

## 全情報源に対する適応型検索システムの提案および 蔵書構成に基づく検索サービスの実現

近藤 育雄 飯田 記子 上島 紳一

関西大学総合情報学部

### 1. まえがき

ネットワーク上に存在する多種多量の情報を効果的、合理的に処理するためには、利用可能な多くの情報の中から必要なもののみを適切に取り出すことのできる仕組みが必要である。本稿では、ネットワーク上で利用可能なすべての情報源に対して、OSやDBMSなどの実装方式に制約されず、質問から回答まで継ぎ目なしの検索を可能にする適応型検索システムを提案し、その要件について述べる。また、そのサービス機能の一部となる、図書館における蔵書構成の考え方を導入した検索サービスの試験的実装についても述べる。

### 2. 検索システムの概要

#### 2.1 適応型検索システムの要件

適応型検索システムにおける検索手法は、すべての情報源を対象として、現実的な費用で検索が完了し、かつ将来の検索機能の高度化に対応できる必要がある。ここでは、この条件を満たす検索手法を全部検索と呼び、この手法が満たすべき3つの要件について述べる。

##### 要件1：情報源のデータ構造に制約されない

全部検索では、情報源のデータ構造の如何によらず、検索が可能でなくてはならない。ここでいうデータ構造とは、MIMEなどで表現されるファイルタイプのほか、データベースにおける各デー

タの並び方などを含んでいる。この要件の目的は、検索対象となる情報源のエンコード形式を透過的に扱うことである。データ構造の種別ごとに検索システムを使い分ける煩雑さが解消し、様々なデータ構造を持つ情報源を対象にした検索結果を、網羅的に得ることができる。

##### 要件2：情報源の物理的属性に制約されない

全部検索では、利用者に対して、情報源の位置や重複、障害、更新を仮想化し、複数のサーバによる透過的なサービスの提供が行われなければならない。この要件により、実際には複数のサーバから構成されるサービスを、仮想的にひとつのサービスであるかのように利用できる。システムは、蔵書構成などの情報に基づき、検索結果の処理に必要な情報源やサービスの位置およびそれらへの接続手順などを仮想化して提供する。

##### 要件3：検索対象の選択ができる

全部検索において、実際に検索操作が行われる情報源は、情報源に関する属性に基づき選択される。ここで情報源に関する属性とは、情報源の状態や利用条件を示した情報を指し、蔵書構成やPICSラベル、robots.txtなどメタデータや、OSや各サーバ・アプリケーションが提供するファイル属性などが含まれる。検索システムは、これらの情報に基づいて、検索する情報源を限定したり、他の独立した検索サービスに質問を依頼する。

#### 2.2 配布/蓄積/表現の分離原則

適応型検索システムの機能は、下位層から順に配布、蓄積、表現の三層より階層的に構成される(図1)。この階層化により、各層の機能を互いに独立させることができ、機能の拡張や変更の影響を同一階層内にとどめることができる。各層を提供するソフトウェアは、新規に作成されたもので

---

Adaptive retrieval system for every data  
source all over the internet and trial  
implementation retrieval service based on the  
collection development of library  
Ikuo Kondo, Noriko Iida, Shinichi Ueshima  
Faculty of Informatics, Kansai University

も、既存のものを修正したものでもよい。

### 2.3 蔵書構成

図書館の通常業務として、図書館資料の収集がある。この業務は、ある一定の機能を果たすために必要となる各主題別の蔵書量を形成することを目標に行われる。ある時点での各主題概念に属する蔵書の量を蔵書構成という[1]。この考えに基づき、蔵書構成リストというある専門性を持った情報源のリストを定義し、検索システム間の選択協調や情報源の維持管理に利用する。

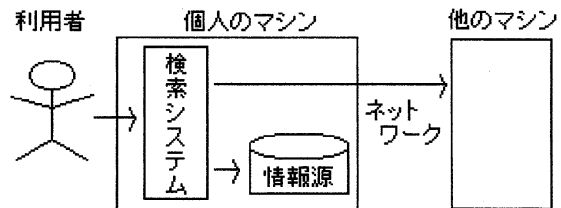
各蔵書構成リストは、リスト作成方針をもつリスト作成者により作成され、そのリストには、一定の機能を果たすために必要な情報源が記されている。ユーザは任意の時点で、任意の内容のリストを作成することができる。

図書館が図書館間協力によって利用者の要求に応えるように、適応型検索システムでは、各サーバ間での蔵書構成リストの交換による資源共有により、システム全体として資源の効率的な利用を図るために協力する。

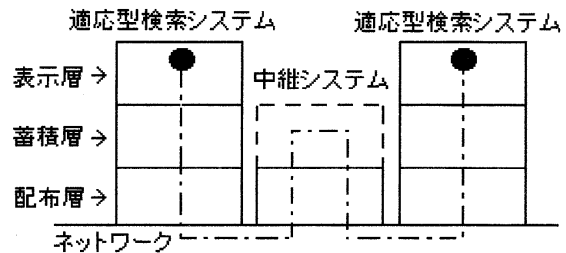
蔵書構成リストによるサーバ間協力の方法は、他の検索システムにも適用できる。例えばWWWサーバ上のデータを高速に収集する目的で提案されている分散型WWWロボット[2]への応用である。同システムでは、分散配置されたロボットが、それぞれ、ping最小応答値から予測されるネットワーク的に近いドメインの収集を担当し、全体としてWWWデータを高速かつ重複なく収集する。これに対して、収集の分担方法に蔵書構成リストを利用し、自ドメイン内においては網羅的収集を行い、加えて蔵書構成リストに従った収集を行うという拡張を考えることができるであろう。

### 3. 試験的実装

これまでの検討を踏まえた試験的実装として、蔵書構成の考え方を導入した検索システムを計画した。これは、各利用者がそれぞれ個人的に作成した蔵書構成リストを持つWeb-DB連携アプリケーションとして作成され、ウェブディレクトリサ



(a) 質問の流れ



(b) 階層化のモデル

図1. 適応型検索システムにおける機能の階層化

ービスとして利用できるものである。利用者間で蔵書構成リストを交換し合うことにより、データベースの内容を互いに補完することができる。

蔵書構成リストは、そのリストに関する情報を記したヘッダ部と、個々のURLアイテムを列挙した部分からなる。ヘッダ部にはリストの表題、作者、要約、最終更新日などの情報が含まれる。各アイテムには、URL、表題、ファイルタイプ、作者、要約などの情報が含まれる。

蔵書構成リストの交換方法には、ftp等でファイルをやり取りする方法のほか、交換用のサーバを介して行う方法が考えられる。今回の試験的実装では、作業の簡便さから前者の方法を採用した。

今後、試験的実装を完成させた後、情報配信を他のシステムに委託する機能を加え、発展させてゆく予定である。

### 参考文献

1. 河合弘志編：蔵書構成と図書選択 新版，日本図書協会，(1992)。
2. 山名ほか：分散型WWWロボットによるWWW情報収集，<http://www.etl.go.jp/~yamana/Publications/ABST/DEWS98/24.html>