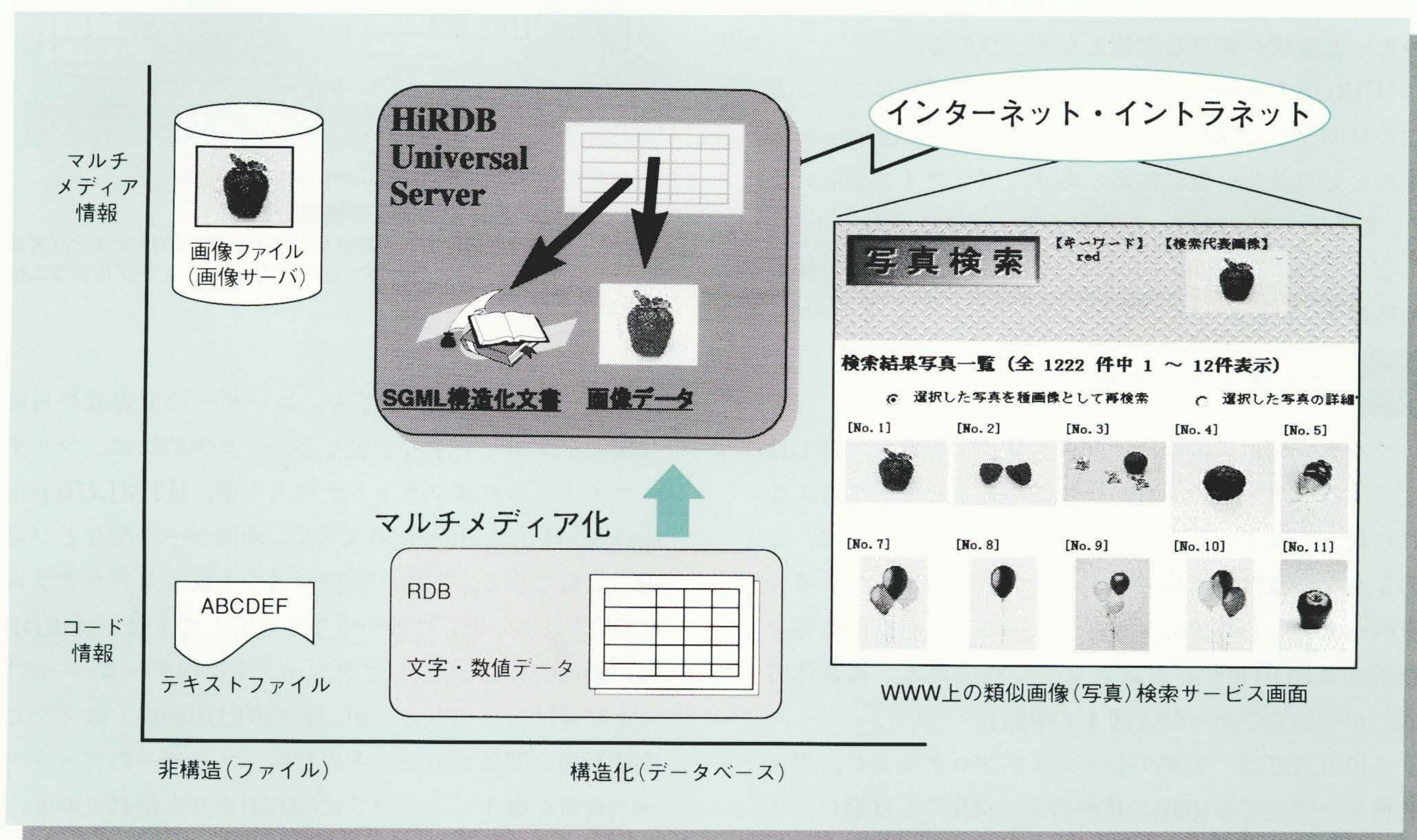


マルチメディアデータの一元管理を実現する “HiRDB Universal Server”

Integrated Multimedia Database “HiRDB Universal Server”

鳥居俊一 *Shun'ichi Torii* 松澤 茂 *Shigeru Matsuzawa*
加藤寛次 *Kanji Katô* 正井一夫 *Kazuo Masai*



注：略語説明 RDB (Relational Database), SGML (Standard Generalized Markup Language), WWW (World Wide Web)

マルチメディア情報のデータベース化による一元管理

“HiRDB Universal Server”は、従来のRDBが格納するコード情報に加えて、マルチメディア情報も表形式で格納できる。標準データベース言語“SQL (Structured Query Language)”を用いて検索・更新が簡単に記述できるだけでなく、ネットワーク上でのマルチメディア情報検索サービスや統一的な運用管理環境も、複雑なアプリケーションプログラムを開発せずに容易に実現できる。

ネットワークコンピューティング時代には、画像や音声、動画といったマルチメディアデータを取り込んだ情報システムの構築が必要不可欠となりつつある。マルチメディアデータは、データ量が大容量だけでなく、メディア操作や格納形式が多様化かつ改良拡張していくところに大きな特徴がある。従来のRDB (Relational Database) やメディアごとに別サーバを持つ方式では、システム運用やアプリケーションプログラムの開発に課題があった。

この課題に対して“HiRDB Universal Server”では、それぞれのマルチメディア情報に応じて、検索や格納な

どの操作ライブラリを自由に追加登録できるプラグイン機能を開発した。プラグインとしては、SGML (Standard Generalized Markup Language) 文書の構造検索や画像内容検索をはじめとして、今後、地図情報や映像検索などを拡充する。HiRDB Universal Serverでは、RDBとして実績のある“HiRDB”が持っている広範なスケーラビリティ、高いパフォーマンス、高信頼性などとを組み合わせることにより、従来のRDB資産を生かしつつ、マルチメディアデータの一元管理により、情報システムの新たな可能性が開けるものと考ええる。

1 はじめに

インターネットに代表されるネットワークコンピューティング時代では、より多くの人々に親しみの感じられる情報を提供する形態が重要であり、画像などのマルチメディアを利用するニーズが高まっている。インターネットへの情報発信のベースとなる情報システムでも、このような動向に対応するため、マルチメディア情報も含めた一元管理が重要な課題となりつつある。

“HiRDB Universal Server”は、この課題にこたえるために開発したORDB(Object Relational Database)を中心とした新しい製品体系であり、プラグイン機能により、多様なメディアの一元的な管理・検索が可能である。

ここでは、HiRDB Universal Serverプラグイン機能の特徴と、ネットワーク環境での応用例について述べる。

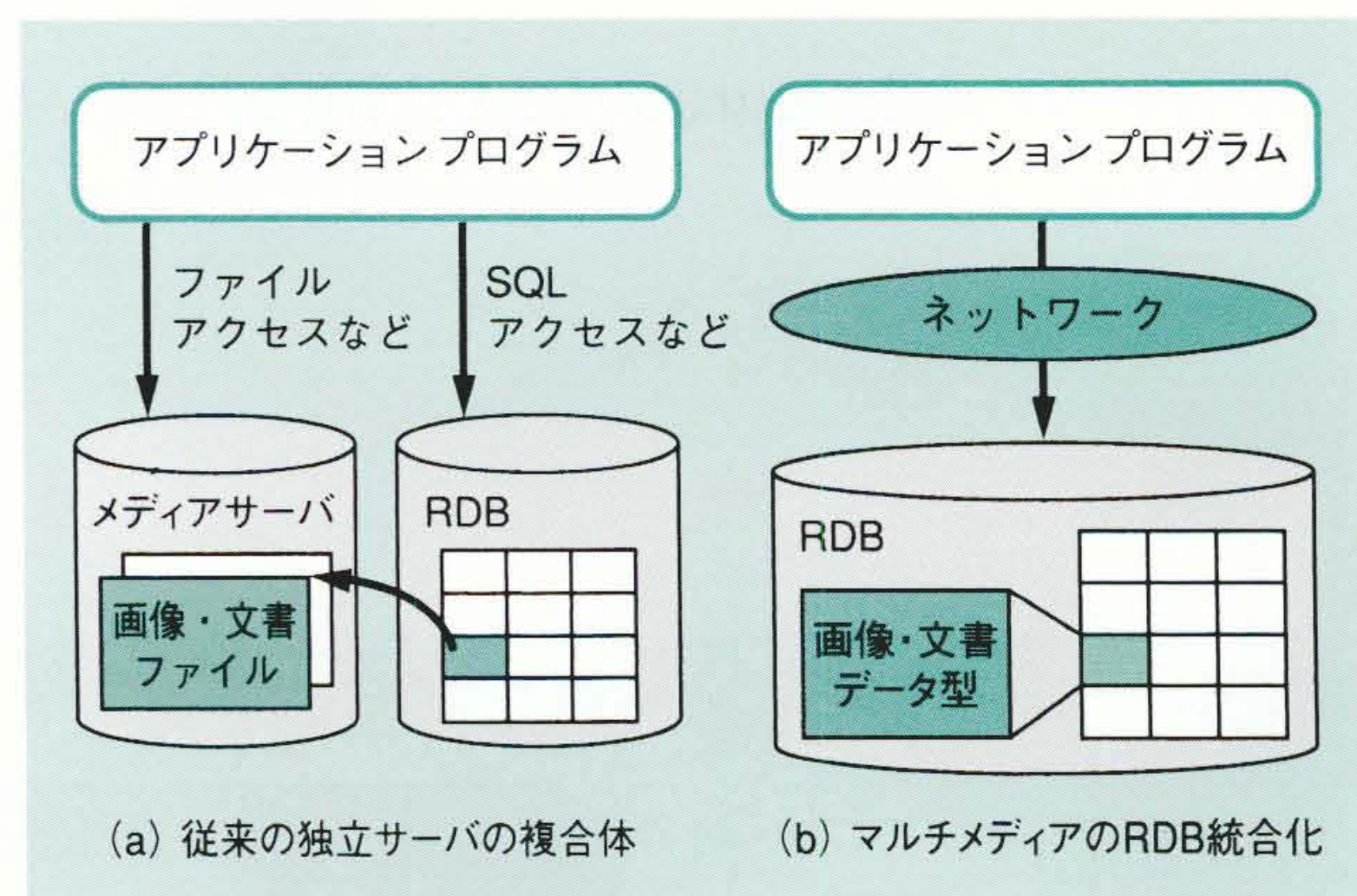
2 マルチメディア情報の特徴

マルチメディア情報は、データ量が大容量なだけでなく、各メディアに固有の操作や格納を必要とする。これに対して従来のRDB(Relational Database)では、このような多様性に対応できず、メディア対応にファイルをベースとした専用のメディアサーバを、書誌的情報を格納するRDBサーバとは別立てで構成する「複合体方式」が一般的であった〔図1(a)参照〕。

この方式では、アプリケーションプログラムが、書誌的情報を管理するRDBの標準言語“SQL”とは別に、各メディアサーバを意識したインタフェース〔図1(a)の例ではファイルインタフェース〕をそれぞれ利用する必要があるため、プログラム開発上に課題があった。また、管理者側もメディアサーバとRDBの内容に整合性の矛盾が無いように管理する必要があった。そのため、同図(b)に示すような、ネットワーク環境でのRDBによる一元管理が、アプリケーション開発者にもデータベース管理者にもそれぞれ強く望まれていた。

3 ORDBとプラグイン

近年話題となりつつあるORDBは、上記の課題を解決しようとして提案されているものである^{1),2)}。ORDBでは、従来のコード情報主体のデータ型に加えて、ユーザーが自由に型を追加定義できる。さらに、型に応じた関数をSQL文で記述できるようにすることにより、メディア対応のデータ型の拡張が可能となり、多様なメディアの出現にも柔軟に対応できるようになった。現在、標準



注：略語説明 SQL (Structured Query Language)

図1 マルチメディア情報の格納方式

画像・文書情報のRDBへの統合化により、従来の複合サーバ方式と比べて、ネットワーク環境下でのアプリケーションプログラム開発が容易となる

化が進んでいるSQL 3³⁾でも、ユーザーの型定義が可能な機能をサポートする予定である。具体的には、マルチメディアであればフラットテキスト型、HTML(Hypertext Markup Language)文書型、画像データ型などが自由に定義できる。圧縮・格納形式や生成したプログラムなどに応じて、さらに細かく型を定義してもよい。ORDBは、機能的にはOODB(オブジェクト指向データベース)にまだ及ばない点もあるが、従来のRDB機能とは完全に上位互換の関係にある。そのため、ユーザーのデータベース資産が継承できる点ではOODBよりも優れており、次世代の主流データベースと考えられる。しかし、現在のSQL3仕様書の範囲では、関数はSQL文で記述するか外部ライブラリを呼び出す必要があり、どちらも記述能力や起動性能などの問題点が十分に解決できていない。

これに対してHiRDB Universal Serverでは、新たにプラグイン機能を開発し、C言語などで記述されたインデクス機能のような高速・高機能ライブラリのORDBへの内部組込みを可能とした。構造を持つ文書に対する国際規格SGML(Standard Generalized Markup Language)で記述された文書の格納が可能な「SGMLTEXT型」の定義例を図2に示す。型の属性値“sgmlcontents”とともに、全文検索用の“contains”関数が、C言語で記述されたプラグイン関数として定義されている。ここで、contains関数の実体は、全文検索用の高速かつ高機能なインデクス機能である。

HiRDB Universal ServerのORDBは、RDBとして実績のあるHiRDB^{4)~6)}の上位互換製品であり、並列RDBサ


```
CREATE TYPE SGMLTEXT (
  PRIVATE sgmlcontents BLOB,
  ...
  PUBLIC FUNCTION contains(...)
  RETURNS ...
  LANGUAGE C
  EXTERNAL NAME '...ライブラリ名...'
  PARAMETER STYLE PLUGIN ... );
```

(a) SGMLTEXT型の定義の一部

```
CREATE TABLE 文書表 (
  文書名 varchar(64),
  本文 SGMLTEXT,
  版番号 integer,
  ... );
```

(b) SGMLTEXT型を含む表定義の一部

```
SELECT 文書名
FROM 文書表
WHERE 版番号>1 AND
      contains(本文, '内容{"日立製作所"}') IS TRUE
```

(c) SGMLTEXT型関数を含む検索例

図2 SGML文書用データ型の使用例

ユーザーが定義したSGMLTEXT型を用いて、文書を文書表の本文カラムに格納すれば、文書の内容に対する全文検索(contains関数)をSQL文で記述することができる。

サーバ上に実現している。データベース規模に応じて、データベースを格納するバックエンドを適宜追加することが可能である(図3参照)。追加されたバックエンドサーバ上で、各プラグイン関数も並列に動作が可能な点は、HiRDB Universal Serverの大きな特徴である。このように、HiRDBの持つ高拡張性(スケーラビリティ)、高信頼性、オープン性、運用・操作性が、HiRDB Universal Serverでも継承されている。

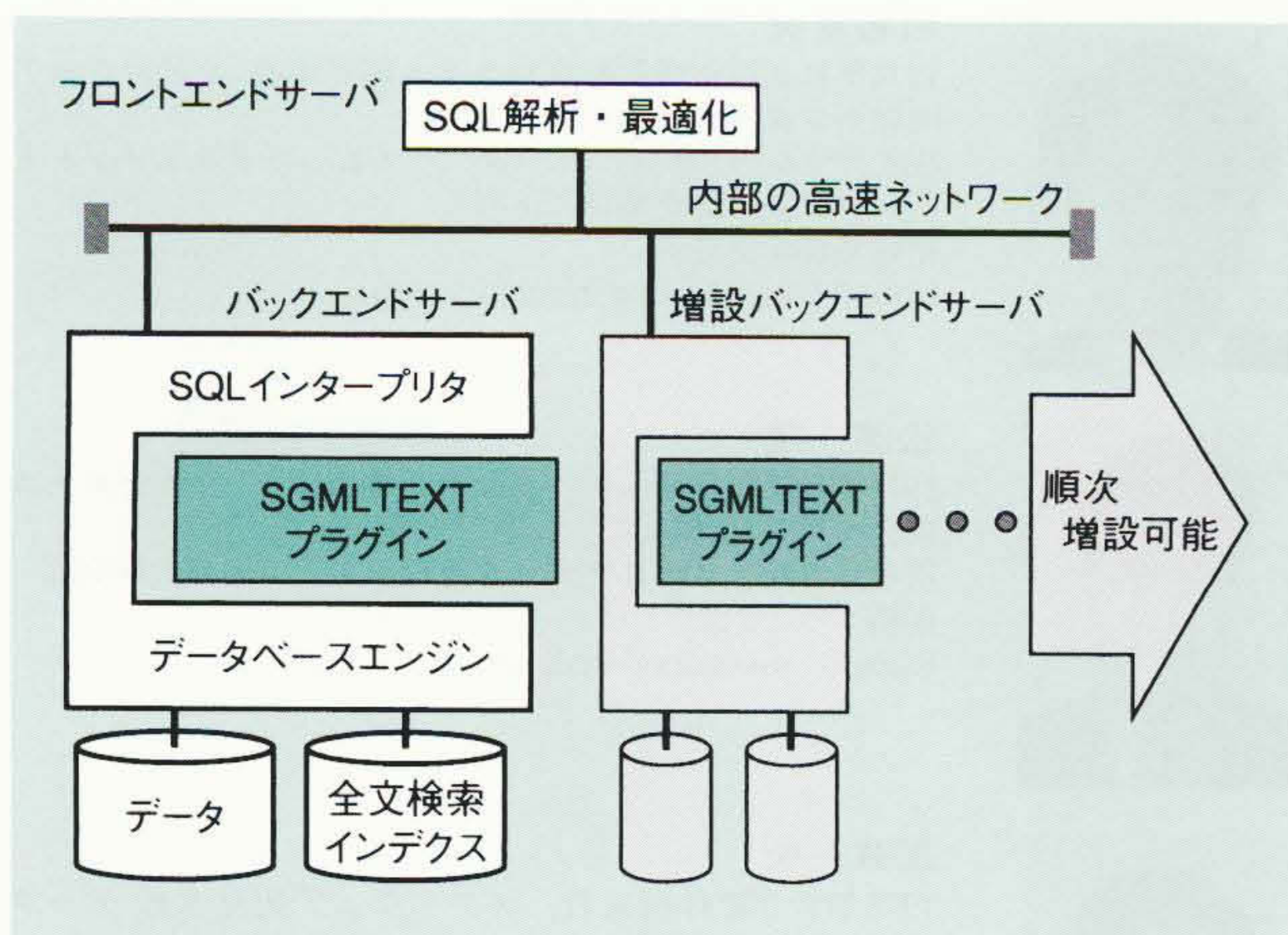


図3 プラグインと並列RDBサーバ方式

プラグイン機能は、並列RDBサーバ方式の広範なスケーラビリティにより、大容量のマルチメディアデータに対しても安定した性能を確保することができる。

4 プラグインを用いたシステム構築例

前述のSGMLTEXTプラグインを用いた企業文書管理基盤“DocumentBroker”の詳細についてはこの特集の別論文に譲り、ここでは、新聞記事データベースへの適用例について述べる。

新聞記事も図4に示すような構造を持っており、これをSGML構造文書としてデータベース化が可能である。同図の「書誌」以下の構造は、従来のRDB機能でも格納、検索することができるが、「本文」以下のテキストデータには全文検索機能が必要である。

「見出し」や「検索」の構造要素を指定した全文検索と、従来型のカラム情報に対する発行日検索を、1個のSQLコマンドで同時に実行する例を図5に示す。ここで使われている全文検索用インデクスプラグインは、日本語全文検索システム“Bibliotheca2 TextSearch”で開発したn-gramインデクス方式をベースとしている。大量の文書に対しても安定した高性能を実現するだけでなく、同図に示したような構造検索、スコアリング、同義語・異表

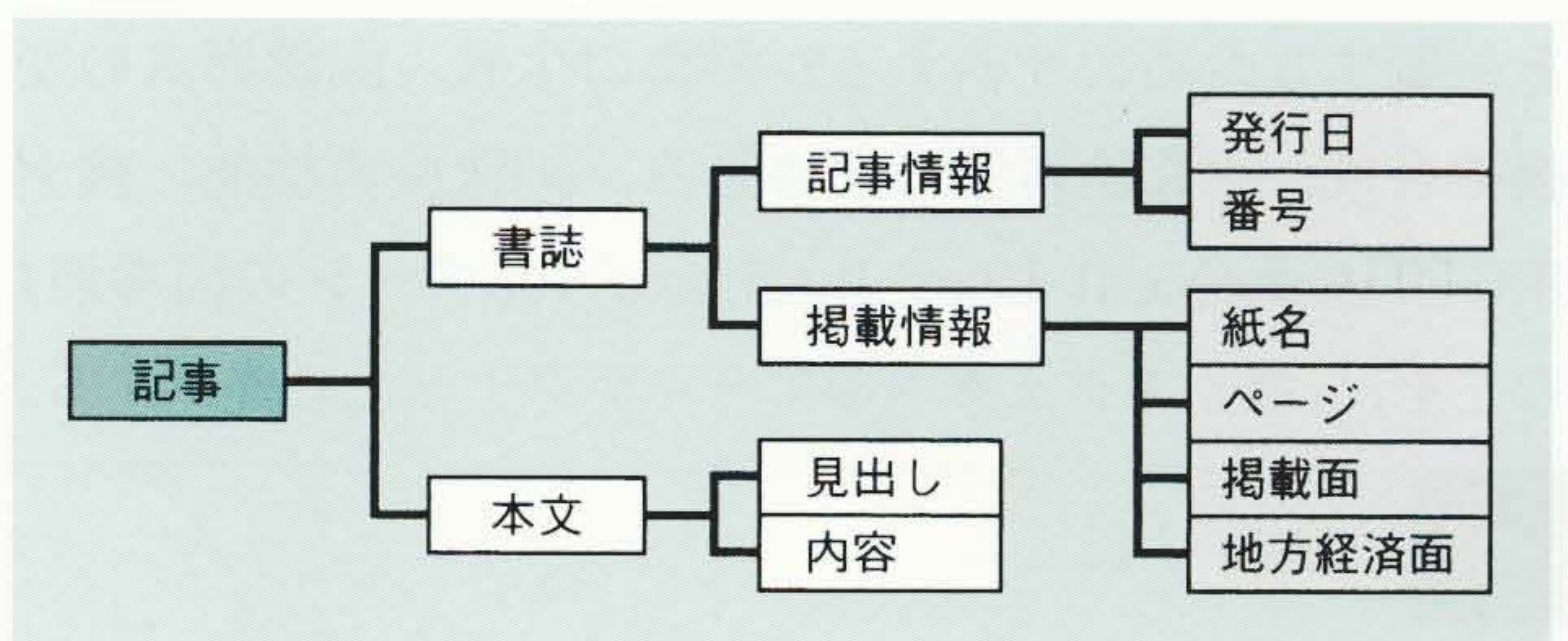


図4 新聞記事の構造

構造化文書を用いることにより、書誌的情報の格納が可能となるだけでなく、精度の高い全文検索機能も実現できる。

図5 構造指定の全文検索画面例

新聞記事データベースから、「見出し」などの構造指定の全文検索と書誌事項(発行日)の検索が、一つのSQL文で実行できる。

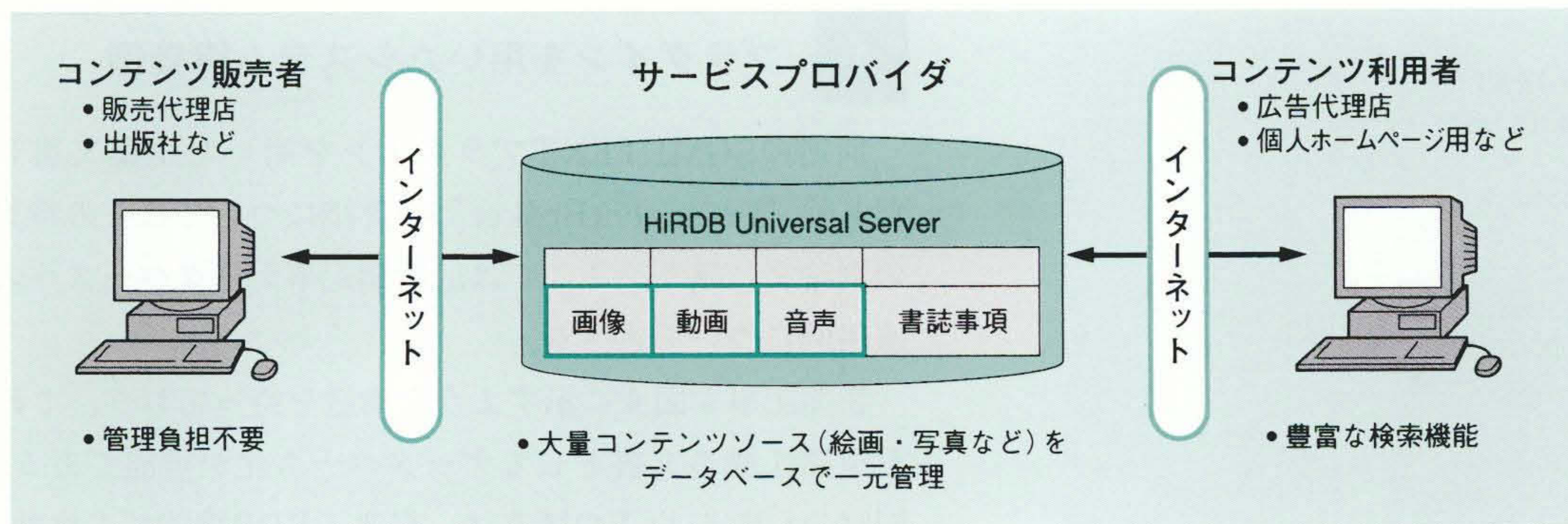


図6 コンテンツ サービス プロバイダの応用例
デジタルコンテンツの
販売で、販売者側の管理経
費を削減するとともに、利
用者はより望ましいコンテ
ンツの入手が可能となる。

記展開などの各種高機能を持っている。

このような高い技術に裏付けられた既存のライブラリを、プラグインとしてORDB本体に容易に組み込める点が、プラグイン機能の利点である。

二番目の応用例として、コンテンツ サービス プロバイダ用の画像管理システムを図6に示す。新たに提供する「画像管理システム構築コンポーネント」を使用して、インターネット上での画像、動画、音声などのデジタルコンテンツの一元管理、販売、および販売促進を行うシステムが構築できる。コンポーネントとしては、画像を一覧するためのサムネイル画像の生成、画像形式の変換や合成、著作権情報の埋め込み、画像検索機能、検索用GUI(Graphical User Interface) (65ページの図参照)などを、分散オブジェクトのアーキテクチャ上で提供する。

5 おわりに

ここでは、従来のRDB資産を継承しつつ、マルチメディア情報の一元管理を実現した“HiRDB Universal Server”について述べた。

このHiRDB Universal Serverは、ORDB本体のHiRDB Version 5.0(Object Option付き)を中心に、ここでは述べなかったデータベースアクセスツール“DABroker”や“DBPARTNER”などをも含んだ新しい製品体系である。プラグインも、SGML構造文書や画像管理に加えて、映像管理、地図情報などを整備していく計画である。

今後も、OLTP(Online Transaction Processing)業務やデータウェアハウス業務で実績のあるHiRDBの特徴を生かしつつ、ネットワーク時代の新しい企業情報システムのデータベース基盤を提案していく考えである。

参考文献

- 1) M. Stonebraker: Object-Relational DBMSs, The Next Wave, Morgan Kaufmann Publishers(1996)

- 2) D. Chamberlin: Using the New DB2; IBM's Object-Relational Database System, Morgan Kaufmann Publishers(1996)
- 3) Jim Melton(editor): 1997.ISO/IEC JTC1/SC21 N11106, N11107, N11137, N11138, ANSI SC21 Secretariat
- 4) 正井, 外: 更新処理を並列実行するUNIX向けDBMSを開発, 日経エレクトロニクス, 1995.2.27, No. 630, 101~114
- 5) 正井, 外: 並列データベース適用による新金融情報システムスケラブルデータベースサーバ“HiRDB”-, 日立評論, 77, 6, 405~410(平7-6)
- 6) 鳥居, 外: 高拡張性を目指した並列RDBサーバ, 電子情報通信学会技術研究報告, DE94-49, 41~48(1994-9)

執筆者紹介



鳥居俊一

1973年日立製作所入社, システム開発本部 所属
現在, データベース管理システムの研究開発に従事
情報処理学会会員
E-mail: torii@isrd.hitachi.co.jp



加藤寛次

1974年日立製作所入社, システム開発本部 ミドルウェア開発センタ 所属
現在, マルチメディアデータベースおよびドキュメントミドルウェアの研究開発に従事
情報処理学会会員
E-mail: kanji@isrd.hitachi.co.jp



松澤 茂

1974年日立製作所入社, システム開発本部 分散システム開発センタ 所属
現在, マルチメディアデータ管理システムの研究開発に従事
E-mail: matsuza@isrd.hitachi.co.jp



正井一夫

1978年日立製作所入社, ソフトウェア開発本部 データベース設計部 所属
現在, データベース管理システムの開発に従事
ACM会員, 電子情報通信学会会員, 情報処理学会会員
E-mail: masai_k@soft.hitachi.co.jp