

## 5N-09 放送型情報配信サービスを含む情報源の統合利用

梶野 智行<sup>†</sup> 水口 弘紀<sup>††</sup> 石川 佳治<sup>†††</sup> 北川 博之<sup>†††</sup>

<sup>†</sup>筑波大学 理工学研究科 <sup>††</sup>筑波大学 工学研究科 <sup>†††</sup>筑波大学 電子・情報工学系

### 1 はじめに

ネットワークの発達によって、各種情報源へのアクセスが容易になってきた。それにともない、膨大な情報の中から新鮮な情報を素早く入手することが重要になっている。このような要求から、近年放送型情報配信サービスが数多く提供されている。しかし、これらのサービスは配信する側が主体となって情報を配信するので、ユーザの希望しない情報が配信されることもある。また、数多くの配信サービスを統合して扱うことは重要な課題であると考えられる。

一方、我々の研究グループでは、WWW、RDB、構造化文書を対象とした異種情報源統合利用環境 *InfoWeaver* の研究開発を行ってきた。*InfoWeaver* では、我々が提案した統合データモデル WebNR/SD を用いている。WebNR/SD では、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構造化文書型 (SD 型) という抽象データ型 (ADT) を導入し、また、構造化文書とリレーション構造を相互に変換するための演算子や、WWW のページ群を扱うための演算子群を提供している。

本研究では、*InfoWeaver* の枠組を利用し、配信情報の再構築、他の情報源との統合、ユーザの指定したタイミングでの情報の配信といった放送型情報源の高度利用について検討する。

### 2 統合利用例

ここで、放送型情報源を含む情報源の統合利用の例を示す。企業に関するニュースを配信するサービスと、各企業の株価の情報を 1 日 1 回配信するサービスの 2 つがあるとす。また、企業の名前、URL、電話番号などの情報が格納されているリレーションがあるとす。この時ユーザは、A 社の株価が  $\alpha$  以下になったらすぐに、その日の株価と、A 社に関する過去 1 週間分のニュースをあわせて配信して欲しい。また、リレーションに格納されている A 社のホームページの URL も添付して欲しい。

### 3 放送型情報源のモデル化

#### 3.1 放送型情報源を扱う際の問題点

先に述べたように、WebNR/SD では情報源をリレーションとして扱っている。しかし、配信は次々に行なわれ、また配信されてくる情報には順序があると考え

Information Integration Incorporating Dissemination-based Information Sources  
Tomoyuki Kajino<sup>†</sup>, Hironori Mizuguchi<sup>††</sup>, Yoshiharu Ishikawa<sup>†††</sup>  
and Hiroyuki Kitagawa<sup>†††</sup>

<sup>†</sup>Master's Degree Program in Sci. and Eng., Univ. of Tsukuba

<sup>††</sup>Doctoral Degree Program in Eng., Univ. of Tsukuba

<sup>†††</sup>College of Info. Sci. and Elec., Univ. of Tsukuba

られるため、配信情報のある単位 (例えば記事) を 1 タプルとするようなリレーションとして単純にモデル化することはできない。そこで本研究では、このような情報を StreamRelation と呼ばれるデータ構造でモデル化する。

#### 3.2 StreamRelation

StreamRelation は、配信情報を扱うために、通常のリレーションにタプルの順序という概念を採り入れたデータ構造である。

##### 3.2.1 データ構造

StreamRelation のデータ構造は図 1 の様になっている。

O	A <sub>1</sub>	A <sub>2</sub>	...	A <sub>n</sub>
1				
2				
⋮				

図 1: StreamRelation のデータ構造

O は、ordering domain と呼ばれ、各タプルごとに position と呼ばれるユニークな整数値が格納されている。A<sub>1</sub>, ..., A<sub>n</sub> は StreamRelation の属性で、各属性には整数値、文字列などの他に、WebNR/SD で導入されている SD 型の値、リレーション構造の値なども格納することができる。

配信されてくる情報のある単位を、StreamRelation の 1 つのタプルとみなす。StreamRelation のデータ構造は、通常のリレーションと類似しているが、ordering domain によってタプルの順番を保証している。また、この StreamRelation にはタプルが動的に追加されるという点でも通常のリレーションとは異なる。

##### 3.2.2 演算子

StreamRelation は、タプルの順序が保証されていることを除けば、通常のリレーションと同様に扱える。したがって、Nest, Unnest のような入れ子型リレーショナル演算子においてもこの点を考慮すれば、リレーショナル代数を使用することによって処理することができる。ただし、タプルの順序を扱うためのいくつかの演算子は新たに定義する必要がある。

これらの演算子を使って最初に述べた統合利用例をどのように実現するかを図 2 に示す。

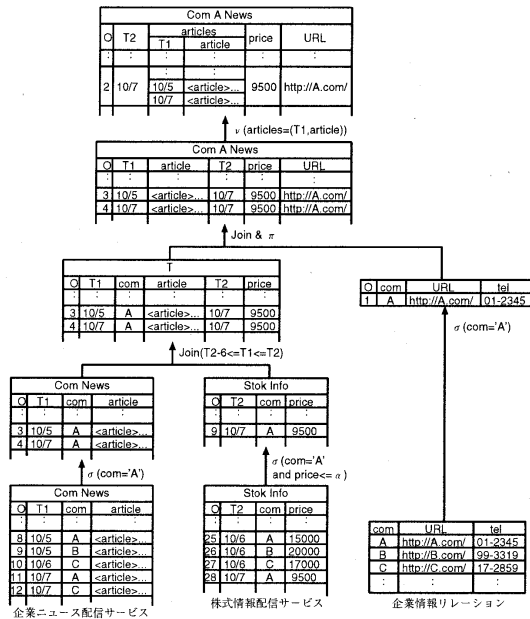


図 2: 統合利用例

#### 4 問合せ処理

StreamRelation へのタブルの追加が終了した時点で問合せ処理を行えば、ユーザが要求するデータを返すことができる。しかし、タブルの追加が終了するまで待っているのはユーザが望むタイミングでの情報の配信ができない。そこで、情報を配信するタイミングをECAルールによって記述し、その時点での部分的な結果をユーザに配信する。

##### 4.1 ECA ルールの記述

3.2.2 節で述べたように、StreamRelation に対する処理は基本的にリレーショナル代数を使用して記述することができる。この演算を用いて、2 節で述べた統合利用例におけるユーザの要求する情報の生成は以下のようなECAルールで記述することができる。ここで on 節はイベント部、if 節はコンディション部、do 節はアクション部を表す。また、*app* は追加されたタブルを表す。

```

on: arrival(株式情報)
if: app.com = 'A' ∧ app.price ≤ α
do: ∇articles=(T1,article) (
    ∏T1,price,T2,article,URL(
        (企業ニュース) ⋈T2-6 ≤ T1 ≤ T2 ∧ com='A'
        app ⋈com='A' (企業情報)))

```

##### 4.2 ユーザによる要求の記述

前節で述べたように、ECAルールを記述することにより、ユーザは要求した情報を取得することができ

る。しかし、ECAルールを直接記述することは、ユーザにとって負担が大きいと考えられる。そこで、宣言的に記述された要求を、自動的にECAルールに変換するような仕組みが必要であると考えられる。

例えば、以下のような記述を与えた時に、前節であげたようなECAルールに変換するための変換規則を定義したい。

```

select T2, articles(T1, article), price, URL
from 企業ニュース, 株式情報, 企業情報
when immediate
where com = 'A' ∧ price ≤ α ∧ T2 - 6 ≤ T1

```

#### 4.3 問合せ処理の際の問題点

図2の例では、A社の株価がα以下になったらすぐに結果を配信することを要求している。必要データが到着した時点で処理を行なう。それに対し、例えば一週間ごとに結果を配信して欲しいというような要求や、クライアントがデータを取りにきた時に処理をして欲しいという要求も考えられる。

これらの2つの処理では、保持しておく必要のあるデータの量は通常異なる。特に後者の処理の場合、保持しておかなければならないデータが膨大になる可能性がある。従って、ユーザが記述した要求から、保持しておくべきデータ量を決定することが必要となる。

#### 5 まとめと今後の課題

我々が提案した統合データモデルWebNR/SDに、StreamRelationという新たなデータ構造を導入することによる、放送型情報源を含む情報源の統合利用について提案した。また、StreamRelationに対する問合せ処理と、その際の問題点について考察した。特に、ユーザの宣言的な要求の記述をECAルールに変換し、ルールに基づく問合せ処理を行なうことを提案した。

今後の課題としては、ユーザによる宣言的な要求の記述方法や、ユーザの記述からECAルールへの具体的変換規則の導入がある。

#### 参考文献

- [1] Praveen Seshadri, Miron Livny, Raghu Ramakrishnan. *SEQ: A Model for Sequence Databases*. Proc. IEEE Conference on Data Engineering, March 1995.
- [2] Praveen Seshadri, Miron Livny, Raghu Ramakrishnan. *The Design and Implementation of a Sequence Database System*. Proc. VLDB Conference, September, 1996.
- [3] Norman W. Paton, Oscar Diaz. *Active Database Systems*. ACM CS, March, 1999.
- [4] A. Morishima and H. Kitagawa. *InfoWeaver: Dynamic and Tailor-Made Integration of Structured Documents, Web, and Databases*. Proc. ACM Digital Libraries '99, August 1999.
- [5] 水口 弘紀, 森嶋 厚行, 石川 佳治, 北川 博之. *異種情報源統合利用環境 InfoWeaver におけるプッシュ型情報配信*. 情報処理学会 第 59 回全国大会, 1999.