

クライアント-サーバモデルによる情報検索システムの提案

3R-6

西澤 格
東京大学工学部安達 淳
学術情報センター研究開発部

1 はじめに

分散処理の進展と共に、広域ネットワーク上で分散処理との親和性の高い情報検索システムが求められるようになった。この場合、必要な情報は1つのデータベースに集中して蓄えられているのではなく、複数のデータベースに分かれて蓄積されていると考えるほうがより自然であろう。また複数のデータベース上でデータの構造・意味がすべて同じだと考え難い。

本稿ではこれらの複数の情報源を統一的に扱うことが可能な情報検索システムの構成方法についての提案を行う。

2 提案するシステムの特徴

従来の情報検索システムはそのほとんどが情報の蓄えられているデータベースにログインし、その上に実装されているデータ操作言語や組み込みコマンドを用いて検索を行うという、いわゆる TSS ベースのものである。このような情報検索システムは集中管理型のデータベースを前提としており、必要な情報が複数のデータベース上に分かれて蓄積されている場合には、以下のような問題が生じる。

- 1.各システムにデータにアクセスする際のコマンド体系が異なるため、ユーザは異なるデータベースにアクセスするたびに、そのシステム固有のコマンドを使用しなければならない。
- 2.各データベースによって、蓄えられているデータの構造および意味が異なるため、ユーザはそれぞれの意味・構造を把握しづらい。
- 3.2の問題があるにもかかわらず、SQLなどのデータ操作言語ではデータ構造を明示的に記述しなければならないため、ユーザは問い合わせを書くにあたり、そのデータベースのデータ構造(スキーマ記述)を十分に理解しておく必要がある。

これらの問題を考慮し、本システムでは分散データベース向けの情報検索のモデルとしてクライアント-

サーバモデルを基礎とし、情報検索過程を“クライアントであるユーザがサーバである各データベースと通信を行うこと”によって実現しようとする。筆者は本モデルにしたがって前述の問題を解消するようなシステムの構築を進めている。

3 本システムの構成

本システムの全体像を図1に示す。クライアント-サーバモデルという立場から本システムの情報検索過程の概略を述べると、“ユーザインタフェースを介してユーザからの入力を受けとったクライアントは、ネットワーク上に散在する情報資源であるリソースオブジェクト(以下RO)からサーバである各データベースにアクセスする際に必要な各種の情報を得ながらサーバとセッションを行うことにより情報検索を行う”ということになる。

情報検索を行うにあたり、本システムは従来の手法と比較して以下のような利点を持たせている。

- 1.データの総称化機能 複数のデータベースにおけるデータの意味と呼称の違いをユーザができるだけ気にすることなく情報検索を行うことができること。
- 2.情報源の位置透明性の実現 ユーザができるだけ情報の所在場所を認識する必要なく情報検索を行うことができること。
- 3.利用者インタフェースのユーザ支援機能 ユーザができるだけ具体的なデータ操作言語を知る必要なしに情報検索を行うことができること。

上述の利点を実現するための2つの柱は4節で説明するROおよびマネージャ(図1参照)の働きである。これらの働きは直接ユーザからは見えない。

4 検索のシナリオとシステムの挙動

本節では情報検索を行う際のシナリオとそれにもなうシステムの挙動について説明する(図1参照)。

4.1 データベース選択と問い合わせ生成

ここでは情報検索の過程を2つのフェーズに分け、まずデータベースの選択とそのデータベース用の問い合わせの生成までの過程について説明する。

A Proposal for Client-Server Model in Information Retrieval System

Itaru NISHIZAWA¹, Jun ADACHI²

¹Faculty of Engineering, The University of Tokyo

²Research & Development Department, National Center for Science Information Systems

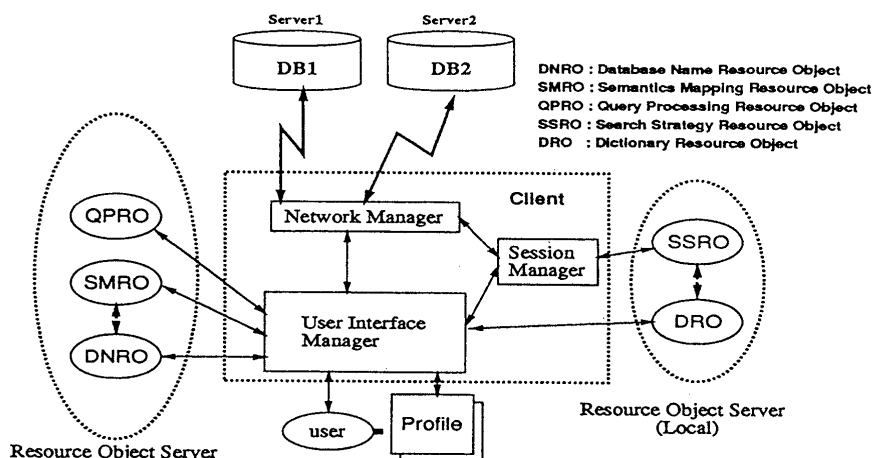


図 1: システムの全体像

1. まず、ユーザはクライアント側検索システムを起動する。
2. ユーザが検索を行ないたい情報を選択するとユーザインタフェースマネージャ(以下 UIM) は DNRO とセッションを行ない、その情報が蓄積されているデータベースの名前および検索を行ないたい情報源の項目に関する概念的な情報(文献データベースにおける著者、タイトルなど)を得る。
3. UIM はこの情報をユーザ側に表の形で返す。
4. ユーザからの表への入力を得た UIM はその情報を用いて SMRO にアクセスする。SMRO では概念的な項目情報を特定のデータベース上の項目情報にマッピングを行い、UIM に返す。
5. UIM は 4 で得た情報を用いて QPRO にアクセスし特定のデータベース用の問い合わせを得る。

この問い合わせ生成過程と 3 節の特徴を関連付けて説明すると、2 によって情報源の位置透明性が確立され、2, 4 によってデータの総称化と名前の解釈が行われるといえよう。また、利用者インタフェースのユーザ支援機能の本質的な部分は 2, 4, 5 における RO とマネージャのセッションによって確立される。

4.2 セッションおよびフィードバック

次にデータベースとのセッションおよびデータベースからの検索結果に応じてシステムがフィードバックを行いながら所望の情報を得るまでの過程について説明する。

1. UIM で生成された問い合わせを受けとったネットワークマネージャ(以下 NM) は問い合わせを発す

るデータベース毎に 1 つのプロセスを生成し、セッションを行う。本システムではこの際クライアント側およびサーバ側にそれぞれタイムを用意し、ネットワーク等の障害に備える。

2. 各サーバは検索結果としてヒットしたレコードの数を NM に返す。レコード自体を転送しないのはデータ転送量を抑えるための配慮である。この情報はセッションマネージャ(以下 SM) に送られる。SM は返されたレコード数に応じて SSRO とセッションを行ない、DRO を用いるあるいはユーザにさらなる検索項目の入力を促すなどのシステムの次の挙動を決定し検索の支援を行う。ユーザに新たに入力を促す際には SM は保持している履歴をユーザに示し、ユーザの検索項目入力を支援する。
3. レコード数が適当な数になるまで 2 を繰り返した後、実際にレコードを転送する。最後に UIM は得られたレコードを概念的な項目にマッピングしユーザに示す。

5 おわりに

以上、複数の情報源を統一的に扱うことが可能な情報検索システムの構成方法として、クライアント-サーバモデルに基づいたシステムの提案を行った。

現在システムの各々の構成要素の詳細を組み立て、システムのプロトタイプを実現しつつある。

参考文献

- [1] American National Standard Z39.50-1988, "Information Retrieval Service Definition and Protocol Specifications for Library Applications", developed by The National Information Standards Organization.