反映して、分野間で生じると仮定すると、OPACサーバ毎に得意な分野、不得意な分野を定めることができる。更に検索語がどの分野に属するかが分れば、その検索語に対してより

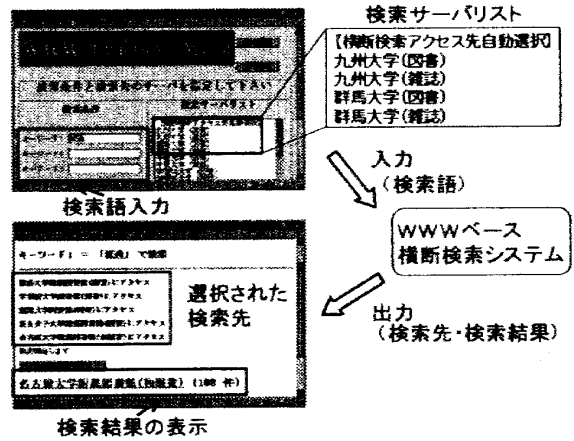


図1: 検索処理の手順

多くの検索結果が得られるOPACサーバを推定できる。本システムは上記の仮定に基づき検索先選択機能を構成したものである。

本システムは、入力された単語が属する分野を推定する知識と、分野ごとにどのOPACサーバが適しているかを評価した知識の2種類の知識を持っており、入力された検索語から分野を推定し、分野をもとに検索先の絞り込みを行うことと、利用者が行った一回の検索処理ごとに検索先の評価の修正を行うことを特徴としている。

2 検索処理

図1に検索の流れを示す。WWWベース横断検索システムはWWWブラウザから利用できる。検索先の自動選択を行う時には検索語を入力し、検索サーバリスト内の「横断検索アクセス先自動選択」を選択して検索を行う。システムは入力された検索語をもとに登録されているOPACサーバの中から絞り込みを行い検索先を決定する。選択されたOPACサーバは利用者に表示されるとともに、各OPAC

サーバに対して検索要求を出す。このときシステムでは選択された OPAC サーバごとに検索式を作成し同時並列に検索を行う。検索が終了した OPAC サーバについてはサーバ名と検索結果の件数を表示する。

3 システム構成

本システムの構成を図 2 に示す。本システムでは、検索先を絞り込む際に利用する知識として分野推定データベースと検索先評価データベースの 2 種類の知識を持っている。

・分野推定データベース

分野推定データベースは利用者が入力した検索語からその単語が属する分野を決定するための知識である。シソーラスは分野間の上位/下位関係を表現するために用いており、各分野にはその分野に属している単語が対応付けられている。例えば、キーワードとして「流通」が入力された場合、分野として「経済」を返すような処理を行う。

・検索先評価データベース

検索先評価データベースは分野推定データベースで得られた分野を用いて、検索先の絞り込みを行う知識である。OPAC サーバの評価は、それぞれ分野ごとに含まれる代表的ないくつかの単語を用いて検索を行い、サーバから得られた検索結果の件数をもとに決定されている。ある分野における評価値を修正するために以下のような式を用いる。

$$W_C^i = \alpha W^i + f_i(R_C^1, \dots, R_C^i, \dots, R_C^N) \quad (1 > \alpha \geq 0)$$

W_C^i : 検索語 C の結果で修正を行った検索サーバ i の評価値

W^i : 検索語 C で選択を行う際に用いた検索サーバ i の評価値

$f_i()$: 結果の件数より決まる検索サーバ i の評価に対する修正量

R_C^N : OPAC サーバ N から得られた検索結果の件数

もし利用者が入力したキーワードが分野推定データベースによって分野の推定が出来ない場合、デフォルトとして指定されている検索サーバ群を検索先として利用する。

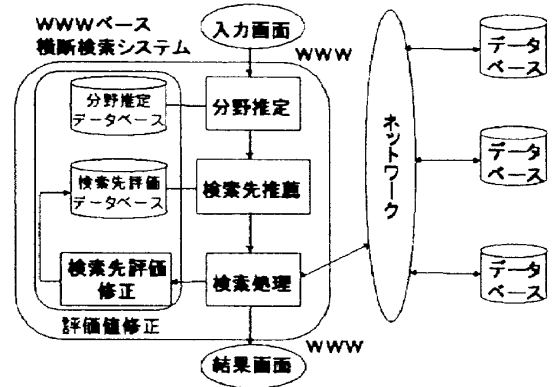


図 2: システム構成

検索先評価データベースの知識更新には OPAC サーバの情報が必要であるが、検索結果の件数のみを利用しているため、利用者が行った検索処理の結果をそのまま利用することができる。また、検索語ごとではなく分野ごとに評価されているので、利用者が今まで検索を行ったことのない検索語を入力した場合にも、分野推定データベースに登録されている他の検索語での検索結果を利用して絞り込みが行える上、評価値の修正の際に分野ごとのサンプル数の増加が期待できる。

4 おわりに

検索サーバに関する知識を用いて利用者の検索要求に応じた検索サーバを推定する機能と毎回の検索結果の件数を用いて、検索サーバに関する知識を更新する機能を実現した。それにより、WWWベース横断検索システムにおける検索先の自動選択機能を実現し、利用者の負担を軽減することができた。

参考文献

- [1] 市山 他 多様な情報源を対象とする WWW ベース電子図書館システム、「デジタル図書館」ワークショップ 第 7 回、pp.32-49、1996
- [2] 柳本 他 WWW ベース図書館情報横断検索システム、情報処理学会全国大会第 54 回、1997