

DBSENA:情報資源管理辞書を用いた異種情報源検索システム

ベンチャー 4

町原 宏毅, 網川 光明, 星野 隆

NTT 情報通信研究所

{hiroki, tunakawa, hoshino}@dq.isl.ntt.co.jp

1.はじめに

近年のネットワークのオープン化に伴い、これまで個別に蓄積されてきた企業データベースをイントラネットやエクストラネットに接続し、有効活用のため、当該データベースを社内や社間等で共有する動きが活発化している。このような背景のもと、急速な情報源の増加に対しても、容易かつ安価に複数の情報源に対し検索を可能とする技術が必要とされている。しかし、個別に運用されてきた情報源には、多種多様な意味的異種性が存在し、これまでの統合モデルを用いたマルチデータベースの実現方式では、急速な情報源の増加に時間的、稼働的に対応しきれない。これに対し、我々は、情報資源管理技術を用いる方式で異種情報源検索を実現する DBSENA[1]のプロトタイプを開発した。

本稿では、情報資源管理技術を利用した情報源情報検索システム、DBSENA について紹介する。

2. 異種情報源

2.1 異種情報源の環境

個別に蓄積されてきた企業の情報源をイントラネットやエクストラネットに接続し、当該情報源を社内や社間等で共有し、地域別の商品の売れ筋情報を求めるなど、経営戦略的な情報を得るための情報源の利用が加速している。経営戦略的な情報を得るために情報源を用いる場合、1つの情報源の情報だけでは不十分な場合があり、たとえば顧客情報と売り上げ情報を組み合わせるといったように、複数の情報源の情報を結合して利用する必要がある。昨今の急速な技術革新に追従し、時代に即したシステムを構築するため、異種情報源環境の構築や維持管理(情

報源の追加/修正/削除)が、柔軟かつ容易に実現可能な技術が必要とされている。

2.2 情報源の異種性

既存の独立した情報源をもとに異種情報源が混在する環境を構築/運用するには、以下のような情報源の意味的異種性[2]を考慮しなくてはならない。

(a) アクセス方法の相違

ネットワーク上には、さまざまな検索方法(インタフェース)を提供している情報源が存在する。RDBに直接アクセスするには、SQL と呼ばれる言語が標準化されており、Web ページを用いている場合は、URL を指定したhttpによるアクセスが必要である。

(b) データの名称の相違

データの名称だけで、その所在や内容を特定することができない。例えば、同じ名称で内容が異なっていたり、異なる名称で内容が異なっていたりする。

(c) データ構造の相違

検索命令は、データ構造に併せて生成する必要がある。つまり、データ構造が異なる場合、データ構造毎に検索命令を生成する必要がある。

(d) データ表現の相違

データ表現形式に合わせて、検索条件を生成する必要がある。データの表現形式には、コード化、単位、データ型、等の様々な相違がある。

2.3 従来技術

これまでのマルチデータベース環境の実現方式として、統合モデルを用いる方式があげられる。当該方式は、マルチデータベースを構成する全てのデータベースを解析し、統合モデルを設計し、個々のデータモデルと統合モデルのマッピング定義を行うことで、データベースの異種性を解消する。当該方式は、構築時にシステム開発者の負担が大きく、維持管理面

でも急速なデータベースの増減に時間的、稼働的に対応しきれないという問題点があった。

3. DBSENA 概要

我々は、情報源の様々な意味的異種性を情報資源管理技術を用いる方式で解消し、マルチデータベース検索システムを実現する DBSENA を開発した。DBSENA のシステム構成を図 1 に示す。

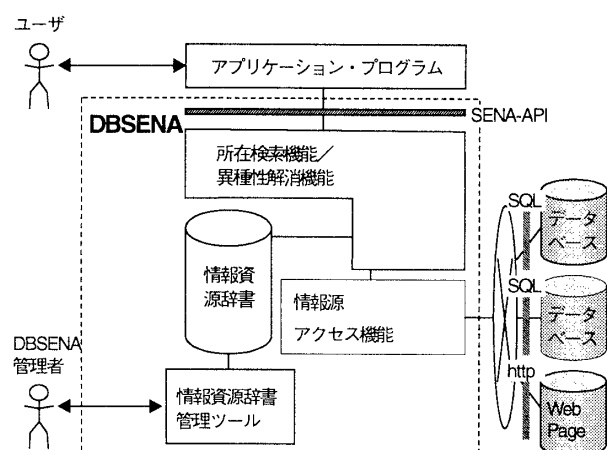


図 1. DBSENA のシステム構成
DBSENA は、以下の要素から構成される。

(a) 情報資源辞書

検索対象となる情報源のデータ構造、データの表現形式、列の取り得る値の範囲、等のメタ情報や統計情報を個別に管理する。

(b) 所在検索機能/異種性解消機能

ユーザが指定した検索要求に対し、情報資源辞書を用いて、その異種性を解消し、情報源アクセス機能を介して複数の情報源を検索する。

(c) 情報源アクセス機能

実際に個々の情報源に対してアクセスを行い、検索要求の送信、検索結果の受信、等を行う。

(d) 情報資源管理ツール

情報資源辞書の維持管理を行う。

(e) SENA-API

従来のような SQL インタフェースではなく、キーワードと検索条件のみを指定するよりシンプルな API により、DBSENA 機能をアプリケーションプログラムから利用するためのインタフェースを提供す

る。

このような機能を実現することにより、DBSENA としては、以下のような特徴を持つことができる。

① システム運用の容易性

新たな情報源の追加/削除が辞書の変更のみで実施でき、キーワード指定の SENA-API により DB 構造の変更に AP が影響を受けない。

② システム開発期間の短縮

DB 設計と独立して AP 開発が可能となるため、並行して開発が可能となる。

4. まとめと今後の課題

本稿では、様々な異種性を持つ情報源に対して、情報資源辞書を用いて異種性を解消する検索システム、DBSENA を提案した。

今後は、さらに独自のインタフェースを持つさまざまな情報源を扱えるよう拡張していくとともに、マルチメディア情報の検索などの検討を行っていく予定である。また、現在は、1つの DBSENA で全てのメタ情報を管理するアーキテクチャをとっているが、複数の会社や組織でそれぞれ DBSENA を利用して情報資源管理を行っているような環境においても適応できるように、DBSENA がお互いに協調しあって運用していくような分散 DBSENA の検討も行っていきたい。

参考文献

- [1] 星野 隆, 綱川 光明, 町原 宏毅, "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.
- [2] 上林 弥彦, "マルチデータベースの研究開発動向", 情報処理, 35(2), 1994.
- [3] Ulman, J. D., "ユーザインタフェースとしての普遍関係(第 9 章)", in データベースシステムの原理(第 2 版), 1985.
- [4] 鈴木源吾, 町原宏毅, "データベースの値の範囲の管理法とその普遍関係ユーザインタフェースへの応用", 電子情報通信学会 第 97 回データ工学研究会, 1997.