

化合物データベースの統合検索方式について

7H-5

川野 繁一, 高橋 修, 吉武 静雄, 河岡 司

(NTT 電気通信研究所)

1. はじめに

各種の分散するデータベースを共通のネットワークで結合する分散データベースシステムでは、利用者の利便性の観点から、複数のデータベースをあたかも一つのデータベースと見立てて検索できる方式(統合検索方式と呼ぶ)が有効である。本稿では、科学技術庁の科学技術振興調整費による「ネットワーク共用による化合物情報等の利用高度化に関する研究」の一環で実現した化合物共用ネットワークにおける化合物データベースの統合検索方式について述べる。

2. 化合物共用ネットワークの概要

(1) ネットワークの構成

化合物共用ネットワークは、異機種種の計算機からなる複数の化合物データベースセンタをDDXバケット交換網で相互に接続した構成をしている(図1)。本ネットワークは昭和56年度から検討されており、61年度に評価実験を行うことになっている。

(2) プロトコル

各センタは異機種計算機上に異種データベース(構造、検索手順等が異なる)を持っており、それらの相互利用を可能とするため共通プロトコルを定めている。共通プロトコルは、異機種計算機間通信用に開発されたDCNAのデータベースアクセスプロトコルに準拠している。DCNAでは、種々の実データベースをネットワーク内で共通な構成を持つ仮想的なデータベース(VDB)にモデル化し、VDBに対するアクセス方法を统一的に定めている。

(3) 統合検索の位置付け

各センタ相互間は共通プロトコルが適用されるのに対し、利用者端末と依頼側センタ間のプロトコルは、ローカルインタフェースであり各センタ固有のプロトコルを適用して良い。本稿では、NTTセンタが依頼側となる場合の利用者端末と依頼側センタのアプリケーションで実現した統合検索方式について述べる。

3. 統合検索の実現方式

3.1 統合検索のタイプ

統合検索のタイプには依頼側センタ上の処理方法から見て次の2種類がある。

・検索タイプA:ある検索条件式に基づき、複数のデータベースへ並行して検索要求を送出し個々に検索結果を取り寄せ、一括して出力する形態。

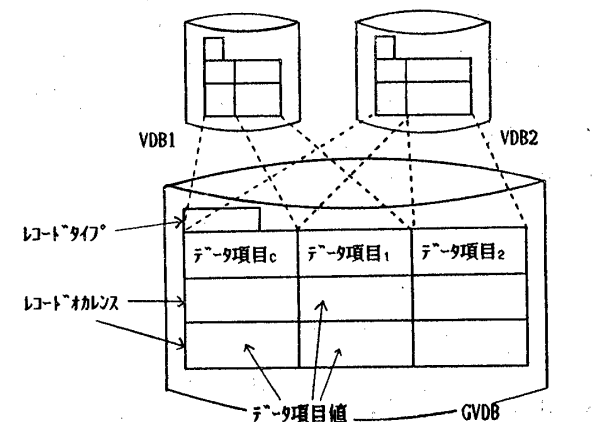
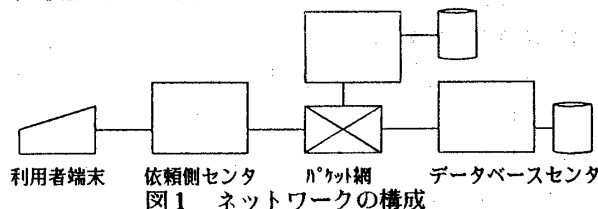
・検索タイプB:あるデータベースの検索結果に基づき他のデータベースの検索条件を定める等、複数のデータベース上のデータを検索条件式により相互に関連付けて検索する形態。

特に化合物については、参加するデータベースセンタの種々の分野の情報を横断的に検索する要求が強く、タイプBを実現する必要がある。以下では、タイプBの統合検索を実現するための依頼側センタにおけるアプリケーションの実現方式について述べる。

3.2 実現方式

(1) GVDBの構成

共用ネットワーク内の各データベースを利用者端末から見て一つのデータベースに見せるために、依頼側センタ上に各センタのデータベースを統合した形のグローバルなデータベース(GVDB)を置くこととし、個別検索とアクセス方法の統一性を図るためVDBと同じ構成とした。図2に利用者から見たGVDBのイメージを示す。依頼側センタでは、メモリ資源の有効利用の観点からGVDB、個



別VDBの制御表等の管理情報を持ち、検索処理は各データベースセンタに依頼する。

(2) 検索制御

依頼側センタでは、利用者が設定した検索条件式(等号関係等の単位比較条件式、AND/OR等の論理条件式)を分解し、関連する複数のVDBの検索を制御し、検索結果(結果データ)を利用者端末に転送する。

複数VDBの検索結果を相互に関連付けるキーとして、レコードオカレンスを一意に識別するための共通のデータ項目が必要である。本ネットワークでは各VDBに共通なデータ項目として化合物の登録番号があることから、これをキーとして使用することとした。

利用者端末から検索要求があるとその条件式を解析し、単位比較条件式の左辺(データ項目名)をVDB対応に分類する。次に条件式をAND/ORの結合式に分解して、演算優先度ごとに並び替えを行い、個別VDBごとにまとめる。

各VDBの検索制御方式には、1フェーズ検索と2フェーズ検索の二つの方式がある。

- ・1フェーズ検索: あるVDBに対する指定された条件による結果データを次のVDBの検索条件に加え、順次VDBを検索していき、目的の結果データを得る(図3a)。
- ・2フェーズ検索: 各VDBに対する指定された条件による共通データ項目(登録番号)を一旦得て(一次検索)、登録番号のAND/OR演算を行い、最終的に得た登録番号を検索条件にして各VDBを検索し(二次検索)、結果データを得る(図3b)。

各方式の比較を表1に示す。

ここでは、検索条件式の生成処理の容易さ、共用ネットワークで転送される結果データ量、結果データ保持用のメモリ量等を重視して2フェーズ検索方式を採用した。

(3) 利用者端末との機能分担

利用者端末に対して個別データベースの検索と同じに見せるため統合検索用の機能はすべて依頼側センタに分担させている。ただし、次に示すような検索結果等の画面表示機能等はすべて利用者端末に分担させている。

- ①対センタ通信機能
- ②ディスプレイ表示/プリンタ出力機能
- ③メニューを用いたコマンド入力機能
- ④データベース検索依頼機能

表1 検索制御方式の比較

	1フェーズ検索	2フェーズ検索
n個のVDBに対する検索回数	n	2n
各VDBの検索順序	逐次	一括
各VDBの検索順序制御	要	不要
各VDBへの検索条件式の生成処理	逐次	一括
フェーズ間の同期処理	不要	要
共用ネットワークを転送される結果データ量	大	小
結果データ保持用のメモリ量	大	小

4. おわりに

異種データベースの統合検索方式について述べた。本方式の特徴は、次のとおりである。

- ①GVDBの概念を導入することにより、利用者からみて単一のデータベース検索と同じインタフェースで実現したこと。
- ②ネットワークに共通なプロトコルを用いることにより、統合するセンタに管理情報のみを持ち、実際の検索処理はすべて各データベースセンタで行うようにしたこと。
- ③統合処理は、すべて依頼側センタのみで行い、各データベースセンタには統合のための処理を不要としたこと。

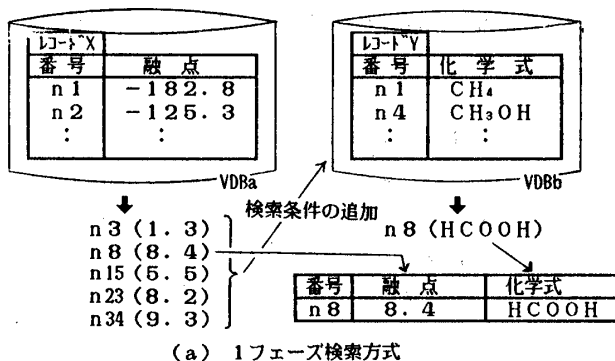
今年度は、予定されている化合物共用ネットワークの総合評価の一環で、本方式の評価を行う予定である。

最後に、本方式の検討にあたって協力して頂いた関係各位に感謝する。

[参考文献] 情処学会第29回全国大会予稿7H-1~6

検索要求: $0^{\circ}\text{C} \leq \text{融点} \leq 10^{\circ}\text{C}$ でかつ化学式に $-\text{COOH}$ を含むもの
 VDB a のデータ項目 VDB b のデータ項目

VDB a に対する検索: $0^{\circ}\text{C} \leq \text{融点} \leq 10^{\circ}\text{C}$ の化合物とその登録番号
 VDB b に対する検索: 登録番号 n3, n8, n15, n23, n34 の化合物で化学式に $-\text{COOH}$ を含む化合物



VDB a に対する検索: $0^{\circ}\text{C} \leq \text{融点} \leq 10^{\circ}\text{C}$ の化合物の登録番号
 VDB b に対する検索: 化学式に $-\text{COOH}$ を含む化合物の登録番号

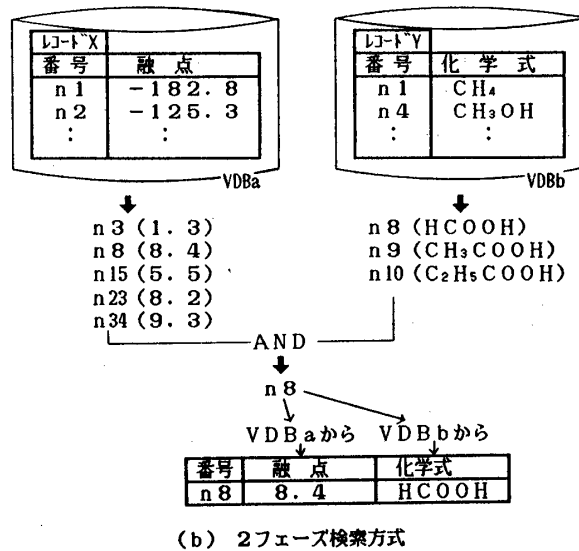


図3 統合検索の例