

マルチデータベース環境におけるマルチメディアデータの情報資源管理と検索方式の実装と実現

3 A a - 6

松井淳一 網川光明 三井一能
NTT情報通信研究所

1. はじめに

オープンなネットワークの普及により、これまで企業の業務効率化等のために蓄積されたデータベースをネットワークに接続し、有効活用しようとする動きが活発である。しかし、このようなマルチデータベース環境では、データベースの異種性の問題により、目的のデータをネットワーク上から柔軟に検索することは困難であった。これに対し、我々はネットワーク上の個々のデータベースのメタ情報と付加情報を管理する情報資源管理技術を用いて、柔軟なマルチデータベース検索を実現するエンジンであるDBSENA (DataBase SEmantic NAVigation system)を研究開発した[1]。

これまでDBSENAの検索対象は、文字、数値系の関係データベースのみであった。しかし、近年、パソコンの高機能化や低価格化、インターネットの普及により、エンドユーザまで画像データを容易に閲覧/加工できる環境が急速に整いつつある。これと並行して、大量に蓄積された画像データから画像特徴により類似検索を実現する画像データベースの研究開発も活発である。これに伴い、DBSENAの検索対象へのマルチメディアデータベースの適用を検討している。その手始めとして、マルチメディアデータベースの情報資源管理と検索方式を明らかにするため、当研究所で研究開発中の画像データベース管理システムであるHyperMatch[2]との連携を行った。

本稿ではDBSENAにおけるマルチメディアデータベースの情報資源管理と検索方式について述べる。

2. DBSENA概要

これまでのマルチデータベース検索では、データベースの異種性を解消するため、統合モデルを用いる方式が一般的であった。この方式では、データベースを分析して個々のデータベースの対応関係を定義する必要があり、システム管理者に高度な技術を要した。また、ユーザは統合モデルのデータ構造や表現形式を意識して検索する必要があった。これらの問題を解決すべく、DBSENAは、統合モデルを用いずに情報資源管理技術とマルチデータベース検索技術の2つの技術により、柔軟なマルチデータベース検索を実現する。これら技術概要を以下に示す。

2.1 情報資源管理技術

個々のデータベースについて、データベース管理システム種別、データの所在、データ構造、データの表現形式等のメタ情報や付加情報を各データベース管理システムから自動収集する[3]。これらの情報をグローバル情報資源辞書(G-IRD:Global Information Resource Dictionary)と称するDBSENA内部のデータベースで管理し、マルチデータベース検索に用いる。

2.2 マルチデータベース検索技術

SQL文のSELECT節(検索項目)とWHERE節(検索条件)のみからなる簡易な構文により指定されたユーザの検索要求に対し、以下の手順でマルチデータベース検索を実現する。

- (1) 検索要求を受け付ける。
- (2) 検索項目と検索条件で指定された項目の所在を用語辞書を用いて探索する。
- (3) 検索要求を含むデータベースを限定する。
- (4) 検索条件値をデータベース上の表現形式に変換する。
- (5) 表結合の情報を取得し、SQL文を生成する。
- (6) データベースにアクセスし、検索結果を取得する。
- (7) 検索結果をユーザの表現形式に変換する。

3. HyperMatch概要

HyperMatchは、画像データの部分画に対する高速な検索を実現する画像データベース管理システムである。部分画毎の各種特徴量(色相、彩度、明度、外形等)を抽出し、多次元インデックスにより管理する。検索インタフェースとしては、SQLライクな構文を提供する。ユーザは、画像の特徴値(キーID、特徴種別、特徴毎の優先度、類似度の計算方法)を指定することにより、類似する画像の全体画、部分画、キーワード情報等を類似度の高い順に指定件数分、取得することが可能である。

4. 画像データベースの情報資源管理と検索方式

DBSENAは、ユーザが指定する検索条件の表現形式と、ローカルデータベースの列の値の表現形式を管理している。それらが相異なる場合、ユーザ指定の表現形式をローカルデータベースの表現形式に変換し、SQLの生成/検索実行を行う。

一般に、画像データベースでは、ユーザが検索条件で指定する画像特徴(色、形等)に対応する列を保持していない。そこで、存在しない列をG-IRDに仮想的に定義し、その表現形式を管理する必要がある。これにより、DBSENAにおける画像データベース検索が実現可能となる。(図1)

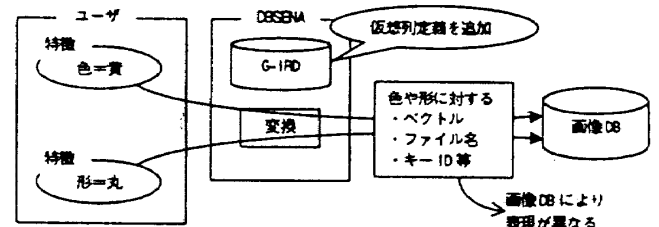


図1. 画像DB検索へのDBSENAの適用

4.1 DBSENAとHyperMatchの連携

DBSENAにおける画像データベース検索への適用性を検証するため、G-IRDに仮想的な定義を追加し、ターゲットをHyperMatchとするプロトタイプを作成した。プロトタイプでは、Webブラウザ上から入力された検索項目と画像特

徴を含む検索条件に対してDBSENAで検索要求の曖昧性を解消し、マルチデータベース環境上のHyperMatchにアクセスし、検索結果を取得/表示する。今回作成したプロトタイプシステムの構成を図2に示す。

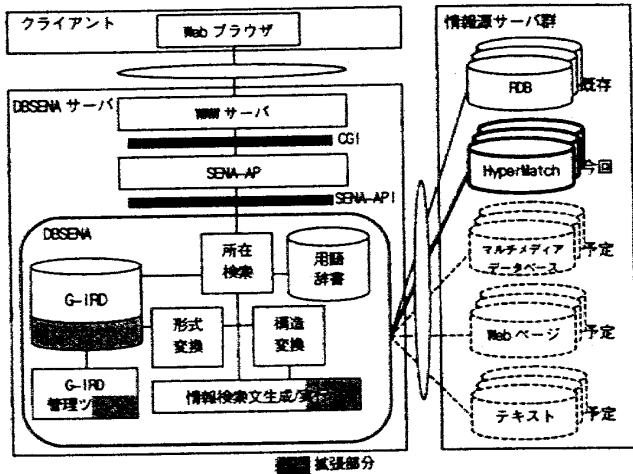


図2. システム構成

プロトタイプでは、検索を実行する前に画像特徴の表現形式定義などの準備を行う準備フェーズと実際の検索を実行する検索フェーズの2段階の処理を行う。

【準備フェーズ】

- (1) HyperMatchのメタ情報として、検索可能な画像特徴に対する項目名(色, 形, 明度, 彩度等)をG-IRDに、そのシソーラスを用語辞書に登録する。
- (2) 検索キーとなる画像ファイルを作成し、画像特徴量を抽出、IDを付与し、HyperMatchに登録する。
- (3) 画像特徴量の言葉による表現(色、形等)から(2)で登録した検索キーのIDに変換するための関数を作成し、G-IRDに登録する。

【検索フェーズ】

準備フェーズで登録された情報を基に2.2節で示したDBSENAの検索処理に基づいて、表1の通り、DBSENA内部で実行する。

表1. プロトタイプの検索処理概要

項番	処理の流れ	処理例
1	検索要求を受け付ける	SELECT 処理 WHERE 色 = 黄
2	検索項目と検索条件で指定された項目の所在を用語辞書を用いて探索する	絵→DB1. 絵画, 全体面 色→DB1. 絵画, 色相
3	検索要求を含むデータベースを限定する	HyperMatch
4	検索条件値をHyperMatchに登録された検索キーIDに変換する	黄→YELLOW_KEY_ID
5	表結合の情報を取得し、HyperMatch問い合わせ文を生成する	SELECT parentfilename FROM 絵画 WHERE 絵画 in similar (YELLOW_KEY_ID, p14ah, 1.0, L1) NUMBER 10
6	HyperMatchにアクセスし、検索結果を取得する	image_id1, image_id2, image_id3
7	検索結果に画像IDが含まれていた場合は当該IDを画像ファイル名に変換/表示する	picture01. jpg picture02. jpg picture03. jpg

4.2 実行例

HyperMatchの管理する風景や人物の写真の部分画、全体画からなる約1000の画像データに対するDBSENAからの検索実行例を図3に示す。ここでの検索条件は、

SELECT 部分画, 全体画
WHERE 色 = '黄'

である。

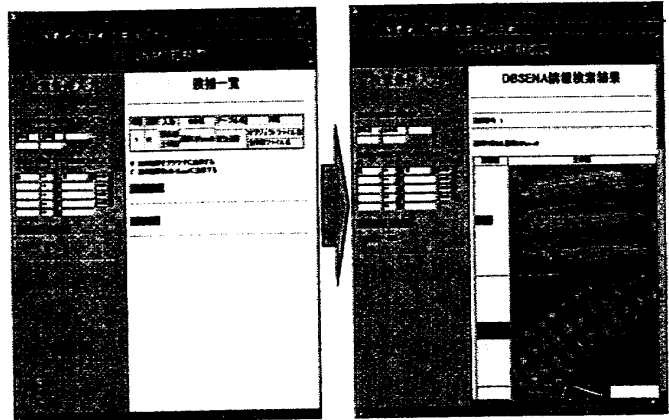


図3. 検索実行例

5. 考察

本実験では、DBSENAのアーキテクチャに仮想列の概念を導入することにより、文字、数値だけでなく、画像データを含むマルチデータベース環境で、DBSENAの検索要求に対し、目的とする画像データが検索可能であることが確認できた。画像データベースの一般的なインターフェースを持つHyperMatchとの連携の確認により、画像データベースを含むマルチデータベース環境へのDBSENAアーキテクチャ適用の有効性を確認した。

また、準備フェーズの作業を通して、HyperMatchで必要な機能(メタ情報の自動収集インターフェース、検索条件での画像特徴量の動的指定インターフェース)があることも確認した。

6. まとめ

本稿では、DBSENAにマルチメディアデータベースを情報源として拡張する際に必要な情報資源管理方法と検索方法を検討し、HyperMatchに対するプロトタイプの作成/検証により、その有効性を示した。

今後は、DBSENAをさまざまな情報源(マルチメディアデータベース、Webページ、テキストファイル等)への拡張、実用化を行っていく予定である。

参考文献

- [1] 星野 隆他: "DBSENA: マルチデータベース環境における情報資源管理と検索方式", 第114回データベースシステム研究会, 1998年1月.
- [2] キャサリン・カーティス他: "多次元特徴ベクトルを用いた画像類似検索エンジン", 第113回データベースシステム研究会, 1997年1月.
- [3] 町原宏毅他: "マルチデータベース環境におけるメタ情報収集方式の検討", 情報処理学会第54回論文集, Vol13, 1996年3月.