

小泉 清一† 水口 弘紀† 梶野 智行†† 佐藤 朗††† 森嶋 厚行†††† 石川 佳治†††† 北川 博之††††  
 †筑波大学 工学研究科 ††筑波大学 理工学研究科 †††筑波大学 第三学群 情報学類 ††††筑波大学 電子・情報工学系

### 1. はじめに

近年、インターネットをはじめとして、ネットワーク環境が急速に発展している。これにともない、ネットワークを介した様々な情報源へのアクセスが容易になりつつある。また、情報源として、伝統的なデータベースに加え、構造化文書、Web データ、そして動画、音声といったマルチメディアデータが急増し、異種分散情報源の統合利用への要求は高まっている。

我々は、このような要求に対応するため、異種分散情報源統合システム *InfoWeaver* の研究開発を行っている。現在、*InfoWeaver* では、Web、リレーショナルデータベース、構造化文書、動画データベースを具体的な統合利用対象としている。

*InfoWeaver* は、統合データモデルとして我々が提案している WebNR/SD データモデルを採用し、メディアータ、ラッパーを用いたアーキテクチャで構成されている。

### 2. 特徴

本アプローチには以下の特徴がある。(1) 対称的なデータ変換により、目的に応じて様々なデータ表現が可能である。例えば、本環境を用いることで、各種情報源を基にして、Web ビュー、リレーショナルビュー、構造化文書ビューを構築することができる。(2) 動的なデータ変換が可能である。これによって、異種情報源に対する統合操作における自由度が増す。例えば、構造化文書のデータを動的にリレーションに変換し、リレーショナル代数演算などを適用できる。(3) 異種情報の統合を目的とした視覚的ユーザインタフェースを用いている。ユーザは、本視覚的操作系を用いて、各情報源の情報を閲覧することでデータ構造と内容を把握し、これを基に視覚的に問合せを作成することが可能である。

### 3. *InfoWeaver* の概要

図1に *InfoWeaver* のアーキテクチャを示す。メディアータは各情報源にまたがった WebNR/SD データモデルに基づく統合的な操作を与える。それぞれの情報源に配置されたラッパーは、各情報源のスキーマ情報やコンテンツ情報などをメディアータに提供する。リレーショナルデータベースと構造化文書リポジトリ、動画データベースに対するラッパーは、ローカルなそれぞれの情報源にアクセスする。また、

Web に対するラッパーは、複数の Web サーバにアクセスすることで情報をメディアータに提供する。GUI<sup>5)</sup> はユーザからの問合せを受けとりメディアータにこれを渡し結果を表示する。本環境では、問合せは以下の手順にしたがって実行される。ユーザは、GUI を介して統合操作要求をメディアータに発行する。メディアータではこれを分析し、各ラッパーで処理可能な部分問合せに変換する。各ラッパーはそれぞれの情報源で処理可能な問合せに翻訳し情報源に投入する。ラッパーはその結果を WebNR/SD 統合データモデルに変換しメディアータにこれを返す。メディアータでは各ラッパーからの部分結果を基に最終的な統合要求を処理する。GUI ではこの結果をユーザに表示する。

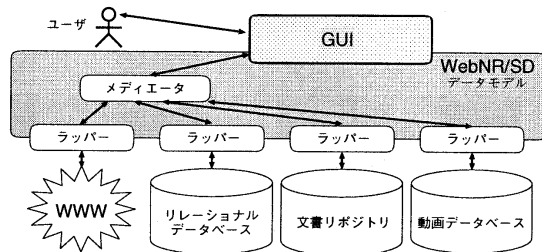


図1 *InfoWeaver*

### 4. WebNR/SD

*InfoWeaver* では、統合データモデルとして、我々が研究開発している WebNR/SD データモデル<sup>1)~4)</sup> を用いている。WebNR/SD データモデルのデータ構造は、入れ子型リレーショナルモデルに構造化文書を扱うための抽象データ型 (SD 型) を導入したデータ構造である (図2)。これにより、リレーション中では、構造化文書は SD 型の値として表現される。

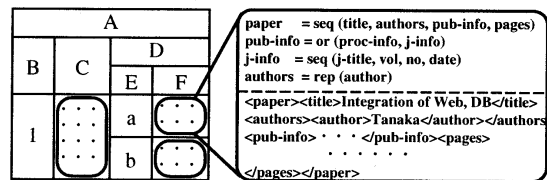


図2 WebNR/SD のデータ構造

WebNR/SD では、入れ子型リレーショナル代数と SD 値を扱うための固有の演算を提供する。これらに加えて、構造化文書と入れ子型リレーションを動的に相互変換するコンバータと、Web を操作する演算子群を持つ。これらの演算

*InfoWeaver*: An Information Integration System on Heterogeneous Information Sources. Seichi Koizumi†, Hironori Mizuguchi†, Tomoyuki Kajino††, Akira Satou†††, Atsuyuki Morishima††††, Yoshiharu Ishikawa†††† and Hiroyuki Kitagawa††††  
 † Doctoral Program in Eng., Univ. of Tsukuba  
 †† Master's Program in Sci. and Eng., Univ. of Tsukuba  
 ††† College of Information Sciences, Univ. of Tsukuba  
 †††† Institute of Info. Sci. and Elec., Univ. of Tsukuba

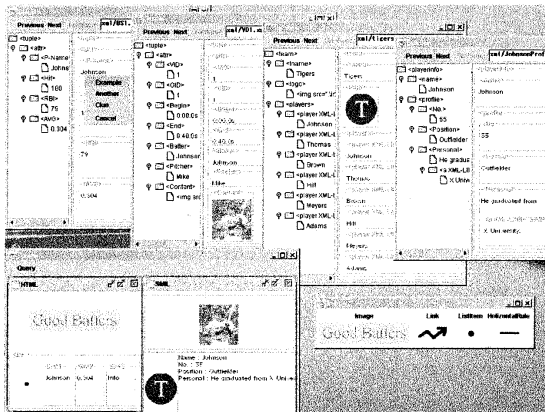


図3 視覚的操作系を用いた操作

子群を用いて Web, リレーショナルデータベース, 構造化文書に対する統合操作を行なう。

## 5. GUI

GUIは次の3種類のウィンドウで構成されている。

- DataBox:情報源中のデータ集合(例えば Web ページやリレーションのタブルの集合)を表現する。画面には集合中の1つのデータ(例えば1つの Web ページや1タブル)のみが表示される。
- Palette:汎用のページ設計用部品が表示されている。
- Canvas:ユーザはここに DataBox および Palette からオブジェクトをドロップし、欲しい情報を描画する。

GUIにおける操作は、主に DataBox での例示オブジェクトの指定と、それに対する Canvas への Drag and Drop 操作によって指定される。例示オブジェクトの指定によって操作に必要なデータの集合と、その集合間の関連が導出される。そして、Drag and Drop 操作によって、生成されるビューのレイアウトが指定される。

## 6. 統合利用例

InfoWeaver を利用した統合操作例を示す(図4)。

(1) ある RDB に野球選手の現在の成績(打率など)が格納されたリレーション BATTING-STATS(P-Name, Hit, RBI, AVG)がある。(2)別の RDB(Video DB)には打席毎の動画オブジェクトとその打席の情報(バッター名, 打席の開始時間, 終了時間など)を格納したリレーション VIDEO(VID, GID, Begin, End, Batter, Pitcher, Contents)がある。(3)Web 上に、選手リストを含む各チーム毎のページ(XML ページ)が存在し、そのページから各選手毎のページにリンクが張られている。各選手毎のページにはその選手のプロフィール情報が入っている。このとき、「打率が3割より上の各選手毎に打席の映像とプロフィール情報と所属チームのロゴ画像をまとめたビュー(SMIL ページ)を作成する。そしてインデックスページ(HTML ページ)を作成し、そこで各選手毎に選手名, 打率, SMIL ページへのリンクを示

す」という例について考える。ここで、SMIL<sup>6)</sup>とはXMLベースのマルチメディア統合言語である。

図3はこの統合操作を GUI を用いて行なっている例である。この統合は GUI を介してメディアエータ, ラッパーの協調動作によって実現される。

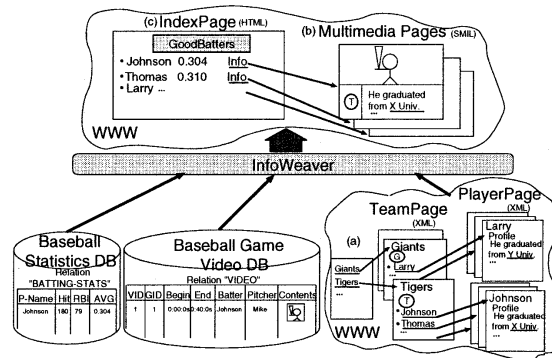


図4 統合操作例

## 7. おわりに

本稿では、異種情報源統合システム InfoWeaver の概要を述べた。現在の実装は、構造化文書リポジトリ用の文献検索エンジンとして OpenText ならびに Namazu, リレーショナルデータベースとして Oracle8 ならびに mSQL を用いている。メディアエータ, ラッパーは Java 言語を用いて開発を行った。また、モジュール間の通信には RMI を用いている。GUI は Java アプレットにより実装され Web ブラウザから実行可能である。

現在、PointCast などの情報配信サービスを統合するための枠組の研究開発と GUI の評価, 改良を行っている。

## 参考文献

- 1) Kitagawa, H., Morishima, A., and Mizuguchi, H. (2000) Integration of Heterogeneous Information Sources in InfoWeaver. *Advances in Databases and Multimedia for the New Century - A Swiss/Japanese Perspective*, World Scientific Publishing (to appear).
- 2) A. Morishima and H. Kitagawa. A Data Modeling and Query Processing Scheme for Integration of Structured Document Repositories and Relational Databases. *Proc. 5th International Conference on Database Systems for Advanced Applications*, pp. 145-154, 1997.
- 3) A. Morishima and H. Kitagawa. Integrated Querying and Restructuring of the World Wide Web and Databases. *Proc. International Conference on Digital Media Information Base*, pp. 262-271, 1997.
- 4) A. Morishima and H. Kitagawa. InfoWeaver: Dynamic and Tailor-Made Integration of Structured Documents, Web, and Databases. *Proc. ACM Digital Libraries '99*, 1999.
- 5) A. Morishima, S. Koizumi, and H. Kitagawa. Drag and Drop: Amalgamation of Authoring, Querying, and Restructuring for Multimedia View Construction *Proc. 5th IFIP 2.6 Working Conference on Visual Database Systems(VDB5)*, 2000.(to appear)
- 6) W3C. "Synchronized Multimedia Integration Language (SMIL) 1.0 Specification." W3C Recommendation, <http://www.w3.org/TR/REC-smil>.