

## 連邦データベースシステムにおけるスキーマ構築の一方式

池田 哲夫<sup>†1</sup> 鈴木 源吾<sup>†1</sup> 町原 宏毅<sup>†2</sup> 安田 浩<sup>†3</sup>

近年、イントラネット・エクストラネットの普及に伴い、複数のデータベースを簡易に連携利用したいという要求が強くなっている。データベースを連携利用する形態の代表的なものに連邦データベースシステムがある。連邦データベースシステムは、複数の要素データベースのスキーマを基に論理的に統合されたスキーマ(連邦スキーマと呼ぶ)を構築し、利用者は連邦スキーマがあたかも通常のスキーマであるかのようにアクセスすることを可能にするシステムである。従来、連邦データベースシステムにおけるスキーマ構築の様々な方式が提案されているが、大規模なスキーマを構築する場合に適用可能な構築方式は確立されていず、より簡易な構築方式が望まれている。筆者らは、構築の各要素作業に対して効率的なユーザ支援機能を提供することが構築全体の簡易化に有効であるという考え方にに基づき、スキーマの一構築方式を提案する。その特徴は、①普遍関係を利用者インタフェースとして利用可能とし、②用語辞書を用いた類似データ項目の分類支援機能を設け、③データ項目間の値変換用の関数を標準関数として設け、さらに④要素データベースのスキーマ自動収集機能を設けることである。

## An Approach to Schema Construction in Federated Database Systems

Tetsuo Ikeda,<sup>†1</sup> Gengo Suzuki,<sup>†1</sup> Hiroki Machihara<sup>†2</sup> and Hiroshi Yasuda<sup>†3</sup>

It is now essential that the cooperation of multiple databases be made easier to achieve. In a federated database system, a typical solution, the schemas of the component databases are logically integrated into a federated schema. Once the federated schema is constructed, users can access data without regard for the individual schemas. A variety of methods for federated database schema construction have been proposed, but all have some weakness when they are applied to large scale construction problems. We think that providing efficient user support functions for each constituent task of the construction is useful in order to simplify the whole construction task. Based on this thought, we propose a method for federated database schema construction. The proposed method has the following characteristics: (i) its user interface adopts universal relations; (ii) it provides a method of clustering similar data items using a word dictionary; (iii) it provides a set of data item value conversion functions; and (iv) it provides an automatic schema information gathering tool.

### 1. はじめに

近年、ネットワークの急速な技術進歩・低廉化に伴い、組織内の複数部門のデータベースを接続するネットワーク(イントラネット)や、組織と限定された外部組織との間でデータベースを接続するネットワーク(エクストラネット<sup>1)</sup>)の構築が盛んになっている<sup>2)</sup>。

組織、特に企業はグローバルな競争に巻き込まれており、経営の効率化や経営戦略の変革を従来以上にスピーディに行うことを迫られている。このような状況のもとでは、イントラネット・エクストラネット上で、

複数のデータベースを連携して顧客管理、事務処理、在庫管理などを効率化することや、複数のデータベースから情報を抽出・集約して経営戦略の策定に活用することをスピーディに実現できることが強く要求されるようになってきている。

データベースを連携して利用する形態の代表的なものに連邦データベースシステムがある。連邦データベースシステムは、複数のデータベースのスキーマを基に、論理的に統合されたスキーマ(連邦スキーマと呼ぶ)を構築し、利用者は連邦スキーマがあたかも通常のスキーマであるかのようにアクセスすることを可能にするシステムである。これ以降、連邦データベースシステムを構成する個々のデータベースを要素データベースと呼ぶ。

従来から、連邦データベースシステムにおけるスキーマの構築は複雑な作業であり、体系的な構築方式及

†1 NTT サイバースペース研究所  
NTT Cyber Space Laboratories.

†2 NTT 第二部門  
NTT Department 2.

†3 東京大学先端科学技術研究センター  
Research Center for Advanced Science and Technology, The University of Tokyo.

びツールによる支援が必要な作業であると認識されており、様々な構築方式・ツールが提案・実現されているが、実体数、データ項目数が多い場合に現実的に適用可能な構築方式・ツールは確立されていないと考えられている<sup>3)</sup>。

未確立である原因として、①多くの従来技術において、実際のスキーマ構築者が習得するには複雑なデータモデルや構築パラダイムを導入していることや、②多くの従来技術において、現実には支援の必要な構築の要素作業の多くを容易に実行可能なものとみなして無視していることなどがあると言われている。

筆者らは、上記を踏まえ、スキーマの構築簡易化には、実際のスキーマ構築者にとって複雑な概念を導入することなく、スキーマ構築の各要素作業に対して効率的な支援機能を導入することが有効であると考え、その考えに基づきスキーマの一構築方式を提案する。その具体的特徴は、①普遍関係を利用者インタフェースとして利用可能とし、②用語辞書を用いた類似データ項目の分類支援機能を設け、③データ項目間の値変換用の関数を標準関数として設け、さらに④要素データベースのスキーマ自動収集機能を設けることである。

また、該方式を実現する連邦データベース検索システム DBSENA(DataBase SEmantic Navigation system)を試作したので、これについても報告する。

以下、本論文の構成を説明する。先ず、2章では連邦データベース技術の現状と問題点について述べる。3章では本研究における要求条件について述べる。4章では要求条件に対処するスキーマ構築方式を提案する。5章では提案方式の一実現形態であるDBSENAについて述べる。6章では関連研究との比較を述べる。最後に7章では提案方式及びDBSENAの評価について述べる。

## 2. 従来技術の問題点

先ず、連邦データベースシステムについて簡単に説明する。

連邦データベースシステムは、複数のデータベースを連携して利用者が統一的なインタフェースでアクセスすることを可能にするシステムであるマルチデータベースシステム<sup>5)</sup>の一形態であり、利用者に統一的なスキーマを提供することと要素データベースの自律性を維持することとを特長とする形態である。複数の要素データベースのスキーマのうち、各データベースの管理者が使用を許可した部分を分析・収集し、一つのスキーマ(連邦スキーマと呼ぶ)に論理的に統合し、利用者は連邦スキーマがあたかも通常のスキーマであ

るかのようにアクセスすることを可能にするシステムである。連邦データベースシステムの管理機能が、利用者からの質問の受理、質問の要素データベース向け質問への分解、要素データベースへの質問要求、要素データベースからの質問結果の組み立てと利用者への返却を行う。利用者は物理的に分散された情報の所在や各要素データベース間に存在する様々な異種性を意識する必要がない。

次に、従来の連邦データベースシステムにおけるスキーマ構築に関する研究の特徴と問題点を説明する。

スキーマ構築は、多くの比較(例:実体・データ項目間の関連特定作業における実体間の比較、データ項目間の比較)を含む作業であり、効率化のためには体系的な構築方式及びツールが必須であることが認識されている。また、ツールは完全自動が望ましいが、完全自動化のためには事前に各要素データベースの実体・データ項目の意味を完全に与えておくことが必要であり、それは現実的には不可能であることから、半自動のツール、すなわち構築者を支援するツールとせざるをえないことも認識されている。このような認識の基でスキーマ構築を支援する多くの方式・ツールが提案されてきた<sup>7)</sup>。

しかしながら、大組織(大企業や大規模地方公共団体)の大規模なスキーマ、すなわち実体数、データ項目数が多い場合のスキーマを簡易に構築可能な構築方式・ツールは未だ確立されていないと考えられている。Navatheら<sup>3)</sup>は未確立である原因として、①多くの従来技術において、実際のスキーマ構築者が習得するには複雑なデータモデル(例えば、関連や版の概念を追加してオブジェクト指向モデルを拡張したデータモデル<sup>8)</sup>などが見られる)や構築パラダイム(例えば、Prologにおける単一化(unification)と同様の技法を含む構築パラダイム<sup>9)</sup>などが見られる)を導入していることや、②多くの従来技術において、現実には支援の必要な構築の要素作業の多くを容易に実行可能なものとみなして無視していること(例えば、実体・データ項目の関連特定作業は複雑な作業であるが、その作業にとって重要な名称の類似性判定支援機能を扱った研究は少ない)などを挙げている。

## 3. 要求条件

前章で述べた問題点を踏まえ、要求条件を明確化する。

先ず、研究の前提条件を述べ、次いでスキーマの構築方式に要求される要求条件を説明する。要求条件は、複数の要素データベースのスキーマを基にスキーマを

構築するための基本的な要求条件（異種性解消条件）と、従来技術の問題点を解決するための条件（スキーマ構築容易化条件）の二つに大別される。

### 3.1 研究の前提条件

対象とする要素データベースは関係データベースとする。

これは以下の知見に基づく。組織（企業、公共団体）の内部のデータベースにおいては、関係データベースで構築されたデータベースが殆ど（8割以上）である<sup>2)</sup>。オブジェクト指向データベース、オブジェクトリレーショナルデータベースなども徐々に普及しつつあるが、組織でのビジネス用途においては、暫くの間は関係データベースが主流であり、他のデータベースは関係データベースを補完する位置付けであると考えられる。従って、要素データベースを関係データベースに限定しても研究の有用性が損なわれることは少ないと考える。

また、連邦スキーマのデータモデルも関係モデルとする。これは、連邦データベースの使い方も、個別の要素データベースと大きく変わるものではなく、関係モデルとするのが有用性が高いと考えるからである。

### 3.2 基本的要求条件

既存の要素データベースはそれぞれ他の要素データベースと異なる設計方針に基づき設計されているため、様々な異種性が生じる。

データ異種性の分類に関しては、マルチデータベースシステムにおいて、スキーマ定義言語・質問言語に要求される記述能力を明らかにすることを狙いとして、関係モデルにおけるデータ異種性を分類したKimの分類<sup>10)</sup>(表1)がよく知られている。

連邦データベースシステムにおいては、利用者からは同一の実体あるいは同一のデータ項目がたとえ異なる要素データベースに分散して存在しても表現方法の違いを意識すること無くアクセス可能なように、表1に挙げた異種性を解消する機能が要求される。

なお、表1に挙げた異種性のうち、誤データ（誤入力データ、陳腐化データ）の解消は本論文の対象外とした。誤データに関しては、初期データ入力時に入力データチェックを行うことと、初期データ投入を行う要素データベースを一個所に限定し該要素データベースから他要素データベースへ自動的にデータ流通を行うことにより、解消すべきであると考えられる。実際にその考え方にに基づき、筆者らはデータ流通プラットフォームシステムDB-STREAMを研究・開発<sup>11)</sup>し、誤データ異種性を解消可能なことを示した。

### 3.3 スキーマ構築容易化条件

スキーマの構築が容易であることが望まれる。

表1 データ異種性の分類  
Table 1 Classification of data heterogeneity

項番	異種性の分類	大分類	中分類	小分類	説明	Kimの分類との対応	
1	スキーマ的異種性	レコード対レコード	レコード対レコード	名称の差異	同名異義あるいは異名同義。	I.A.1.a Table name conflicts	
2				構成の差異	属性の多寡。	I.A.1.b Table structure conflicts	
3					DBシステム間でレコードの対応が1対多あるいは多対多。	I.A.2 Many-to-many table conflicts	
4		1データ項目対1データ項目	1データ項目対1データ項目	名称の差異	同名異義あるいは異名同義。	I.B.1.a. Attribute name conflicts	
5				データ型の差異	例: CHAR対INTEGER	I.B.1.c.1 Data Type conflicts	
6		複数データ項目対複数データ項目	複数データ項目対複数データ項目		DBシステム間でデータ項目の対応が1対多あるいは多対多。	I.B.2 Many-to-many attribute conflicts	
7				レコード対データ項目		片方のDBシステムでのレコードが他方のDBシステムのデータ項目に対応。	I.C Table-versus-attribute conflicts
8		体系的異種性	誤データ	誤入力データ	(片方あるいは両方のDBシステムのデータに)正しくない値が入力されたこと。	II.A.1 Incorrect entry data	
9				陳腐化データ	異なるDBシステムのデータ間において更新の同期がとれていないこと。	II.A.2 Obsolete data	
10				値の表現方法の差異	異なる表現	例: 会社名での正式名称対略称	II.B.1 Different expressions
11					異なる単位	例: 長さでのm対cm	II.B.2 Different units
12					異なる精度	領域(domain)の基数が異なること。例: 長さでの整数表示対1重、中、細の3段階表示	II.B.3 Different precisions

注1)スキーマ定義文(SQLのスキーマ定義文)の記述の差異として表現できる異種性。

注2)スキーマ定義文では差異を記述できない異種性。

実際のスキーマ構築者にとって複雑な概念を導入することなく、スキーマ構築の各要素作業に対して効率的な支援機能を備えたスキーマ構築方式が望まれる。

## 4. スキーマ構築方式

要求条件を満たす、スキーマ構築の一方式を提案する。

まず、スキーマ構築方式を説明し、次いで該方式に対応するスキーマと連邦データベース管理機能とについて説明する。

### 4.1 スキーマ構築方式

まず、スキーマ構築の主要な作業の概要について説明する<sup>7)</sup>。

- ・ 共通データモデルへの変換  
要素データベースのスキーマを収集し、共通データモデルを用いたスキーマに変換する。
- ・ 実体・データ項目間の関連の特定  
実体間の同義や包含や、データ項目間の同義や多対多対応など、実体・データ項目間の関連を検出し、整理する。
- ・ 統合スキーマの作成  
関連を特定された実体・データ項目を基に、論理的に統合されたスキーマ（連邦スキーマ）を作成する。統合スキーマ作成と同時に様々な異種性の解消も行う。

以下、上記の作業フェーズに沿って、本論文で提案

するスキーマ構築方式を説明する。

### (1) 共通データモデルへの変換

スキーマ構築作業全般を通して実体レベルの作業の稼働の削減を図るため、利用者インタフェースとして普遍関係<sup>12)</sup>を採用することとする。普遍関係とは、データベース中の全データが必ずその中に含まれる単一の(仮想的な)関係を言う。普遍関係は、関係データベースでの質問を単純化すること、即ちデータ項目と関係の組を意識すること無くデータ項目だけを意識すれば質問可能とすることを主たる狙いとして導入・研究されてきた概念である。

普遍関係の採用に伴い、統合スキーマ作成フェーズ(3)における実体レベルの統合作業が不要となり、従って本フェーズではその準備作業の一つである要素データベースのスキーマの共通データモデルへの変換が不要となる。

本フェーズでは要素データベースのスキーマの収集のみ必要となる。

要素データベースのスキーマの収集に関しては、自動収集機能を設けて効率化を図ることとする。同じ関係データベースでも製品毎にスキーマ構造は少しずつ異なる。そこで、代表的な商用関係データベース管理システム毎にスキーマ自動収集機能を設けることとする(具体的な対応製品名は省略する)。このようにしたのは、ビジネスの場において使用される商用関係データベース管理システムの種類はほぼ収束しており、それらに対応しておけば新たな機能の作り込みが必要になる可能性は小さいと考えたためである。

また、スキーマ的異種性のうちの1データ項目対1データ項目の場合のデータ型の差異と、データの異種性のうちの値の表現方法の差異とを解消するために、ドメイン概念を設け、各データ項目毎にドメイン値を設定することとする。

ドメインとは、筆者らの属する研究グループで研究・開発したデータベース設計法<sup>13)14)</sup>におけるデータ項目値の管理情報と同じものであり、具体的には、データ型と、精度と、null不可情報と、値範囲情報と、自然言語による説明とからなる。ドメインとして管理する情報の種類をこのようにしたのは、データベース設計法の使用経験、特に既存の複数データベースを更改して新規データベースを設計した経験から、データ型と値の表現方法の異同を判断するには実用上十分であると判断したためである。

### (2) 実体・データ項目間の関連の特定

普遍関係を利用者インタフェースとして採用することにより統合スキーマ作成フェーズ(3)における実体

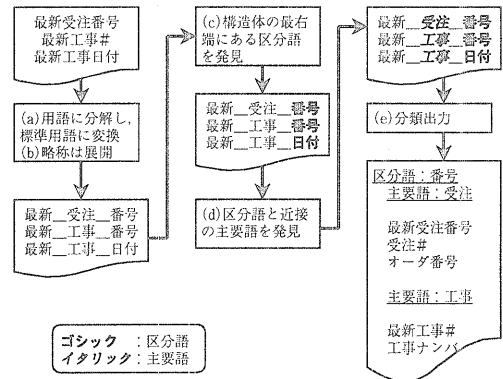


図1 類似データ項目分類の手順

Fig.1 Procedure of clustering similar data items.

レベルの統合作業が不要となることに伴い、本フェーズではその準備作業の一つである実体間の関連特定が不要となる。

本フェーズではデータ項目間の関連特定のみ必要となる。

筆者らの属する研究グループにおいて、データ標準化ツールの一部として開発した、類似データ項目の分類支援機能<sup>15)</sup>を用いてデータ項目間の関連特定の効率化を図ることとする。

類似データ項目の分類支援機能を簡単に説明する。

- ・ 予め汎用の類義語辞書をベースに用語辞書(類義語辞書)を作成する。
- ・ 予め用語辞書において、全ての用語を Durell の分類<sup>16)</sup>に基づき、値の種類を表す用語である区分語(例:番号,日付)と、データ項目が管理する対象を表す用語である主要語(例:顧客,回線,工事)と、データ項目の内容を区別する語である修飾語(例:新規,重要,上位)との三つに分類する。
- ・ 分類支援機能実行時に自動的に、データ項目の名称を形態素解析し、分類対象のデータ項目名集合の中から、名称の区分語及び主要語が互いに類義語の関係にあるものを類似データ項目の候補として分類する。アルゴリズムの概要を示す(図1)。
  - (a) データ項目名を用語に分解し、それぞれを用語辞書を用いて標準用語に分類する。
  - (b) 略称(#など)がある場合、略称を標準用語に展開する。
  - (c) データ項目名を構成する用語を末尾から順に左に探し、区分語になりうる標準用語を探す。
  - (d) 見つかった区分語から、さらに左方に主要語を探す。

- (e) 区分語を大分類，主要語を小分類として，データ項目を分類する。

該支援機能で類似データ項目の分類が行われた後は，人間の判断により，分類されたデータ項目のグループの中からデータ項目間の関連を特定する。

この支援機能を用いたのは，他の類似データ項目分類支援の方式と比較して効率的に優れた方式であることが検証されている<sup>15)</sup>ためである。

(3) 統合スキーマの作成

普遍関係を利用者インタフェースとして採用することにより，実体レベルの統合作業，即ちテーブルの設計作業が不要となる。

本フェーズではデータ項目レベルの作業のみ必要となる。データ項目レベルの異種性解消作業は，特定された関連の種類に応じて作業を行う。作業例を以下に示す。

- ・ 同義のデータ項目に関しては，先ず対応する連邦スキーマのデータ項目を設け，次いで連邦スキーマのデータ項目と要素データベースのデータ項目との間の変換関数を決定し，必要ならば変換関数の作成を行う。
- ・ 複数データ項目と複数データ項目の対応に関しても同様の作業を行う。例えば，関連特定において，ある要素データベースのデータ項目「販売日付」が，他の要素データベースのデータ項目「販売年」「販売月」「販売日」と対応していることが特定できたとする。この時，例えば，先ず対応する連邦スキーマのデータ項目としては「販売日付」を設けることとし，次いで該データ項目と要素データベースのデータ項目との間の変換関数を決定し，必要ならば変換関数の作成を行う。

ここで，変換関数作成の稼働削減を図るため，予め標準的な変換関数を設けることとする。変換関数の品揃えとしては，データ流通プラットフォームシステムDB-STREAM<sup>11)</sup>で設けた変換関数(表2)をそのまま設けることとする。DB-STREAMでは，既存の8つのデータベースシステムをサンプル調査し，データベースシステム間でのデータ変換機能を実現するための機能としてデータ型変換関数や文字列操作関数など58種類の変換関数を抽出し，それらの関数を関数ライブラリとして設けた。DB-STREAMで設けた関数群をそのまま設けることとしたのは以下の知見による。変換関数は，新たなドメインが現れる毎に必要ななりうる(例:年月日の全く独自の表現方法)が，DB-STREAMの使用経験を通して，上記関数群を用意しておけば，追加が必要となる頻度はそれほど大きくないと期待できることが判

表2 変換関数(抜粋)

Table 2 Data item conversion functions (extract)

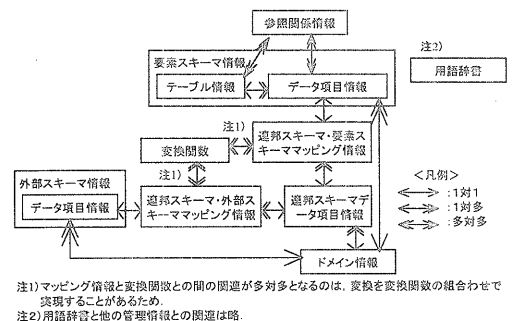
データ項目値変換関数の分類			
項番	大分類	中分類	
1	データ型の変換	データ型変換	INT->CHAR
			CHAR->INT
			>SHORTINT
			その他30種
2	データ項目の統合・分解	文字列操作	文字列結合
			部分文字列
		算術演算	左シフト/右シフト
			その他11種 (文字探索/置換等)
3	データの値の体系の変換	対応表変換	四則演算
			剰余
			最大,最小
			対応表変換

明しているためである。

なお，普遍関係を利用者インタフェースとして用いる場合は，予め参照関係(キーデータ項目と外部キーデータ項目の関係)情報を登録しておく必要がある。これは，質問処理において，参照関係情報を基に結合条件を生成しなければならない場合があるからである(4.3節で説明する)。ここで，参照関係情報の登録方式は，以下の方式とする。(2)で説明した類似データ項目の分類支援機能による分類が行われた後，人間の判断により分類されたデータ項目のグループの中から先ずテーブルのキーであるデータ項目を抽出する。次いで，グループの中から同義のデータ項目(すなわち外部キー)を抽出し，キーデータ項目と外部キーデータ項目の組を登録する。この方式では，人間の判断が必要となるが，類似データ項目の分類支援機能の利用により判断対象が大きく絞り込まれるため，それほど大きな負担なく判断可能と考える。

4.2 連邦データベースのスキーマ

連邦データベースのスキーマにおいては以下の情報



注1) マッピング情報と変換関数との間の関連が多対多となるのは，変換関数を変換関数の組合わせて実現することがあるため。  
注2) 用語辞書と他の管理情報との関連は略。

図2 連邦データベースのスキーマ  
Fig. 2 Schema of a federated database.

を管理する(図2).

- ・ 要素スキーマ情報  
要素データベースのスキーマ情報. 主な管理対象はテーブルとデータ項目である. テーブルに関しては名称, 構成データ項目を管理し, データ項目に関しては名称, ドメイン情報を管理する.
- ・ 参照関係情報  
要素データベースにおける異なるテーブルのデータ項目間の参照関係を管理する.
- ・ 連邦スキーマ情報  
連邦スキーマのスキーマ情報. 連邦スキーマに含まれるデータ項目を管理する. 各データ項目に関しては名称とドメイン情報を管理する.
- ・ 外部スキーマ情報  
質問を発行する利用者向けのスキーマ情報. 該当スキーマに含まれるデータ項目を管理する. 各データ項目に関しては名称とドメイン情報を管理する.
- ・ マッピング情報  
要素スキーマ, 連邦スキーマ, 外部スキーマのデータ項目間の対応関係を管理する. 併せてそれらの間の変換用の変換関数名も管理する.
- ・ 変換関数  
データ項目の値を変換する関数.
- ・ 用語辞書  
類似データ項目の分類支援のために用いる用語辞書.

#### 4.3 連邦データベース管理機能

利用者が連邦スキーマを用いて連邦データベースにアクセスするのを可能にするための, 連邦データベース管理機能について説明する.

なお, 本研究においては, 連邦データベースの更新機能は将来課題とし, 検索関連機能のみを研究対象とした.

##### 4.3.1 スキーマ管理機能

スキーマを管理する機能である.

通常のデータベース管理システムのスキーマ管理機能と同様な機能, 即ちスキーマデータを格納効率良く格納する機能と, 無矛盾性・耐障害性を保証する機能と, 実行時処理機能とのインタフェース機能と, スキーマ入力・編集機能とを設ける. さらに既述の要素スキーマ自動収集機能を設ける.

##### 4.3.2 実行時処理機能

実行時の処理を行う機能である.

##### (1) 質問の構文

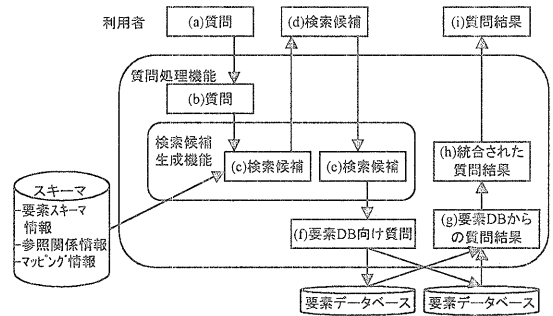


図3 質問処理

Fig.3 Query processing.

質問の構文は, SQL<sup>17)</sup>の構文を簡易化した以下の構文とする. 普遍関係を採用したため, FROM節は不要となる.

項目名リスト WHERE { (項目 比較演算 項目 or 定数) を論理演算子で結んだもの }

##### (2) スキーマ間変換機能

外部スキーマ, 連邦スキーマ, 要素スキーマ間でデータ項目の名称と値の変換を行う機能である.

##### (3) 質問処理機能

利用者から発行された質問を処理する機能である.

図を用いて機能を説明する(図3).

- ・ 利用者からの質問(図3(a))をスキーマ間変換した質問(図3(b))に対して, 検索候補生成機能が検索候補(図3(c))を作成する. その際, (4)で説明する判断支援情報を検索候補に付加する.
- ・ 提示された検索候補(図3(d))から, 利用者に所望の検索候補を選択させる.
- ・ 選択された検索候補(図3(e))を要素データベースに対して発行する. 検索候補が複数データベースに跨る場合は, 各データベース向けの質問(図3(f))に分解してから発行する.
- ・ 各要素データベースからの質問結果(図3(g))を統合し(図3(h)), さらにスキーマ変換を施して利用者に返却する(図3(i)).

##### (4) 検索候補生成機能

これは, 質問処理機能の部分機能であり, 普遍関係を利用者インタフェースを採用した場合に問題となる曖昧性の問題を解決するための機能である.

曖昧性とは, 質問の構文中に現れるデータ項目が複数のテーブルに跨る場合に質問処理機能において結合条件(複数の結合条件の接続もありうる)の付加が必要となるが, その際結合条件の候補が複数生じうる事象を言う<sup>18)</sup>. 付加される結合条件は, テーブル間の参照関係の辿り方を意味するが, それが, 質問を発行した

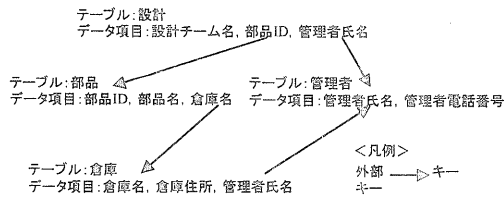


図 4. 要素データベースの例  
Fig. 4 Example of an element database.

利用者の意図と一致しているとは限らないことが問題となる。

曖昧性を例で説明する。図 4 に示すように、設計、部品、倉庫、管理者の 4 つのテーブルからなる要素データベースを考える。4 つのテーブルの間には図中の矢印で示す参照関係が存在すると仮定する。この時、以下の質問を利用者が発行することを考える。

管理者氏名, 管理者電話番号  
WHERE 部品ID=" abc123"

2 通りの結合条件の組み合わせが考えられ、それぞれの場合の質問の意味は異なる。

- ① 部品と倉庫とを倉庫名で結合し、併せて倉庫と管理者とを管理者氏名で結合する。これは部品を管理する倉庫の管理者情報の質問を意味する。
- ② 部品と設計とを部品IDで結合し、併せて設計と管理者とを管理者氏名で結合する。これは部品を設計する部門の管理者情報の質問を意味する。

曖昧性の問題に関しては以下の方法で対処することとする。

検索候補生成機能において検索候補を提示する際に、質問発行者の判断支援用に以下の情報（判断支援情報と呼ぶ）をスキーマから取得し付加する。

- ・ 質問文に陽に現れるデータ項目に関してその属するテーブル名とデータベース名
- ・ 結合条件を付加する場合、結合条件及び各結合キーデータ項目の属するテーブル名とデータベース名

質問発行者が、判断支援情報を用いて、提示された複数の候補の中から所望の候補の選択を行う。

曖昧性の説明の①を例に用いて、具体的な判断支援情報の提示方法を以下に示す。データ項目の属するテーブル名は、<検索項目>と<検索条件>とにおいてデータ項目名を修飾する名称として示される（データベース名は簡単化のため略してある）。結合条件及び各結合キーデータ項目の属するテーブル名は、<関連情報>として示される。

<検索項目>管理者, 管理者氏名, 管理者, 管理者

電話番号

<検索条件>部品, 部品ID=" abc123"

<関連情報>部品, 倉庫名=倉庫, 倉庫名 and 倉庫, 管理者氏名=管理者, 管理者氏名

本方式では、質問発行者による候補選択作業が必要となるが、判断支援情報を明示することにより、大きな負荷なく候補の選択が可能と考える。

### 5. 連邦データベース検索システム DBSENA

4 章で述べたスキーマ構築方式を実現する一システムとして、連邦データベース検索システム DBSENA を試作した。

#### 5.1 アーキテクチャ

DBSENA は、大きくスキーマと連邦データベース管理機能からなる。DBSENA のアーキテクチャを図 5 に示す。

#### 5.2 スキーマ

スキーマは 4.2 節で説明したスキーマを実現するものである。実現容易性を考慮して市販のデータベース管理システムを用いて作成した。

#### 5.3 連邦データベース管理機能

連邦データベース管理機能は 4.3 節で説明した連邦データベース管理機能を実現するものである。

利用者（アプリケーションプログラム）が実際に連邦データベースにアクセスする際に用いるインタフェースである、DBSENA のアプリケーションプログラムインタフェース (API) について特徴を説明する。

API としては、コールレベルインタフェース<sup>17)</sup>を採用した。これは、他の有力なインタフェース方式である埋め込み質問言語インタフェースと異なり、コールレベルインタフェースではプリプロセスが不要となることにより、アプリケーションの構築が単純化され、デバッグ作業が容易化されるためである。具体的には C

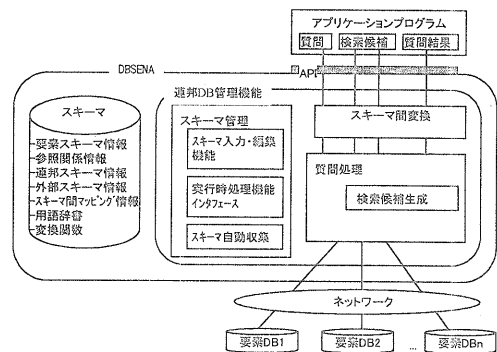


図 5 DBSENA のアーキテクチャ

Fig.5 Architecture of DBSENA.

表3 DBSENA-API(抜粋)  
Table 3 DBSENA-API(Extract)

分類	API名 (C関数名)	説明
質問の操作	DBsena_Query_createQuerySet	質問を生成し、DBSENAに発行する。
検索候補の操作	DBsena_QuerySet_getQuery	判断支援情報を付加された検索候補を取得する。
質問結果の操作	DBsena_RelultSet_next	質問結果内で参照ポインタを次のレコードに設定する。
	DBsena_RelultSet_getCharArray	カレント参照レコードの指定された位置の文字列データを取得する。

関数のライブラリとして API を提供した。また、API 仕様の策定にあたっては、将来 JAVA などのオブジェクト指向言語に向けて API を拡張する際の拡張容易性を考慮してサン・マイクロシステムズ社が開発した JDBC (Java Database Connectivity)<sup>19)</sup>を参考にした。

API は、操作対象のデータ種別によって、質問の操作、検索候補の操作、質問結果の操作などに大別される。主な API を表 3 に示す。

## 6. 関連研究

本研究の特徴のうち、類似研究例の少ない、普遍関係の利用と用語辞書を用いた類似データ項目分類支援とに関して、関連研究との比較を行う。

### 6.1 普遍関係の利用

普遍関係を利用者インタフェースとして利用する研究を中心に関連研究を説明する。

Zao<sup>18)</sup>は、製造業の複数社のデータベースを連携する方法を提案している。そのような環境ではデータベースの自律性が強く要求されることから、普遍関係を用いることにより緩やかな協調を図る方法を提案している。

利用者は普遍関係のデータ項目名を用いて質問することが可能である。質問結果には以下のような情報(文脈情報と呼ぶ)が付加される。これは、4章で説明した普遍関係における曖昧性の問題を解決するための情報である。

- ・ データ項目が含まれる要素データベース名
- ・ アクセスパス情報(質問の実行において結合演算が必要になる場合どのような結合演算が用いられるかを示す)
- ・ (金額の単位などの)スケール情報

利用者は文脈情報を用いることにより自分の意図と一致する質問結果を選択する。以上により、利用者が簡単に質問文を作成し、解釈容易な結果情報を返却してもらうことが可能になる。

この方式の問題点は、データ項目レベルのスキーマ構築作業に関わるものである。この方式においては、データ項目間の関連を判定する際の支援機能が無く、また、データ項目間の変換機能がすべて利用者記述(ルール或いは汎用プログラミング言語による関数作成)となっている。データ項目レベルのスキーマ構築作業が筆者らの方式と比べて同等あるいは容易とは考え難い。

Reck<sup>ら</sup><sup>20) 21)</sup>は、異種情報源アクセスへの透過的なアクセス(情報源の種類・所在などを気にせずに済むアクセス)を可能とするためのアーキテクチャを提案している。その一部として、普遍関係に基づく利用者インタフェースを提案している。

Reckらのアプローチは、スキーマを構築するアプローチではなく、メディエータと呼ばれる意味的異種性を解消するソフトウェアを複数構築することにより、複数情報源へのアクセスを可能にするアプローチである。

スキーマの構築にほぼ相当するのは、メディエータの作成作業と情報源をアクセス可能とするためのソフトウェア(ラッパ)を作成する作業とである。

この方式の問題点は二つある。一つは、普遍関係における曖昧性の問題を解決するための手段が示されていないことである。もう一つは、以下に示すようにデータ項目レベルのスキーマ構築作業に関わるものである。

- ・ メディエータ作成作業の一部として、シソーラス、辞書などを用いてデータ項目の同一性判定を行うとあるが、既存の研究の参照に止まっており、筆者らの方法に比べて同等あるいは効率的な方法が実現されているとは考え難い。
- ・ ラッパ機能の一部として値の表現形式の異種性解消のための変換機能を含むとあるが、変換機能作成方法や事前の変換機能の品揃えに関する記述は無い。筆者らの方式(事前に標準の変換関数を提供する方式)に比べて同等あるいは容易にラッパを作成可能とは考え難い。

また、普遍関係を用いないスキーマ構築の研究においては、要素データベースのスキーマの共通データモデルへの変換作業と、実体間の関連特定作業と、関連を特定された実体を基にテーブル設計を行う作業とを行う必要があるという問題があると考えられる。

### 6.2 データ項目間の関連の特定

データ項目間の関連の特定を完全に自動化する方式は見当たらない。汎用的な知識を用いて関連の特定を支援する方式(特定を半自動化する方式)の研究を中心



に関連研究を説明する。

Yuら<sup>22)</sup>は概念階層を用いたデータ項目間の関連特定の支援方法を提案している。概念階層は特定の業務に依存しない汎用的な概念の階層(上位がより抽象的な概念を表し、下位がより具体的な概念を表す)として構築される。データ項目は、この概念階層を用いて以下のように特徴付けられる。この特徴づけは、概念階層の利用者が手作業で行うことを前提にしている。

- ・ システム中の概念数を  $n$  とすると、 $n$  次元ベクトルで特徴付けられる。
- ・ ベクトルの各次元の値は対応する概念がデータ項目にあてはまるならば 1、あてはまらないならば 0 とする。

データ項目の近似度は、対応するベクトル間の近似度で計算される。近似度の大きいデータ項目が関連を有するデータ項目の候補として提示される。利用者はそれらの候補の中から関連付けを行う他のデータ項目を手作業で決定する。

この方式の問題点は二つある。一つは、概念階層の構築・維持管理に関するものである。業務非依存という性質を維持しつつ、大企業の大規模データベースを連携する際にように実体数、データ項目数が多い実際の現場において適用可能な概念階層の構築方法が不明である。もう一つは、データ項目毎の  $n$  次元ベクトルの作成に関するものである。前述したように基本的には手作業で作成することを前提としており、実体数、データ項目数が多い実際の現場においては多大な作業量が必要になると考える。半自動化の方式を提案しているが、方式の説明が詳細度に欠けており、その方式によってどの程度稼働が削減されるのか不明である。実用に耐えうるレベルの削減が可能とは考え難い。

Brightらの研究<sup>23)</sup>は、名称の類似度に基づくデータ項目間の関連特定の支援方法を提案している。既存の汎用のシソーラスを基に用語のネットワークを構成する。ネットワークのリンクには用語間の関連の種類(類義語、上位語、下位語)に依存して重みが付けられている。名称間の類似度は、ネットワーク上での重み付きの距離として表される。距離が小さいほど類似度が大きいことを表す。Brightらの類似実体・データ項目の検出処理においては、距離が閾値以下の用語が一つ見つかりと処理が終了するが、距離が閾値以下の用語の集合を、関連を有する実体・データ項目の候補として

提示できるように機能拡張することは容易であると考えられる。

この方式の問題点は、適用業務分野向けにネットワークを拡張しようとする際の方法が不明であり、実体数、データ項目数が多い実際の現場においては多大な作業量が必要になると思われることと、名称が複合語(複数の単語の組み合わせからなる語。例えば、「新規顧客番号」は新規、顧客、番号という三つの単語の組み合わせからなる複合語である)である場合の類似度の判定方法が不明なことである。

Colletら<sup>24)</sup>はCarnotと呼ばれるスキーマ統合ツールを提案している。ツールの中心となるのは、CyCと呼ばれる知識ベースである。約50000の実体と関連とが格納されている。CyCは規範となる大局的スキーマとして機能する。

スキーマ統合の最初のフェーズにおいて、要素データベースのスキーマがCyCに対応付けされる。要素データベースの方が豊富な実体或いはデータ項目を有している場合は、CyCを拡張する。

実体・データ項目の対応付けの基本機構は、名称或いは名称の同義語との文字列一致である。名称或いは名称の同義語と文字列一致した実体・データ項目の候補の中から利用者が適切な候補を選択する。

この方式の問題点は、CyCの拡張頻度や拡張容易性が不明であり、実体数が多い実際の現場においてはCyCの維持管理稼働が大きくなると思われることと、名称が複合語である場合の対応付け方法が不明なことである。

また、概念階層、用語のネットワーク、知識ベースなどの汎用的な知識を用いることをしない関連特定の研究においては、汎用的な知識を用いる場合と同程度の効率で類似データ項目を分類するのは困難であるという問題があると考えられる。

## 7. 評価

### 7.1 評価方針

以下の方針で評価を行う。

まず、DBSENA使用時、未使用時のそれぞれの場合でスキーマの試作を行い、スキーマ構築稼働を評価する。次いで、構築容易化のために提供した機能の十分性を評価する。

7.2 試作

複数の既存社内データベースシステムを基に、DBSENA利用時、未使用時のそれぞれの場合でスキーマの試作を行い、スキーマ構築稼動を評価する。

既存社内データベースシステムの諸元を以下に示す。

- ・ データベース数：5
- ・ 総テーブル数：25
- ・ 総データ項目数：627

DBSENA未使用時の構築方法を中心に、試作におけるスキーマの具体的な構築方法を簡単に説明する。

(1) スキーマ収集

DBSENA未使用時は各要素データベースのスキーマをエクスポートツールなどを用いてエクスポートし、そのデータをスキーマ入力・編集ツールを用いて担当者が入力を行う。

DBSENA使用時・未使用時とも、スキーマ以外に存在する設計情報(設計ドキュメント中の、データ項目の自然言語での説明など)に関しては、スキーマ入力・編集ツールを用いて担当者が入力を行う。

(2) スキーマ間関連特定

DBSENA未使用時は担当者が自ら考案した以下のヒューリスティックスを用いて、類似データ項目の分類を行った。

- ・ 18のカテゴリ(例：電話、日付、住所)に項目を分類する。
- ・ 各カテゴリ内では、担当者の判断で類似項目のグルーピングを行う。

DBSENA使用時・未使用時とも、データ項目の各グループ内での項目間の関連判定は、ドメインの値に基づき担当者が行う。ドメインの値からでは不明な場合は、業務担当者へのヒアリングにより実施する。

なお、用語辞書の維持作業は、用語辞書の導入に伴い必要になった作業であることから、稼動の評価においては、DBSENA使用時のデータ項目の関連特定作業の一部と扱うこととする。

(3) 統合スキーマ作成

DBSENA使用時・未使用時とも、同義データ項目間の変換関数の設定などを行う。必要ならば変換関数の作成を行う。

変換関数提供による効果を明確化するため、DBSENA未使用時は変換関数をすべて新規作成要とした。新規作成要の関数の数と種類は、スキーマ間関連特定で、同義だが変換要と判明したデータ項目の組み合わせから導く。

また、普遍関係の採用による効果を明確化するため、DBSENA未使用時はテーブル設計を行う。筆者らが属す

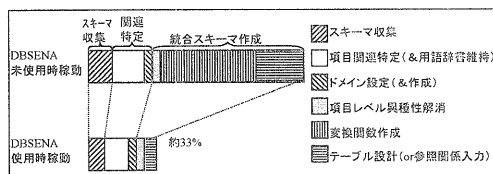


図6 スキーマの構築稼動の比較

Fig.6 Comparison of man-months required to construct a federated database schema.

る研究グループで開発した、データベース設計法<sup>13) 14)</sup>を用いてテーブル設計を行う。

なお、参照関係の投入作業は、普遍関係の導入に伴い必要となった作業であり、テーブル設計とトレードオフの関係にある作業と考えられる。従って、稼動の評価においては、DBSENA使用時の参照関係の投入稼動はDBSENA未使用時のテーブル設計稼動と比較することとする。

7.3 スキーマ構築稼動の評価

DBSENA未使用時に比べると、スキーマ構築稼動が大幅に削減され、約1/3になった(図6)。

提案方式の4つの特徴、即ち①普遍関係を利用可能とし、②用語辞書を用いた類似データ項目分類支援機能を設け、③項目間の変換用の関数を予め標準関数として設け、④要素データベースのスキーマ自動収集機能を設けることが稼動削減に効果のあることが実証された。特に、普遍関係の利用可能化と変換関数の提供による効果が大きい。

スキーマ構築の作業フェーズ毎に削減効果の分析を行う。

(1) スキーマ収集

スキーマ自動収集機能の導入により、DBSENA未使用時に比べスキーマ収集の稼動を約1/3削減できた。要素データベースのスキーマの外部に存在する設計情報が多々ありDBSENA利用時も手入力稼動を大きくは削減できない現状においては、約1/3の削減は十分大きな削減であると考ええる。

(2) スキーマ間関連特定

用語辞書を用いた類似データ項目分類支援機能の導入により、DBSENA未使用時に比べスキーマ間関連特定の稼動を約1/4削減できた。

他の特徴に比べると相対的に削減効果は小さいが、約1/4の削減は有意義な削減であると考ええる。なお、削減効果が小さかったのは以下の要因による。

- ・ DBSENA未使用時の担当者が効率的なヒューリスティックスを考案した。
- ・ 収集した情報(スキーマ情報と設計ドキュメント)

だけでは同義性の判断に必要な情報が不足しており、不足情報を得るために業務担当者へのヒアリングが必要となった要素データベースシステムがいくつかあり、そのヒアリング稼動が大きかった。これは、DBSENA使用時・未使用時に共通的なオーバーヘッドであるが、使用時の効果を相対的に小さくする方向に働く。

### (3) 統合スキーマ生成

DBSENA使用により、DBSENA未使用時に比べ統合スキーマ生成の稼動を約8割削減できた。

関数の作成稼動が、DBSENA未使用時の稼動の半分近くを占め、標準関数の導入による稼動削減効果は大きいと言える。10種類の関数を新たに作成した。主なものを以下に示す。

- ・ 電話番号、年号などのハイフンの削除/挿入
- ・ 金額の単位変換
- ・ 日付の書式変換

また、普遍関係の利用可能化の効果も大きいと言える。DBSENA未使用時のテーブル設計の稼動と比べると、DBSENA使用時の参照関係入力稼動は約1/4であった。

### 7.4 構築容易化のために用意した機能の十分性

#### (1) 用語辞書

新サービスの名称など約10語の追加が必要になった。業務担当者との確認をとりつつ、容易に、用語の登録を行うことができた。

今後も新製品・新サービスの追加や業務の進め方の変更などに伴い、用語は漸増し、それに伴い辞書への登録が必要となるが、スキーマのデータ管理者が業務担当者との連携を密にとることにより、用語の登録はそれほど困難でないと考えられる。

#### (2) データ項目値変換関数

全ての変換に必要な関数は、予め提供された関数あるいは関数の組み合わせにより記述できた。

今後も必要に応じて追加する予定であるが、DBSTREAMの使用においてもさほど追加の必要が生じなかった経験をも踏まえ、追加の必要が生じる可能性は少ないと考える。

#### (3) スキーマ自動収集機能

既存データベースのスキーマは、全てスキーマ自動収集機能で収集できた。

今後も必要に応じて機能を追加する予定であるが、ビジネスで使用される関係データベース管理システムの種類が増えるとは考え難いことと、既存の関係データベース管理システムのスキーマ構造が変わることは稀であると考えられることから、機能追加の生じる可能性は少ないと考える。

## 8. 終わりに

連邦データベースシステムにおけるスキーマの構築の容易化を狙いとして、新たなスキーマ構築方式を提案した。その特徴は、①普遍関係を利用者インタフェースとして利用可能とし、②用語辞書を用いた類似データ項目の分類支援機能を設け、③データ項目間の値変換用の関数を標準関数として設け、さらに④要素データベースのスキーマ自動収集機能を設けることである。

また、該方式の一実現形態として、連邦データベース検索システムDBSENAを試作した。

実際にDBSENAを用いてスキーマの構築を行い、スキーマの構築稼動を大きく削減できることを示した。特に、変換関数の提供と、普遍関係の利用可能化による効果が大きいことを示した。

今後は以下のような課題に取り組むことを考えている。

- ・ ビジネスの現場において徐々に普及しつつある関係データベース以外の情報源(オブジェクト指向データベース、全文検索エンジン、半構造化文書など)の検索を可能にする。
- ・ データウェアハウスなどにも適用分野を広げるために言語能力を拡張する(集約関数の導入など)。
- ・ 更新を可能にする。

## 参 考 文 献

- 1) Baker, R. H. : EXTRANETS, McGraw-Hill, New York(1997).
- 2) 財団法人データベース振興センター編: データベース白書 1998, 財団法人データベース振興センター, 東京(1998).
- 3) Navathe, S. and Savasere, A.: A Schema Integration Facility Using Object-Oriented Data Model. in 4), pp.105-128.
- 4) Bukhres, O. A. and Elmagarmid, A. K. ed.: Object-Oriented Multidatabases Systems: A Solution for Advanced Applications. Prentice-Hall, New Jersey (1996).
- 5) Bouguettaya, A., Benatallah, B. and Elmagarmid, A.: An Overview of Multidatabase Systems: Past and Present. in 6), pp.1-32.
- 6) Elmagarmid, A., Rusinkiewicz, M., and Sheth, A. ed. : Management of Heterogeneous and Autonomous Database Systems. Morgan Kaufman Publishers, San Francisco, CA (1998).
- 7) Ram, S. and Ramesh, V. : Schema Integration: Past, Present, and Future. in 6), pp.119-156.
- 8) Gotthard, W., Lockemann, P. and Neufeld, A.: System Guided View Integration for Object-Oriented Databases. *IEEE Trans. on Knowledge and Data Engineering*, Vol.4, No.1, pp.1-22 (1992).

- 9) Bouzeghoub, M. and Comyn-Wattiau, I.: View Integration by Semantic Unification and Transformation of Data Structures. *Proc. Int. Conf. on Entity-Relationship Approach (ER'90)*, pp.413-430 (1990).
- 10) Kim, W. and J. Seo: Classifying Schematic and Data Heterogeneity in Multidatabase Systems, *Computer*, Vol.24, No.12, pp.12-18 (1993).
- 11) 池田哲夫, 伊土誠一, 石垣昭一郎, 村田達彦: データ流通プラットフォームシステム: DB-STREAM, 情報処理学会論文誌, Vol.38, No.12, pp. 2552-2565(1997).
- 12) Ullman, J. D.: Principles of Database and Knowledge-Base Systems: The New Technologies. Vol.2, W H Freeman & Co. (1989).
- 13) NTT 情報通信研究所(池田哲夫, 川下満, 坂田哲夫, 関根純, 中川優, 村田達彦): データベース概念設計第2版, 阿部出版, 東京(1996).
- 14) NTT 情報通信研究所(大久保成隆, 川下満, 関根純, 中川優, 町原宏毅, 村田達彦): データベース論理設計第2版, 阿部出版, 東京(1996).
- 15) 関根純, 川下満, 町原宏毅, 中川優: 体系的な DB 構築のための用語辞書を用いたデータ標準化手法, 情報処理学会論文誌, Vol.30, No.3, pp.457-467(1993).
- 16) Durell, R. W.: データ資源管理, 日経マグローヒル(1987).
- 17) Darwen, H. and Date, C. J.: A Guide to the SQL Standard : A User's Guide to the Standard Database Language SQL, Addison-Wesley(1997).
- 18) Zhao, J. L., Segev, A. and Chatterjee, A.: A Universal Relation Approach to Federated Database Management. *International Conference on Data Engineering (ICDE)*, pp.261-270 (1995).
- 19) 日本サン・マイクロシステムズ株式会社: Java プログラミング JDBC Java Database Connectivity, サイエンス社(1997).
- 20) Reck, C. and Hillebrand, G.: Implementing a Universal Relation Interface Using Access Scripts with Binding Patterns. Technical Report 1996-40, Universitat Karlsruhe(1996).
- 21) Reck, C. and Konig-Ries, B.: An Architecture for Transparent Access to Semantically Heterogeneous Information Sources. *Proc. 1st Cooperative Information Agents (CIA) 1997*, Springer-Verlag, Berlin, pp. 260-71 (1997).
- 22) Yu, C. T., Sun, W., Dao, S. and Keirse, D.: Determining Relationships among Attributes for Interoperability of Multi-Database Systems, *International Workshop on Interoperability in Multidatabase Systems (RIDE-IMS91)*, Kyoto, pp251-257(1991).
- 23) Bright, M. and Hurson, A. : Linguistic Support for Semantic Identification and Interpretation in Multidatabases. *International Workshop on Interoperability in Multidatabase Systems (RIDE-IMS91)*, Kyoto, pp.306-313 (1991).
- 24) Collet, C., Huhns, M. N. and Shen, W.: Resource Integration Using a Large Knowledge Base in Carnot. *Computer*, vol.24, No.12, pp. 55-62 (1991).

(平成 11 年 6 月 20 日受付)

(平成 11 年 8 月 3 日採録)

(担当編集委員 中川 優)



池田 哲夫(正会員)

1979年東京大学理学部情報科学科卒業。1981年東京大学大学院理学系研究科情報科学専攻修士課程修了。同年日本電信電話公社(現NTT)電気通信研究所入所。現在、NTTサイバースペース研究所主任研究員。この間、プログラム言語の意味論の研究、データベース管理システムの研究開発などに従事。ACM, IEEE CS 各会員。



鈴木 源吾(正会員)

1988年東北大学理学部数学科卒業。1990年東北大学大学院理学研究科数学専攻修士課程修了。同年日本電信電話株式会社通信網総合研究所入社。現在、NTTサイバースペース研究所研究主任。通信網オペレーション情報モデル化/ビジュアル化、スキーマ統合、マルチデータベースシステムの研究開発などに従事。電子情報通信学会会員。



町原 宏毅(正会員)

1987年慶応義塾大学理工学部数理学科卒業。同年日本電信電話株式会社情報通信研究所入所。現在、NTT持株会社移行本部第二部門担当課長。研究所在籍中、データベース設計、異種情報源情報検索システムの研究開発などに従事。



安田 浩(正会員)

1967年東京大学工学部電子工学科卒業。1972年東京大学大学院工学系研究科博士課程修了。同年日本電信電話公社(現NTT)電気通信研究所入所。以来、画像信号符号化、画像信号処理の研究実用化に従事。1997年理事、情報通信研究所長を最後にNTT退社。そのまま東京大学先端研究所教授として現在に至る。専門はマルチメディア符号化、サービス開発、知的財産権処理など。IEEE Fellow。電子情報通信学会会員。工学博士。