

WWWベース図書情報横断検索システム

6R-6

柳本 豪一 柳生 弘之* 山門 純** 細見 格

NEC 関西C&C研究所 * NEC情報システムズ ** NECソフトウェア関西

1 はじめに

現在、ネットワーク上で多くの検索サーバが公開されてサービスの提供を行なっている。例えば、図書館の図書目録(OPAC: Online Public Access Catalog)の検索が行なえる図書検索サーバなどである。しかし、各サーバのプロトコルや操作インターフェースが異なっており、利用者が検索を行なうためにはそれぞれの操作方法を習得しなくてはならず余分な負担をかけていた。また、得られた検索結果も各サーバ毎に異なった形式で表示されるため、サーバをまたがって検索結果を容易に比較することが困難であった。このため、効率的に各サーバの検索を行ない、結果を比較できる技術が必要であった。

本稿では統一したインターフェースを持ち、複数のOPACサーバに一括して検索を行なうWWWベースの横断検索システムについて述べる。各OPACサーバ毎に適応知識を用意することにより、インターフェースの違いを吸収する。また、検索処理は複数のOPACサーバに対し、同時並列に行なう。検索結果の表示は各サーバの形式に依存しない統一した形式に変換して利用者に提示する。また利用者が結果の比較を行ないやすくするため、得られた検索結果間で重複したデータの削除などを行なう統合表示機能を持つ。

2 検索処理

本システムの検索画面を図1に示す。利用者が横断検索システムを利用する際、WWWブラウザ上でキーワードを入力し、検索を行なうOPACサーバを選択する。検索のためのキーワードとしては、タイトルや著者名などの書誌情報を利用することができる。また、キーワードは複数指定することができ、指定された場合各キーワ

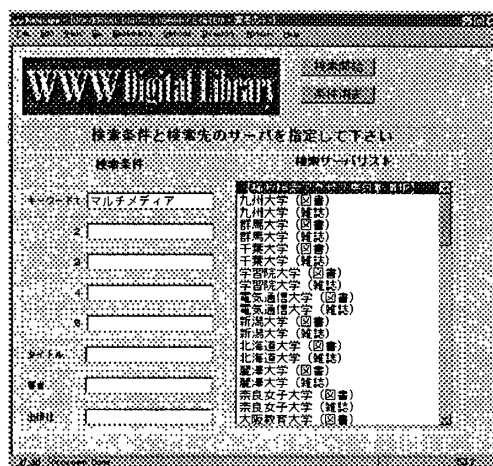


図1: 横断検索システムの検索画面

ドのAND結合を条件として検索を行なう。OPACサーバについては一度に複数選択することが可能である。OPACサーバを選択する時、各サーバのプロトコルやインターフェースの違いを利用者は一切考慮する必要がない。

2.1 各サーバへの対応

検索手順などのサーバごとの違いを吸収するため、各サーバごとに検索スクリプトと呼ぶ検索手順に関する知識ファイルを用意している。この検索スクリプトには、各サーバに対応した検索式のテンプレートが用意されており、利用者からの検索要求に応じて各サーバに送る検索式を作成する。他にも運営時間や漢字コードなどデータベース知識も有しており、利用者が要求した条件と各サーバの運営状態のチェックを行なうことができる。

2.2 並列検索

指定されたOPACサーバが複数の場合、順番に検索を行なうのではなく同時並列に検索を行なう。利用者に対し検索が終了したOPACサーバから早いもの順でサーバ名の表示を行なうことにより、待ち時間を短縮する。早く検索が終了したサーバについては、利用者が検索結

WWW Library System with Cross Domain Search

Hidekazu YANAGIMOTO, Hiroyuki YAGYUU*

Jun YAMAKADO**, Itaru HOSOMI

Kansai C&C Research Laboratories, NEC Corporation

*NEC Information Systems

**NEC Software Kansai, Ltd.

果を参照することができる。また正常に検索が終了しなかった場合、利用者に検索が失敗した原因を調査し通知する。

3 結果表示

3.1 標準構造化

検索により得られた結果は各サーバごとに異なった形式を持っている。そのまま利用者に結果として表示したのでは、他のサーバから得られた検索結果と比較する際に見にくくなってしまいます。そこで、図書の識別に有効で、かつどのサーバにも含まれている項目を切り出し一定の形式で利用者に結果を表示する。このデータ形式を標準構造化と呼ぶことにする。標準構造化を用いることにより、各サーバの検索結果を相互に比較しやすくなる。標準構造化としては以下のものを採用している。

[タイトル, 著者名 or 巻号, 出版者]

図書の場合は著者名を雑誌の場合は巻号を取る。図書と雑誌で異なっているのは、目的のキーワードを含む記事が載っている雑誌の巻号まで明示する必要があるためである。実際の検索結果の画面例を図2に示す。各サー

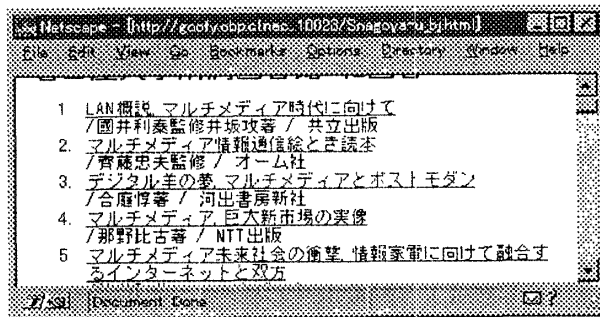


図2: 検索結果の表示画面

バから得られた検索結果から標準構造化として採用した項目を切り出す知識を定義し、これを標準構造抽出フィルタと呼んでいる。標準構造抽出フィルタは各サーバごとに用意しており、検索結果の項目の並びが記述されている。この標準構造抽出フィルタを用いることにより標準構造化として必要な項目の切り出しを行なう。

3.2 統合表示

複数のOPACサーバを対象として検索を行なうと、どうしても各サーバ間で同一の図書が含まれてしまう。

このため、利用者が各検索結果を見比べる際には、このような重複を取り除く必要がある。判断基準として標準構造化として切り出された項目のうち、タイトル以外にも一つの項目が一致する場合同一の図書であると判断する。これはOPACサーバから得られる検索結果の表示形式が異なるため、すべての項目が完全に一致することを要求すると、同一の図書を違う図書と扱ってしまうためである。よって冗長性を判断するためのルールを緩めてやることにより対処した。

4 システム構成

本システム構成を図3に示す。本システムは国内40ヶ所、海外10ヶ所を対象に検索が行なえる。対応しているプロトコルはhttpとtelnetである。また、新規サーバ登録用に知識作成ツールも用意している。

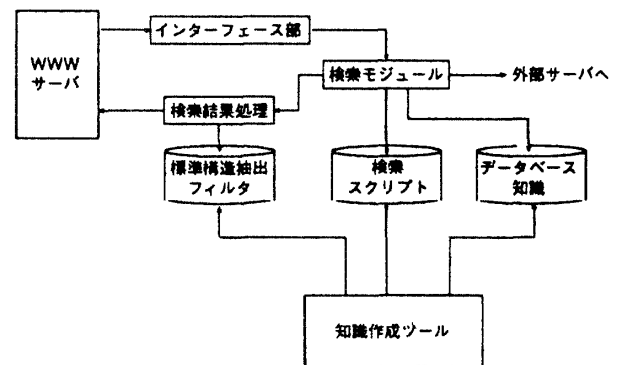


図3: システム構成

5 おわりに

本システムによって、複数のOPACサーバの違いを利用者が意識せずに検索を行なえる。また、複数のOPACサーバから得られた検索結果を比較しやすい表示形式を提供した。

今後は、既存のOPACサーバだけでなく他の検索サービスや一次情報を含むサーバに対応する。また、統合表示の表示規則も利用者の要求に応じて切り替えて行くような機能を提供する。さらに、利用者の検索要求に応じて最適な検索先を自動的に選択する機能も提供する。

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