

5. 結論

本章では，研究のまとめ，将来展望，および今後の課題を示す．

5.1. 本研究のまとめ

データベース技術の発展・普及に伴い，既存の複数のデータベースシステムを連携して利用したいという要求が生じている．

ここで，既存の，互いに異種性を有する（即ち，データモデル，質問言語，データ項目の値の表現形式などが同一とは限らない）データベースシステム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，企業アプリケーション統合（EAI：Enterprise Application Integration），データウェアハウス（DWH:Data Warehouse）がある．

連邦データベースシステムとは，複数のデータベースシステムのスキーマを収集・分析し，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を言う．連邦データベースシステムの基本技術には，スキーマ（連邦スキーマ）の構築技術や，複数のトランザクションを矛盾無く並列走行させる技術（トランザクション同期制御技術と呼ぶ）などがある．

EAIとは，データベースシステムのデータとアプリケーションを統合する技術，またはそれらの技術を用いて構築されたシステムを言う．EAIは，統合の対象となるものの種別に応じて4つのタイプに大別される．データレベルのEAIと，アプリケーションインタフェースレベルのEAIと，メソッド（共有および再利用可能なロジックをメソッドと呼ぶ）レベルのEAIと，ユーザインタフェースレベルの技術のEAIとである．それらのうち，異種データベースシステム連携に最も関連の深いものはデータレベルのEAIである．本論文ではデータレベルのEAIをデータ交換システムと呼ぶ．連邦データベースシステムとEAIとの違いは，連邦データベースシステムでは連邦スキーマを介した仮想的な連携を実現するのに対し，EAIではデータベースシステム間で直接的にデータを交換して連携を実現する点である．

データウェアハウスとは，意思決定支援用の大量データを格納したデータベースシステムを言う．データウェアハウスの基本技術には，ソースデータベースシステムの

スキーマを統合してデータウェアハウス向けのスキーマを構築するスキーマ構築技術や、ソースデータベースシステムからデータを抽出・変換・格納するデータロード技術や、データウェアハウスのデータを様々な視点から分析する多次元分析技術などがある。連邦データベースシステムやEAIとの違いは、連邦データベースシステムやEAIは、顧客管理や在庫管理などの組織の基幹業務を効率良く行うためのデータベースシステム（基幹業務系データベースシステムあるいは基幹系データベースシステムと呼ぶ）の連携を実現するものであるのに対し、データウェアハウスは基幹系データベースシステムからデータを抽出して構築されるものである点である。

近年、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、連邦データベースシステムを用いて複数のデータベースシステムを連携して顧客管理、在庫管理などの基幹業務を効率化することや、EAIを用いて複数のデータベースシステムを連携して、データ入力作業の軽減、二重投入の回避、データ品質の向上、企業内を流れるデータの速度の向上などを可能とすることや、既存の基幹系データベースシステムからデータを収集してデータウェアハウスを構築して的確な経営上の意思決定を可能にすることなど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。

既存のデータベースシステムを連携することは一般に難度が高く、多くの稼働を要する作業であると認識されている。難度が高いものとなっている要因は大別して二つある。データベースシステムの自律性(autonomy)と、データベースシステム間に存在する異種性(heterogeneity)とである。自律性とは、個々のデータベースシステムが、複数データベースを連携したシステムとは独立に自システムの設計や実行方法を決定できることを言う。異種性とは、データベースシステム間に存在する様々な差異を言う。自律性を維持しつつ異種性を解消して、既存のデータベースシステムを効率的に連携する技術が求められている。

筆者はこのような状況を踏まえて、異種データベースシステム連携の基本技術の研究を研究課題として設定した。具体的には、ニーズの面からの重要性も高まっており、技術的にも重要と認識されている以下を研究課題として設定し、その解決方法を提案した。

- ・ 連邦データベースシステムにおけるスキーマの構築方式
- ・ 企業アプリケーション統合におけるデータ交換システムの構築方式
- ・ データウェアハウスにおけるオンラインデータロードの高速化方式

以下に研究成果のまとめを述べる。

5.1.1. 研究成果の有効性

まず、有効性について論じる。

(1) 連邦データベースシステムにおけるスキーマの構築方式

連邦データベースシステムのスキーマ（連邦スキーマ）の簡易な構築方式が要求される。

この要求に応えるため、スキーマ構築方式として、以下を特徴とする方式を提案した。

- ・ 連邦スキーマの関係として普遍関係(universal relation)を導入する。
- ・ 用語辞書を用いた類似データ項目分類支援機能を設ける。

提案方式を実装する連邦データベース検索システムDBSENAを作成し、さらにDBSENAを用いて、連邦データベースシステムのスキーマを構築し、提案方式の有効性を評価した。

上記特徴を有しないスキーマ構築方式（実体レベルの構築作業を含み、類似データ項目の分類支援機能を有しない構築方式）を用いて構築した場合に比べて構築稼動が約4割削減されることを示した。

また、上記特徴の採用に伴い、用語辞書の維持管理稼動が必要になるが、その維持管理稼動はさほど大きくないと期待できるが分かった。

以上より、提案する構築方式は、構築簡易化において優れた方式であると言える考える。

(2) 企業アプリケーション統合におけるデータ交換システムの構築方式

様々なデータベースシステム間でのデータ交換を可能にするためには、データベースシステム間のデータ異種性を解消することに加えて、各データベースシステムの自律性を維持する必要がある。

ここで、自律性維持の一つの形態として、交換先のデータベースシステムの既存機能や性能への影響を極小化しつつデータ交換を行う必要があるケースがある。典型的なケースとして、交換先のデータベースシステムがレガシーシステム[Brodie95]であり、そのため既存機能の更改が困難であり、データ交換システムの構築時に、交換先

のデータ入力方法を改造無しに用いることを要求されるケースがある。既存データ入力方法には、下記に挙げる制約を有するものがしばしば見られる。

- ・ 複数関連データの一括入力

特定の関連を有する、可変個の関連データをまとめて入力しなければならないという制約。

この課題に対する解決法を与えた研究は見当たらない。

交換元・交換先データベースシステム間のデータ異種性を解消することに加えて、交換先データベースシステムの自律性を維持すること、特に交換先のデータベースシステムの既存入力方法への影響を極小化しつつデータ交換を行うことを可能にするデータ交換システムの簡易な構築方式が求められる。

この要求に応えるため、先ず、様々な形態のデータ交換に必要な機能をモデル化するデータ交換処理モデルを提案し、次いで該モデルに沿ったデータ交換システムの開発・実行環境を提案した。

データ交換処理モデルは、形式的なデータと機能のモデルの考案と、それに基づく操作機能の導出とを特徴とする。

データ交換システム開発・実行環境は、生成方式を適用すること、即ち、以下を特徴とする。

- ・ 高水準な仕様記述言語（シナリオ言語と呼ぶ）を設ける。
- ・ データ交換処理の要素機能（基本操作機能と項目レベルの値変換機能）を再利用可能なソフトウェア部品（メソッドと呼ぶ）として設ける。

データ交換システム開発・実行環境の提案を実装するデータ交換システム開発・実行環境DB-STREAMを作成し、さらにDB-STREAMを用いて複数データベースシステム間のデータ交換システムを試作し、データ交換処理モデルとデータ交換システム開発・実行環境の有効性を評価した。

交換元・交換先データベースシステム間のデータ異種性を解消し、交換先データベースシステムの自律性を維持するデータ交換システムが構築できた。データ交換システム処理モデルはデータ異種性解消、自律性維持に有効であると言えると考える。

データ交換システム開発・実行環境に関しては、個別プログラミングによって構築した場合に比べて構築稼動が約4割になった。データ交換システム開発・実行環境は構築簡易化に有効であると言えると考える。

(3) データウェアハウスにおけるオンラインデータローダの高速化方式

データウェアハウスの一つの形態にオペレーショナルデータストア（ODS：Operational Data Store）がある。ODS は、遅延の少ない意志決定支援を可能にすることなどを目的に、基幹系データベースシステムと同じ環境に構築されるデータウェアハウスである。

ODSへのデータローダは、しばしばトランザクション技術を用いて構築され、そのアクセス特性は、多量のデータを基幹系データベースシステムから検索し、少量の更新をODSに行うことである。ODSのデータローダを始めとする、多量の検索と少量の更新を行うグローバルトランザクションを大量検索型グローバルトランザクションと呼ぶ。

大量検索型グローバルトランザクションと、少量のデータの検索・更新を行う通常のトランザクションとが共存する環境向けのトランザクション同期制御方式は確立していない。

上記環境向けの高性能なトランザクション同期制御方式が要求される。特に、ミッションクリティカルなものが数多く存在する通常のトランザクションの高性能な実行を可能とする方式が要求される。

併せて、異種性解消方式、即ちローカルなデータベース管理システム間の異種性を簡易に解消して、提案するトランザクション同期制御方式を実現する方式が望まれる。

グローバルトランザクションが検索専用の場合は、同期制御に起因する副作用が小さい（アボートは一切発生せず、他のトランザクションとの間で待ちが発生するのはREADの実行時に、未完了のトランザクションによって作成された版を選択しそのトランザクションの完了を待つ場合のみ）という特長を持つ方式として多版時刻印方式があることに着目し、該方式と同期制御方式の主流である2相ロック方式とをベースに新たな方式を提案した。提案した方式は以下の特徴を有する。

- ・ 通常のトランザクションのREAD／WRITEの制御には2相ロック方式を修正適用する。
- ・ 大量検索型グローバルトランザクションのREADの制御には多版時刻印方式を修正適用する。
- ・ さらに、大量検索型グローバルトランザクションはそのWRITESET（トランザクションでの更新対象テーブルの和）に関して、他のトランザクションと直列に実行させることによって、全トランザクションの直列化可能性を保証する。

また、異種性解消方式として、ローカルなデータベースシステムの質問処理機能および同期制御機能に改造を加え、外付けの機能と協調動作させることによって提案する同期制御方式を実現する方式を提案した。

提案方式と従来方式（２相ロック方式）との比較をシミュレーションによって行い、以下の結果を得た。

- ・ 通常のトランザクションの性能（スループット）は提案方式の方が大幅に（最大約 7 割）良い。但し、大量検索型グローバルトランザクションの性能は提案方式の方が多少（最大約 3 割）悪い。

このことより、提案方式は、通常のトランザクションと大量検索型グローバルトランザクションとが共存する環境において優れた同期制御方式である、特に、ミッションクリティカルなものが多数存在する通常のトランザクションの高性能な実行を可能とする同期制御方式であると言える。と考える。

また、異種性解消方式に関しては、既存機能への改造が、①アルゴリズム的にさほど困難ではなく、②規模はさほど大きくなく、かつ③質問処理機能とトランザクション同期制御機能との残余の機能と独立性の高い（即ち残余の機能への影響の小さい）機能であることを示した。提案する異種性解消方式により簡易に異種性を解消可能であると言える。と考える。

5.1.2. 研究成果の一般性

次に、研究成果の一般性について論じる。

(1) 連邦データベースシステムにおけるスキーマの構築方式

本研究の一般性は、普遍関係の有効な範囲に関わる。

普遍関係に関しては、全てのデータ項目が単一の関係（普遍関係）に属するようになるため、データ（データ項目）の意味、データ間の関連などがわかりにくくなり、アプリケーションの構築が却って困難になるという批判がある[Codd88]。データ項目数がいくつの場合からアプリケーションの構築が困難という問題が顕在化するのかに関しては研究者間の合意はないが、筆者は、データ項目名に関して特に条件を付けない場合、数十項目を越えるとアプリケーションの構築が困難という問題が顕在化するものと考ええる。

以下、本研究の有効範囲が広いことを示すため、データ項目の集合がある条件を満たすならばデータ項目の数が 200 程度までは提案方式が有効となる。と言えること (a) と、その条件を満たす場合が多々あると言えること (b) とを示す。

(a) 社内のデータベースシステム設計・構築経験者 4 名にヒアリングを行った結果、彼らの設計・構築経験から判断するに以下の普遍関係の有効性に関する見解は成立する

との見解を得た。

- ・ 連邦スキーマのデータ項目の集合が、もし以下で定義する「理解容易性条件」を満たしかつデータ項目の数が約200を越えないならば、各データ項目の意味は容易に理解可能であり、従ってデータベース検索・更新文の作成にあたってどのデータ項目を用いればよいかの判断が容易であり、このことより普遍関係を用いるのと通常の関係を用いるのとではほぼ同程度の稼動でアプリケーションを構築可能である。

ここで、「理解容易性条件」とは以下の2条件であると定義する。

- ・ 個々の用語（データ項目名を構成する用語）が、理解容易であり、かつ意味のぶれが少ない。
- ・ 各データ項目名がそれらの用語を基に、一定の命名規則（例：修飾語*+主要語+区分語）に基づき、命名されていること。

(b)また、データ項目名の集合が理解容易性条件を満たすように構成できる場合は多々あると考える。

企業が情報資源管理の考え方（情報を企業経営資源の一つとみなして、その完全性を維持し、有効利用を図る考え方。情報資源管理の方法論の一部に用語の統制やデータ項目名の標準化などが含まれる。）[堀内 96]に基づき、業務に用いる用語の統制及びデータ項目名命名規則の制定を行っているならば、その企業において提案方式に基づき連邦スキーマを構築する際に、統制された用語を用いることにより理解容易性条件の第1の条件が保証され、制定された命名規則を用いることにより第2の条件が保証され、従って、理解容易性条件を満たすように、データ項目名の集合を構成することが可能となる。

ここで、IRM 研究会（Information Resource Management 研究会。日本における情報資源管理の研究会。主に企業の実務担当者が会員になっている会。）[堀内 96]の活動などにみられるように、情報資源管理の考え方に基づき、業務に用いる用語の統制及びデータ項目名命名規則の制定を行っている企業が多々あることが知られている[中村 87] [林 95]。

以上をまとめると、提案方式は以下に示す広い適用領域を有するものと考ええる。

- ・ 企業が情報資源管理を実施しており、スキーマのデータ項目名の集合を各データ項目の意味が容易に理解可能なように構築できる場合の、総データ項目数が約2

00程度までのスキーマの構築。

- ・ (スキーマのデータ項目名の集合を各データ項目の意味が容易に理解可能なように構築できるか否かに関わらず), 総データ項目数が数十程度までのスキーマの構築。

(2) 企業アプリケーション統合におけるデータ交換システムの構築方式

本研究が有効となる明確なケースは、以下である。

- ① 交換先の既存データ入力方法への影響極小化のために交換ユニット（複数の関連するレコードの集まり）が必要になるデータ交換システムの構築。

ところで、一般に、歴史のある中・大規模の会社におけるデータベースシステムの構築は以下の特性を持つと言われている[Brodie93]。

- ・ 多数のデータベースシステムを有する。
- ・ 各業務分野毎に独自の設計指針に基づいて設計されたデータベースシステムの割合が多い。全社共通的な指針に基づいて設計されたものはさほど多くない。

上記の特性を持つということは、多くのデータベースシステム間に多様なデータ異種性が発生している可能性が高いことを示し、また、既存データベースシステムを含む情報システム群の段階的更改などにおいて、既存取入力方法への影響極小化などの自律性維持要求が発生する場合が少なからずあることを示すものと考ええる。

従って、一般に歴史のある中・大規模の会社において、①のケースはしばしば生じるものと考ええる。

また、適用対象のデータ交換システムで必要となる機能の多寡に依存して、競合製品・研究と比べて、本研究の成果が最適となりうる（稼動最小で構築できる）ケースが存在する。以下に示す。

- ② 業務として定常的にデータ交換を行うシステムの構築。
- ③ 情報システムの更改において、初期データの構築及び移行を行うシステムの構築
- ④ データウェアハウス構築システムの構築。

- ・ データローダ

- ・ データクリーニング（データクレンジング）システム

具体的社内事例において、上記③、④に分類される事例が数例ずつある。これらの社内事例においては、開発組織が機能と構築容易性を総合的に判断して本研究の成果（DB-STREAM）の適用を決めた。また、適用の結果、データ交換システムの構築が簡

易化されたことと、3.7.1 節で今後の課題として述べた機能を除き新たな機能要求はなかったことが確認されている。このことは、本研究の成果の有効性を裏付けるものであり、本研究の成果が②～④のケースで最適になる場合が少なからぬ割合であることを示すものとする。

以上をまとめると、本研究の成果は一定の適用領域を有するものとする。

(3) データウェアハウスにおけるオンラインデータロードの高速化方式

性能評価結果より、提案方式は、通常のトランザクションと大量検索型グローバルトランザクションが共存する環境において、双方を高性能で走行させる方式であること、特に、ミッションクリティカルなものが多数存在する通常のトランザクションの高性能な実行を可能とする同期制御方式であることが示された。

また、データベースシステムの設計・構築の経験または設計・構築の支援の経験のある者（3名）にヒアリングを行い、以下の見解を得た。

データウェアハウスの重要性は広く認識されつつあり、社内外で ODS が構築されるケースは今後増えるものとする。

それらのケースのうち、遅延の少ないデータロードを可能にすることを狙いとして、データロードをトランザクション技術を用いて実現するケースは少なからぬ割合で存在するものとする。

従って、通常のトランザクションと大量検索型グローバルトランザクションが共存する環境はしばしば生じうると考える。

以上をまとめると、提案方式は一定の適用領域を有するものとする。

5.1.3. 研究成果の貢献

以上の各研究課題の研究成果は、それぞれ着実に現実のものとなって世の中に貢献してきている。

(1) 連邦データベースシステムにおけるスキーマの構築方式

提案したスキーマ構築方式を実装する、連邦データベース検索システム DBSENA は、NTT 社内のデジタルライブラリシステムに採用され、複数の既存図書情報システムを

跨る検索を可能にした。また、DBSENAは、デジタルライブラリ以外にもNTT社内（グループ企業を含む）のいくつかのシステムに採用され、複数データベースシステムからの検索を可能にしている。それらのシステムにおいては、スキーマ構築が簡易化され、用語辞書の維持管理は容易であり、データ項目値変換関数とスキーマ自動収集機能の追加は不要であったことが確認されている。これは、提案方式の有効性を一層強く裏付けるものと考ええる。

(2) 企業アプリケーション統合におけるデータ交換システムの構築方式

提案したデータ交換システム開発・実行環境の実装DB-STREAMは、NTT内の設備データ交換システムに採用され、設計系のシステムで作成されたデータを自動的に監視系のシステムに交換することを可能にし、併せて監視系のシステムにおいて従来のデータ入力方法を改造することなく利用することを可能にした。また、DB-STREAMは、設備データ交換システム以外にもNTT社内のいくつかのシステムに採用され、複数データベースシステム間のデータ交換を可能にしている。これらの適用事例においては、データ交換システムの構築が簡易化されたことと、3.7.1節で今後の課題として述べた機能を除き新たな機能要求はなかったことが確認されている。これは、提案方式の有効性を一層強く裏付けるものと考ええる。

5.2. 将来展望 -インターネット環境への適用へ向けて-

WWW (World Wide Web) の急速な普及に伴い、WWW 上の情報源 (HTML 文書, XML 文書, プレーンテキスト, Postscript 文書等) をもデータベースシステムと同様に情報源として扱いたいという要求が急速に高まっており、WWW 上の情報源をも連携して使用可能とするシステム (異種情報源連携システムと呼ぶ) の研究が近年盛んに行われている (1.3.5 節参照)。

ここでは、各研究成果のインターネット環境 (異種情報源環境) への適用性について述べる。

(1) 連邦データベースシステムにおけるスキーマの構築方式

インターネット環境への一定の適用性があると考ええる。

異種情報源対応に、データベースと同様にアクセス可能とするためのソフトウェア (一般にラッパと呼ばれる) を作成できれば、スキーマ構築方式は少しの拡張 (情報

源の問合せ回答能力をスキーマで新たに管理する拡張など)で適用可能であり、連邦データベース検索システム DBSENA も少しの拡張(拡張機能例:異種情報源が特定の形式の述語のみ処理する能力を有する場合は、利用者からの問合せに含まれる述語を分解して、異種情報源が処理可能な述語は異種情報源に評価させ、残余の述語は異種情報源検索システムで評価するようにする、質問処理最適化機能)で異種情報源検索システムとできると考える。

実際に、異種情報源検索システムを実現して欲しいとの利用者要望が顕在化しており、その要望に応えるために、HTML 文書および全文検索エンジン向けにラッパを簡易に構築する仕掛けを考案し、DBSENA を拡張して異種情報源検索システムを実現することを始めている[Iizuka00]。

なお、インターネット環境においては、そのオープン性に起因して、研究成果の有効性に以下の制約が生じると考える。

- ・ インターネット環境は、オープンな環境であることから、各情報源の構築者が、実体(相当のもの)やデータ項目(相当のもの)を命名する際に用いる用語の種類およびそれらの用語の意味は、一組織に閉じる場合に比べて、はるかに多様性に富んでいる。従って、実証はしていないものの、用語辞書を用いた類似データ項目分類支援において、①類似データ項目の絞込みの効果が一組織に閉じる場合に比べて小さく、②用語辞書に存在しなかった用語が現れ用語の追加作業が必要となる確率が一組織に閉じる場合に比べて高くなる、と考える。①②は、用語辞書を用いた類似データ項目分類支援機能の有効性を低下させる方向に働く。

(2)企業アプリケーション統合におけるデータ交換システムの構築方式

インターネット環境への一定の適用性があると考ええる。

異種情報源対応に、抽出・反映機能(情報源のデータを授受し交換ユニットの形式に/から変換するための機能)を構築できれば、データ交換システムは適用可能である。

異種情報源の間でデータ交換が必要になるケースは、多々あると考える。さらに、異種情報源間でデータ交換システムが必要になるケースの中には、HTML におけるフォーム入力インタフェースのように、データベースのレコードを単純に入力するのではなく、関連する複数のレコードを組み立てて入力することが必要なケース、すなわち筆者が提案するデータ交換システムが有用であるケースが多々あると考える。

抽出・反映機能を簡易に構築する方法に関しては、情報ブローカシステムにおけるラッパを簡易に構築する方法の研究(TSIMMIS[Garcia-Molina97], DISCO[Tomasic96])

など)を元に考案可能と考える。

(3) データウェアハウスにおけるオンラインデータローダの高速化方式

インターネット環境への適用性は、原理的には有るが、大きくないと考える。

異種情報源にもし2相ロック方式及び2相コミット方式を実装する同期制御機能が備わっていれば、各情報源のオリジナルな同期制御機能および質問処理機能に小規模な改造を施しかつグローバルな同期制御機能および質問処理機能を付加することにより、提案する同期制御方式の実現は可能である。

しかしながら、データベースシステム以外の情報源が2相ロック方式及び2相コミット方式を実装する同期制御機能を備えていることは稀であり、現実的な適用性は大きくないと考える。

5.3. 課題

最後に今後の課題を挙げる。

(1) 連邦データベースシステムにおけるスキーマの構築方式

将来展望(前節)で述べた、インターネット環境(異種情報源環境)へ向けての拡張が課題になると考える。

HTML文書、全文検索エンジン向けの拡張は始めたところであるが、それ以外の情報源、例えば、XML文書などとも簡易に連携できることが強く要求されるようになると予想する。

具体的には、以下が課題になると考える。

(a) 異種情報源対応のラッパの簡易な構築機能の確立

異種情報源をデータベースと同様にアクセス可能とするためのソフトウェア(ラッパ)の簡易な構築方法の確立が課題となる。

(b) スキーマ構築方式の拡張

異種情報源の問合せ回答能力をスキーマで新たに管理するなどの拡張が課題となる。

(c) 連邦データベース検索システムの異種情報源検索システムへの拡張

異種情報源が特定の形式の述語のみ処理する能力を有する場合は、利用者からの問合せに含まれる述語を分解して、異種情報源が処理可能な述語は異種情報源に評価

させ、残余の述語は異種情報源検索システムで評価するように質問最適化機能を拡張する、などの拡張が課題となる。

(2) 企業アプリケーション統合におけるデータ交換システムの構築方式

将来展望（前節）および機能の十分性の考察（3.7.1節）を踏まえ、以下が課題になると考える。

(a) 抽出・反映機能の簡易な構築方法の確立

各種データベースシステムおよび各種異種情報源対応に、抽出・反映機能の簡易な構築方法の確立が課題となる。

(b) データウェアハウスのデータローダ向け機能の拡張

データの集約機能（例：複数の値から平均値、最大値などを求める機能）などを追加することが課題となる。

謝辞

本論文をまとめる機会を頂くと共に、懇切なるご指導と格別のご配慮、温かい励ましを賜りました東京大学先端科学技術研究センターの安田浩教授に心より感謝申し上げます。懇切なるご指導と温かいご支援を賜りました東京大学生産技術研究所の喜連川優教授に心より感謝申し上げます。また、有益なご助言と温かいご支援を賜りました東京大学工学部の石塚満教授、東京大学先端科学技術研究センターの堀浩一教授、東京大学情報基盤センターの山口和紀教授に心より感謝申し上げます。

本論文は、筆者が日本電信電話株式会社（NTT）の研究所において行ってきた研究、および東京大学先端科学技術研究センターとの共同研究プロジェクトとして行った研究をまとめたものであり、この間に多くの方々のご指導、ご支援、ご協力を賜りましたことを厚く御礼申し上げます。

本研究を行う機会を与えて頂き、またご指導を頂いた情報ベースプロジェクト（前データベース研究部）の伊土誠一部長（現情報流通基盤総合研究所）、石垣昭一郎部長（現NTTドコモ）、福岡秀樹プロジェクトマネージャ（現NTTソフトウェア）、吉田孝プロジェクトマネージャ、田中豪グループリーダー（現NTTデータ）、柴垣斉グループリーダー（現サイバーソリューション研究所）、武田英昭グループリーダー（現サイバーソリューション研究所）、櫻井紀彦グループリーダー、佐藤哲司グループリーダーおよびサイバーコミュニケーション総合研究所の寺中勝美企画部長に深く感謝いたします。

本研究は、多くの方々との共同研究や討論を経ています。関根純博士（現情報流通基盤総合研究所）、町原浩毅主任研究員（現第二部門）、川下満主任研究員（現NTTソフトウェア）、星野隆研究主任（現NTT西日本）、鈴木源吾研究主任とは、スキーマ構築方式の研究を進めるにあたって多くの討論を重ね、示唆を頂きました。伊土誠一部長（現情報流通基盤総合研究所）、石垣昭一郎部長（現NTTドコモ）、石黒正典グループリーダー、村田達彦主幹研究員（現NTTドコモ）、岸本義一主任研究員（現NTTソフトウェア）、星野隆研究主任（現NTT西日本）とは、データ交換システムの構築方式の研究を進めるにあたって多くの討論を重ね、示唆を頂きました。寺中勝美サイバーコミュニケーション総合研究所企画部長には、オンラインデータロード高速化方式の研究を進めるにあたって多くの示唆を頂き、さらに大阪大学工学部の西尾章

治郎教授と討論する機会を設けて頂きました。これらの方々に深く感謝します。

スキーマ構築方式の研究の一部の内容は東京大学先端科学技術研究センターとNTT研究所との共同研究の一環として行われたものです。共同研究においては、東京大学先端科学技術研究センターの安田浩教授にご指導、ご協力頂きました。また、NTTの関根純博士（現情報流通基盤総合研究所）、町原浩毅主任研究員（現第二部門）、川下満主任研究員（現NTTソフトウェア）、星野隆研究主任（現NTT西日本）にご協力頂きました。これらの方々に深く感謝します。

また、この他にも多くの方々のご指導、ご支援、ご協力を頂きましたことを厚くお礼申し上げます。

参考文献

- [Adali96] S. Adali, K. S. Candan, Y. Papakonstantinou and V. S. Subrahmanian, "Query Caching and Optimization in Distributed Mediator Systems," Proc. ACM SIGMOD, pp.137-148, 1996.
- [Arens93] Y. Arens, C. Y. Chee, C. Hsu and C. A. Knoblock, "Retrieving and Integrating Data from Multiple Information Sources," Int. Journal of Intelligent and Cooperative Information Systems, Vol. 2, No. 2, pp. 127-158, 1993.
- [Baker97] R. H. Baker, EXTRANETS, McGraw-Hill, New York, 1997.
- [Barquin97] R. Barquin and H. Edelstein, Building, Using, and Managing the Data Warehouse, Prentice Hall, 1997.
- [Bayardo97] R. J. Bayardo Jr., B. Bohrer, R. S. Brice, A. Cichocki, J. Fowler, A. Helal, V. Kashyap, T. Ksiezyk, G. Martin, M. H. Nodine, M. Rashid, M. Rusinkiewicz, R. Shea, C. Unnikrishnan, A. Unruh and D. Woelk, "InfoSleuth: Semantic Integration of Information in Open and Dynamic Environments," Proc. ACM SIGMOD, pp.195-206, 1997.
- [Bernstein80] P. A. Bernstein, et al., "Concurrency Control in a System for Distributed Databases (SDD-1)," ACM Trans. Database Syst., Vol. 5, No. 1, pp.18-51, 1980.
- [Bober92] P. M. Bober and M. J. Carey "On Mixing Queries and Transactions via Multiversion Locking," Int. Conf. on Data Engineering, pp.535-545, 1992.
- [Bontempo98] C. Bontempo and G. Zagelow, "The IBM Data Warehouse Architecture," Comm. ACM, Vol.41, No.9, pp.38-48, 1998.
- [Bouguettaya98] A. Bouguettaya, B. Benatallah and A. Elmagarmid "An Overview of Multidatabase Systems: Past and Present," in [Elmagarmid98], pp.1-32.
- [Breitbart91] Y. Breitbart, D. Georgakopoulos, M. Rusinkiewicz and A. Silberschatz, "On Rigorous Transaction Scheduling," IEEE Trans. Software Engineering, vol.17, No.9, pp.954-960, 1991.
- [Bright91] M. Bright and A. Hurson, "Linguistic Support for Semantic Identification and Interpretation in Multidatabases," Int.l Workshop on Interoperability in Multidatabase Systems (RIDE-IMS91), Kyoto, pp.306-313,

1991.

- [Bright92] M. W. Bright and A. R. Hurson and S. H. Pakzad, "A Taxonomy and Current Issues in Multidatabase Systems," *Computer*, Vol.25, No.3, pp. 50-60, March 1992.
- [Brodie93] M. L. Brodie, "The Promise of distributed computing and the challenges of legacy information systems," *Proceedings of the IFIP TC2/WG2.6 Conference on Semantics of Interoperable Database Systems*, pp.1-31, 1993.
- [Brodie95] M. Brodie and M. Stonebraker, *Migrating Legacy Systems : Gateways, Interfaces & the Incremental Approach*, Morgan Kaufmann, 1995.
- [Brzezinski84] Z. Brzezinski, J. R. Getta, J. Rybnik and W. Stepniewski, "UNIBASE - An Integrated Access to Databases," *Proc. 10th VLDB Conf.*, pp.388-396, 1984.
- [Bukhres96a] O. A. Bukhres and A. K. Elmagarmid ed., *Object-Oriented Multidatabases Systems: A Solution for Advanced Applications*, Prentice-Hall, New Jersey, 1996.
- [Bukhres96b] O. A. Bukhres, A. K. Elmagarmid, F. F. Gherfal, X. Liu, K. Baker and T. Schaller, "The Integration of Database Systems," in [Bukhres96a], pp.37-56.
- [Burch97] G. Burch, "Building High Data Quality Into Your Data Warehouse," in [Barquin97], 1997.
- [Carey84] M. Carey and M. Stonebraker, "The Performance of Concurrency Control Algorithms for Database Management Systems," *Proc. 10th VLDB Conf.*, pp.107-118, 1984.
- [Cattell94] R. G. G. Cattell, *Object Data Management, Object-Oriented and Extended Relational Database Systems*, Revised Edition, Addison-Wesley, 1994.
- [Ceri85] S. Ceri and G. Pelagatti, *Distributed Databases : Principles and Systems*, McGraw-Hill, New York, 1985.
- [Chan83] A. Chan, et al., "Overview of an ADA Compatible Distributed Database Manager," *Proc. ACM SIGMOD*, San Jose, CA, pp.228-237, 1983.
- [Chan85] A. Chan and R. Gray "Implementing Distributed Read-Only Transactions," *IEEE Trans. Software Engineering*, Vol.11, No.2, pp.205-212, 1985.
- [Chawathe94] S. S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y.

- Papakonstantinou, J. D. Ullman and J. Widom, "The TSIMMIS Project: Integration of Heterogeneous Information Sources," Proc. of the 10th Meeting of the Information Processing Society of Japan, pp.7-18, 1994.
- [Chen76] Peter P. Chen, "The Entity-Relationship Model - Toward a Unified View of Data," ACM Trans. Database Syst., Vol.1, No.1, pp.9-36, 1976.
- [Codd70] E. F. Codd, "A Relational Model of Data for Large Shared Data Banks," Commun. ACM, Vol.13, No.6, pp.377-387, 1970.
- [Codd88] E. F. Codd, "Comments, with reply, on 'The universal-relation data model for logical independence' by M. Vardi," IEEE Software, Vol.5, No.4, p. 4-6, 1988.
- [Collet91] C. Collet, M. N. Huhns and W. Shen, "Resource Integration Using a Large Knowledge Base in Carnot," Computer, vol.24, No.12, pp.55-62, 1991.
- [Czejdo92] B. D. Czejdo and M. C. Taylor, "Integration of object-oriented programming languages and database systems in KOPERNIK," Data & Knowledge Engineering, Vol.7, pp.271-298, 1991.
- [DDL73] "CODASYL Data Description Language Journal of Development, June 1973," NBS Handbook, 113, 1974.
- [Darwen97] H. Darwen and C. J. Date, A Guide to the SQL Standard : A User's Guide to the Standard Database Language SQL, Addison-Wesley, 1997.
- [Dayal84] U. Dayal and H. Hwang, "View Definition and Generalization for Database Integration in a Multidatabase System," IEEE Trans. Software Engineering, vol.10, No.6, pp. 628-645, 1984.
- [Du92] W. Du, R. Krishnamurthy and M. Shan, "Query Optimization in a Heterogeneous DBMS," Proc. 18th VLDB Conf., pp.277-291, 1992.
- [Durell87] R. W. Durell, データ資源管理, 日経マグローヒル, 1987.
- [Elmagarmid90] A. K. Elmagarmid, Y. Leu, W. Litwin and M. Rusinkiewicz, "A Multidatabase Transaction Model for InterBase," Proc. 16th VLDB Conf., pp.507-518, 1990.
- [Elmagarmid98] A. Elmagarmid, M. Rusinkiewicz and A. Sheth ed., Management of Heterogeneous and Autonomous Database Systems, Morgan Kaufman Publishers, San Francisco, 1998.
- [Elmasri85] Ramez Elmasri, J. Weeldreyer and A. Hevner, "The Category Concept:

- An Extension to the Entity-Relationship Model," *Data & Knowledge Engineering*, Vol.1, pp.75-116, 1985.
- [Eswaren76] K. Eswaren, et al., "On the Notions of Consistency and Predicate Locks in a Relational Database Systems," *Comm. ACM*, Vol. 19, No. 11, pp.624-633, 1976.
- [Garcia-Molina87] H. Garcia-Molina and K. Salem, "Sagas," 249-259, *Proc. ACM SIGMOD*, San Francisco, CA, pp.249-259, 1987.
- [Garcia-Molina97] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. D. Ullman, V. Vassalos and J. Widom, "The TSIMMIS Approach to Mediation: Data Models and Languages," *Journal of Intelligent Information Systems*, Vol.8, No.2, pp.117-132, 1997.
- [Garcia-Solaco96] M. Garcia-Solaco, F. Saltor and M. Castellanos, "Semantic Heterogeneity in Multidatabase Systems," in [Bukhres96], pp.129-202.
- [Geller92] J. Geller, A. Mehta, Y. Perl, E. Neuhold and A. Sheth, "Algorithms for Structural Schema Integration," *Int. Conf. on Systems Integration*, pp.604-614, 1992.
- [Georgakopoulos91] D. Georgakopoulos, M. Rusinkiewicz and A. P. Sheth, "On Serializability of Multidatabase Transactions Through Forced Local Conflicts," *Int. Conf. on Data Engineering*, pp.314-323, 1991.
- [Gray79] Jim Gray, "Notes on Data Base Operating Systems," in *Advanced Course: Operating Systems*, eds. M. J. Flynn et al., *Lecture Notes in Computer Science*, Vol. 60, pp.393-481, Springer, 1979,
- [Haerder83] T. Haerder and A. Reuter, "Principles of Transaction-Oriented Database Recovery," *Comput. Surv.*, Vol.15, No.4, pp.287-317, 1983.
- [Hammer91] K. Hammer, "Data Conversion: A Database-Driven Approach," *Database Programming and Design*, April, pp.42-47, 1991.
- [Hammer97] K. Hammer, "Migrating Data from Legacy Systems: Challenges and Solutions," in [Barquin97], 1997.
- [Hayne90] S. Hayne and S. Ram, "Multi-user view integration system (MUVIS): an expert system for view integration," *Int. Conf. on Data Engineering*, pp. 402-409, 1990.
- [Hernandez98] M. A. Hernandez and S. J. Stolfo, "Real-world Data is Dirty: Data

- Cleansing and The Merge/Purge Problem,” *Data Mining and Knowledge Discovery*, Vol.2, No.1, pp.9-37, 1998.
- [Hinrichs99] H. Hinrichs and K. Panienski, “Experiences with Knowledge-Based Data Cleansing at the Epidemiological Cancer Registry of Lower-Saxony,” 5th Biannual German Conference on Knowledge-Based Systems, pp. 218-225, 1999.
- [Howells87] D. I. Howells, N. J. Fiddian and W. A. Gray, “A Source-to-Source Meta-Translation System for Relational Query Languages,” *Proc. 13th VLDB Conf.*, pp.227-234, 1987.
- [IBM00a] <http://www.software.ibm.com/data/datajoiner/>
- [IBM00b] <http://www.software.ibm.com/data/dpropr/pdfs/99factsheet.pdf>
- [Inmon96] W. H. Inmon, *Building the Data Warehouse*, 2nd ed., John Wiley & Sons, 1996.
- [Inmon96] W. H. Inmon, “Building the Data Warehouse 2nd edition,” John Wiley & Sons, 1996.
- [Inmon99] W. H. Inmon, “Building the Operational Data Store 2nd edition,” John Wiley & Sons, 1999.
- [JIS87] JIS X 3004:1987, データベース言語 NDL, 1987.
- [Jarke00] M. Jarke, M. Lenzerini, Y. Vassiliou and P. Vassiliadis, “Fundamentals of Data Warehouses,” Springer Verlag, 2000.
- [Java97] 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.
- [Katoh85] N. Katoh, T. Ibaraki and T. Kameda, “Cautious Transaction Schedulers with Admission Control,” *ACM Trans. Database Syst.*, Vol.10, No.2, pp.205-229, 1985.
- [Kim93] W. Kim and J. Seo, “Classifying Schematic and Data Heterogeneity in Multidatabase Systems,” *Computer*, Vol.24, No.12, pp.12-18, 1993.
- [Kolovson89] C. P. Kolovson and M. Stonebraker, “Indexing Techniques for Historical Databases,” *Int. Conf. on Data Engineering*, pp.127-137, 1989.
- [Konopski95] D. K. and O. Shmueli, “W3QS: A query system for the World Wide Web,” *Proc. 21st VLDB Conf.*, pp.54-65, 1995.
- [Kung81] H. T. Kung and J. T. Robinson, “On Optimistic Methods for Concurrency Control,” *ACM Trans. Database Syst.*, Vol.6, No.2, pp.213-226, 1981.

- [Lakshmanan96] Laks V. S. Lakshmanan, F. Sadri, and I. N. Subramanian, "A declarative language for querying and restructuring the Web," Proc. of 6th. Int. Workshop on Research Issues in Data Engineering (RIDE '96), New Orleans, 1996.
- [Larson89] J. A. Larson, S. B. Navathe and R. Elmasri, "A Theory of Attribute Equivalence in Databases with Application to Schema Integration," IEEE Trans. Software Engineering, vol.15, No.4, pp. 449-463, 1989.
- [Levy96] A. Y. Levy, A. Rajaraman and J. J. Ordille, "Querying Heterogeneous Information Sources Using Source Descriptions," Proc. 22nd VLDB Conf., pp.251-262, 1996.
- [Linthicum99] D. S. Linthicum," Enterprise Application Integration," Addison-Wesley, 1999.
- [Litwin90] Witold Litwin, Leo Mark and Nick Roussopoulos, "Interoperability of multiple autonomous databases," Comput. Surv., Vol.22, No.3, pp.267-293, September 1990.
- [Lohman85] G. M. Lohman, C. Mohan, L. M. Haas, D. Daniels, B. G. Lindsay, P. G. Selinger and P. F. Wilms, "Query Processing in R*," in Query Processing in Database Systems, ed. W. Kim, D. S. Reiner and D. S. Batory, pp.31-47, Springer, 1985.
- [Lomet90] D. B. Lomet and B. Salzberg, "The Performance of a Multiversion Access Method," Proc. ACM SIGMOD, pp.353-363, 1990
- [Markowitz93] V. M. Markowitz and A. Shoshani, "Object Queries over Relational Databases: Language, Implementation, and Applications," Int. Conf. on Data Engineering, pp.71-80, 1993.
- [McGee77] W. C. McGee, "The Information Management System IMS/VS Part I through Part IV," IBM Syst. J. Vol.16, No.2, pp. 84-168, 1977.
- [Mehrotra92] S. Mehrotra, R. Rastogi, Y. Breitbart, H. F. Korth and A. Silberschatz, "Ensuring Transaction Atomicity in Multidatabase Systems," Proc. ACM Symp. on Principles of Database Systems, pp.164-175, 1992.
- [Mendelzon96] A. O. Mendelzon, G. A. Mihaila and T. Milo, "Querying the World Wide Web," Int. Conf. on Parallel and Distributed Information Systems, pp.80-91, 1996.
- [Motro81] A. Motro and P. Buneman, "Constructing Superviews," Proc. ACM SIGMOD,

pp.56-64, 1981.

[NTT96a] 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.

[NTT96b] 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.

[Navathe86] S. B. Navathe and R. Elmasri and J. A. Larson, "Integrating User Views in Database Design," *Computer*, Vol.19, no.1, January 1986.

[Navathe96] S. Navathe and A. Savasere, "A Schema Integration Facility Using Object-Oriented Data Model," in [Bukhres96a], pp.105-128.

[Obermarck82] R. Obermarck, "Distributed Deadlock Detection Algorithms," *ACM Trans. Database Syst.*, Vol.7, No. 2, pp. 187-208, 1982.

[Ozsu98] M. T. Ozsu and P. Valduriez, *Principles of Distributed Database Systems*, Prentice Hall, 1998.

[PeopleSoft00] http://www.peoplesoft.com/en/us/products/applications/pt/product_content.html

[Pirahesh 90] H. Pirahesh, C. Mohan, J. M. Cheng, T. S. Liu and P. G. Selinger, "Parallelism in Relational Data Base Systems: Architectural Issues and Design Approaches," *Proc. of the 2nd Int. Symposium on Databases in Parallel and Distributed Systems*, pp.4-29, 1990.

[Pu88] C. Pu, G. E. Kaiser and N. C. Hutchinson, "Split-Transactions for Open-Ended Activities," *Proc. 14th VLDB Conf.*, pp.26-37, 1988.

[Ram98] S. Ram and V. Ramesh, "Schema Integration: Past, Present, and Future," in [Elmagarmid98], pp.119-156.

[Ramamritham96] K. Ramamritham and P. K. Chrysanthis. "A Taxonomy of Correctness Criteria in Database Applications," *The International Journal on Very Large Data Bases*, Vol. 5, No. 1, pp.181-293, 1996.

[Reck96] C. Reck and G. Hillebrand, "Implementing a Universal Relation Interface Using Access Scripts with Binding Patterns," *Technical Report 1996-40*, Universitat Karlsruhe, 1996.

[Reck97] C. Reck and B. Konig-Ries, "An architecture for transparent access to semantically heterogeneous information sources," *Proc. 1st Cooperative Information Agents (CIA) 1997*, Springer-Verlag, Berlin, pp.260-71, 1997.

- [Reed83] D.P. Reed, "Implementing Atomic Actions on Decentralized Data," ACM Trans. Computer Systems, Vol. 1, No. 1, pp.3-23, 1983.
- [Rusinkiewicz85] M. Rusinkiewicz and B. D. Czejdo, "Query Transformation in Heterogeneous Distributed Database Systems," Int. Conf. on Distributed Computing System, pp.300-307, 1985.
- [SAP00] <http://www.sap-ag.de/solutions/r3/index.htm>
- [Selinger79] P. G. Selinger, M. M. Astrahan, D. D. Chamberlin, R. A. Lorie and T. G. Price, "Access Path Selection in a Relational Database Management System," Proc. ACM SIGMOD, pp.23-34, 1979.
- [Sheth88] A. P. Sheth, J. A. Larson, A. Cornelio and S. B. Navathe, "A Tool for Integrating Conceptual Schemas and User Views," Int. Conf. on Data Engineering, pp. 176-183, 1988.
- [Sheth90] Amit P. Sheth and James A. Larson, "Federated Databases for Managing Distributed, Heterogeneous, and Autonomous Databases," Comput. Surv., Vol.22, No.3, pp.183-236, September 1990.
- [Sheth92] A. P. Sheth and V. Kashyap, "So Far (Schematically) yet So Near (Semantically)," Proc. of the IFIP WG 2.6 Database Semantics Conference on Interoperable Database Systems (DS-5), pp.283-312, 1992.
- [Shipman81] David W. Shipman, "The Functional Data Model and the Data Language DAPLEX," ACM Trans. Database Syst., Vol.6, No.1, pp.140-173, 1981.
- [Spaccapietra94] S. Spaccapietra and C. Parent, "View Integration: A Step Forward in Solving Structural Conflicts," IEEE Trans. Knowledge and Data Engineering, Vol.6, No.2, pp.258-274, 1994.
- [Stonebraker77] D. Stonebraker and E. Neuhold, "A Distributed Version of INGRES," Proc. 2nd Berkley Work. on Distributed Data Management and Computer Networks, pp.19-36 , 1977.
- [Sybase00a]<http://www.sybase.com:80/products/entcon/dircon.html>
<http://www.sybase.com:80/products/entcon/omni.html>
- [Sybase00b] <http://www.sybase.com:80/products/system11/repssvr.html>
- [Teorey86] T. J. Teorey, D. Yang and J. P. Fry, "A logical design methodology for relational databases using the extended entity-relationship model," Comp. Surv., Vol.18, No.2, pp.197-222, 1986.

- [Tomasic96] A. Tomasic, L. Raschid and P. Valduriez, "Scaling Heterogeneous Databases and the Design of Disco," Proc. of the 16th Int. Conf. on Distributed Computing Systems, pp.449-457, 1996.
- [Tsichritzis78] D. Tsichritzis and A. C. Klug, "The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems," Inf. Syst. vol.3, No.3, 1978.
- [Ullman89] J. D. Ullman, Principles of Database and Knowledge-Base Systems: The New Technologies, Vol.2, W H Freeman & Co., 1989.
- [Urban91] S. D. Urban and J. Wu, "Resolving Semantic Heterogeneity Through the Explicit Representation of Data Model Semantics," ACM SIGMOD Record, Vol.20, No.4, pp.55-58, 1991.
- [Vality00] <http://www.vality.com/>
- [Vardi88] M. Y. Vardi, "The Universal-Relation Data Model for Logic Independence," IEEE Software, Vol.5, No.2, pp.80-85, 1988.
- [Yu91] C. T. Yu, W. Sun, S. Dao and D. Keirse, "Determining Relationships Among Attributes for Interoperability of Multi-Database Systems," Int.l Workshop on Interoperability in Multidatabase Systems(RIDE-IMS91), Kyoto, pp251-257(1991).
- [Zaniolo79] C. Zaniolo, "Design of Relational Views over Network Schemas," Proc. ACM SIGMOD, pp.179-190, 1979.
- [Zhao95] J. L. Zhao, A. Segev and A. Chatterjee, "A Universal Relation Approach to Federated Database Management," Int. Conf. on Data Engineering, pp.261-270, 1995.
- [Zhuge95] Y. Zhuge, H. Garcia-Molina, J. Hammer and J. Widom, "View Maintenance in a Warehousing Environment," Proc. ACM SIGMOD, pp. 316-327, 1995.
- [Zhuge96] Y. Zhuge, H. Garcia-Molina and J. L. Wiener, "The Strobe Algorithms for Multi-Source Warehouse Consistency," Conf. on Parallel and Distributed Information Systems, pp.146-157, 1996.
- [タスク IT 新書編集部 00] タスク IT 新書編集部, "エンタープライズ アプリケーション インテグレーション," タスクシステム・プロモーション, 2000.
- [データベース白書 00]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 2000, 財団法人データベース振興センター, 東京, 2000.
- [磯田 95]磯田定宏, "ソフトウェア再利用," 情報処理ハンドブック, 情報処理学

- 会 (編), pp.784-793, オーム社, 東京, 1995.
- [横山 94]横山雅明, 合田信久, 堀田英一, “ネットワークデータベースの基盤整備,” NTT 技術ジャーナル, Vol.6, No.9, pp.49-53, 1994.
- [関根 89]関根純, 川下満, 鈴木健司, “ネーミング手法と支援ツール,” 信学技報, DE89-4, pp.25-32, 1989.
- [関根 93]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,” 情報学論, Vol.30, No.3, pp.457-467, 1993.
- [宮崎 84]宮崎正弘, “係り受け解析を用いた複合語の自動分割法,” 情報学論, Vol.25, No.6, pp.970-979, 1984.
- [国語研 64]国立国語研究所, “分類語彙集,” 秀英出版, 1964.
- [黒川 93]黒川清, 中川優, 関根純, “複合語解析技術を用いたデータ項目名称の標準化手法,” 情報学論, Vol.34, No.3, pp.447-456, 1993.
- [植村 79]植村俊亮, データベースシステムの基礎, オーム社, 東京, 1979.
- [数学辞典 85]日本数学会編集, 数学辞典第3版, 岩波書店, 東京, 1985.
- [増永 91]増永良文, リレーショナルデータベース入門, サイエンス社, 東京, 1991.
- [中村 87]中村正弘, “先進ユーザがつかんだデータ項目名標準化の糸口,” 日経コンピュータ, 1987.9.14号, pp.105-113, 1987.
- [日経システムプロバイダ 00]“売れ行きチェック EAI ツール,” 日経システムプロバイダ, pp.36-38, 2000.5.12号, 日経 BP 社, 2000.
- [正田 89]正田定幸, 分散データベースシステム入門, オーム社, 東京, 1989.
- [堀内 96]堀内一監修, IRM 研究会編, “データ中心システム分析と設計,” オーム社, 1996.
- [木本 91]木本晴夫, “日本語新聞記事からのキーワード自動抽出と重要度評価,” 信学論, Vol.J74-D-I, No.8, pp.556-566, 1991.
- [林 95]林泰樹, “企業データの分析とデータ管理—実施例に基づく報告—,” 情報学シンポジウム, pp.129-137, 1995.

関連発表

1. 論文・国際会議

[池田 88]池田哲夫, “分散データベースの同期制御の一方式,” 情報学論, Vol.29, No.7 pp.669-675, 1988.

[池田 97]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,” 情報学論, Vol.38, No.12, pp.2552-2565, 1997.

[池田 99]池田哲夫, 町原宏毅, 安田浩, “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,” 情報学論: データベース, Vol.40, No.SIG 8(TOD 4), pp.29-40, 1999.

[Iizuka00] Yuici Iizuka, Mitsuaki Tunakawa, Shin-ichiro Seo, Tetsuo Ikeda, “An Approach to Integration of Web Information Source Search and Web Information Retrieval,” ACM Symposium on Applied Computing, Como, Italy, pp.289-293, 2000.

2. 研究会, 全国大会

[池田 90]池田哲夫, 岸本義一, 田中一敏, “情報流通のための DB ゲートウェイ機能の一考察,” 情報処理学会第 41 回全国大会, 1990.

[池田 93]池田哲夫, 石垣昭一郎, 村田達彦, “DB 流通の基本方式について,” 情報処理学会第 46 回全国大会, 1993.

[奥村 95]奥村昌和, 池田哲夫, 岸本義一, “DB 流通におけるデータ抽出方式に関する考察,” 情報処理学会第 51 回全国大会, 1995.

[奥村 96]奥村昌和, 池田哲夫, 岸本義一, “DB 流通における更新差分データ抽出方式に関する検討,” 情報処理学会第 52 回全国大会, 1996.

[河合 94]河合宏保, 池田哲夫, 岸本義一, “DB 再構築における DB 流通システムの適用法に関する考察,” 情報処理学会第 49 回全国大会, 1994.

[川下 89]川下満, 関根純, 池田哲夫, “大規模 DB 向け正規化手法,” 情報処理学会第 39 回全国大会, 1989.

[塩原 94]塩原寿子, 池田哲夫, 村田達彦, 井戸正幸, “DB 流通システムとその適用,” 情報処研報(マルチメディア通信と分散処理), 94-DPS-64, pp.51-56, 1994.

[関根 89]関根純, 川下満, 池田哲夫, “一貫性制約を考慮した DB 概念設計法,” 情報処理学会第 39 回全国大会, 1989.

付録 1. 用語辞書構築方法

1. まえがき

本付録では、類似データ項目分類支援機能で用いる用語辞書の構築方法を説明する[黒川 93].

筆者の所属する組織においては、以前にデータ項目名称の標準化の研究を行った[関根 89]. 名称の標準化機能の一つに、新規開発システムのデータ項目名称を統一するための名称チェック機能がある. この名称チェック機能のための核となる技術の一つに、データ項目名称に使用される用語を納める用語辞書の構築技術がある. データ項目名称の標準化研究の一環として、用語辞書の構築方法の研究及び用語辞書の構築が行われた.

筆者の提案する類似データ項目分類支援機能では、データ項目名称の標準化研究において構築された用語辞書を用いている.

ここでは、データ項目名称の標準化研究の一環として提案された用語辞書の構築方法を説明する.

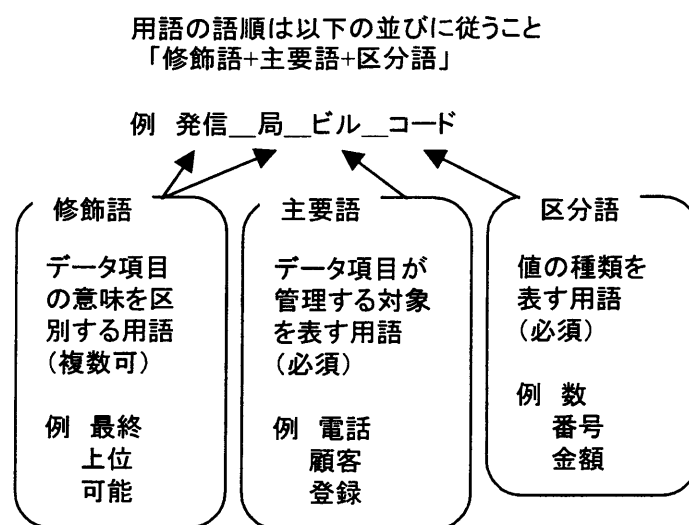
以下、先ず、データ項目名称の命名規則と用語辞書の構成について説明し、次いで用語辞書の初期構築方法について説明し、最後に用語辞書の維持管理方法について説明する.

2. データ項目名称の命名規則

データ項目名称は、用語の組合せ、すなわち、複合語として構成されている. この用語の組合せに構文的制約を加えることで、誰にも分かりやすい名称にするのが命名規則の考え方である. 日本語の特徴を考慮し Durell の命名規則に制約を加えた提案[中村 87]とその例を付図 1 に示す.

例えば、データ項目の名称として「ビル」を登録するものとする. しかし、この名称だけでは「ビル」の「名前」を表すものか「コード」を表すものか分からない. そこで、命名規則ではデータ項目名がどのような値かを表す用語（この例の場合「コード」）、および、その用語の前方に、管理する対象が何かを表す用語（この例の場合「ビル」）が必須であることを規定する. このとき、管理する対象を表す用語を主要語(prime word)、内容を表す用語を区分語(class word)と呼ぶ. さらに、この二つの用語から構

成される名称だけでは、データ項目名称が一意にならない場合、意味を区別する用語である修飾語(modifier word)を任意個、主要語の前に接続させて識別性を高める。



付図1. 命名規則

3. 用語辞書の構成

データ標準化で用いられる用語辞書には、以下のような情報が管理されている。

(1)用語

データ項目名称を構成する最小単位であり、その一語で意味を持つものとする。

(2)標準用語

同義だが表記の異なる用語のうち一つを公式に通用するものとして定めたものを標準用語と呼ぶ。例えば、「ユーザー」、「ユーザ」、「USR」などの用語は、「ユーザ」を標準用語にすると決める。

(3)用語種別

Durell の命名規則における、区分語、主要語、修飾語のような用語の役割の概念が、データ項目名称に出現している用語でどのように使われているかを分析した結果、これら三つの概念は一つに固定されるわけではなく、状況に応じて変わりうる事が分かった。これを分類すると、用語は「修主区」「修主」「修」の3種類に分類・整理できた(付表1)。用語種別が「修主区」の用語は修飾語、主要語、区分語として、また、「修主」は修飾語、主要語として用いることができ、「修」の用語は修飾語としてのみ

使用可能であることが分かった。例を付図2に示す。用語辞書では、用語毎に上記3つのうちのどの分類に属するかを管理することにした。

付表1 用語の役割と用語種別の関係

用語種別 \ 役割	修飾語	主要語	区分語
修主区	○	○	○
修主	○	○	
修	○		

(a)用語種別が「修主区」の例(番号)

電話__番号 登録__番号__区分 番号__更新__日付
 ↑ ↑ ↑ ↑ ↑ ↑ ↑
 主要語 区分語 修飾語 主要語 区分語 修飾語 主要語 区分語

(b)用語種別が「修主」の例(支店)

支店__コード 支店__社員__数
 ↑ ↑ ↑ ↑ ↑
 主要語 区分語 修飾語 主要語 区分語

(c)用語種別が「修」の例(上位)

上位__局__番号
 ↑ ↑ ↑
 修飾語 主要語 区分語

付図2. 用語の役割の変化

(4)代表標準用語

用語辞書においては、類義の標準用語をグルーピングしている。そのうちの代表となる標準用語を代表標準用語と呼ぶ。全ての標準用語は、いずれかのグルーピングに属する。また、各々のグルーピングは、排他的である（標準用語が複数のグルーピングに属することはない。）

代表標準用語と対応する標準用語の例を示す。

・ 代表標準用語：名称

対応する標準用語：氏名，線名，俗称，地名，通称，品目，名，名前，品名，呼称，略称，名称

・ 代表標準用語：時間

対応する標準用語：時，時刻，時分秒，秒，分

- ・ 代表標準用語：住所
対応する標準用語：屋号，号，番地，地，区
- ・ 代表標準用語：電話
対応する標準用語：ダイヤル，データテレホン，ビジネスホン，プッシュホン，
電話機，電話帳，テレホン，ホームテレホン

4. 用語辞書初期構築方法

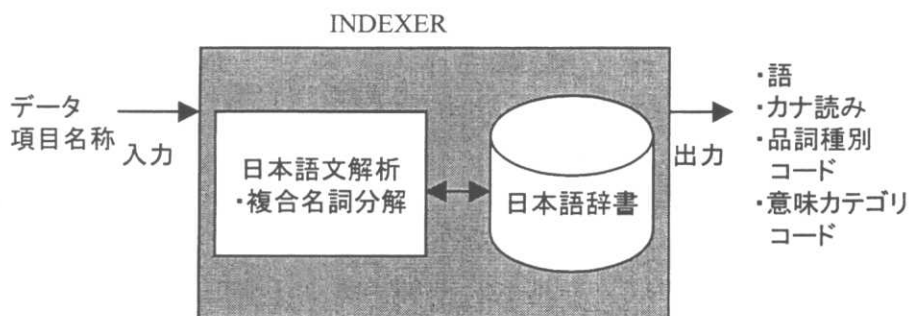
用語辞書の初期構築方法について説明する。

具体的には，用語の区切り方法と，用語種別の判定方法と，類義の標準用語のグルーピングの方法について説明する。

データ項目名は複数の品詞から構成された複合語であり，その複合語解析の研究については，自然言語の研究開発から種々の成果がでてきており，わが社でもその成果の一つとして，キーワード自動抽出システム INDEXER[木本 91]がある．そこで，用語の区切り，用語種別の判定のために，INDEXER の処理結果を利用することを考えた．

4.1 INDEXER の利用

INDEXER は、大量の文献情報データベースから必要な情報を早く引き出すために必要な、キーワードを自動抽出するシステムとして開発された。この INDEXER が提供するいくつかの機能のうち、複合名詞分割機能を用語区切りと用語種別の判定に用いることを考えた。INDEXER の入出力情報と処理の流れを付図 3 に示す。INDEXER にデータ項目名称を入力すると、INDEXER 内部で形態素解析および意味解析が行われる。複合語であるデータ項目名称は、複合名詞解析ルーチンにおいて、語と語の前後関係を解析する係り受け解析法[宮崎 84]によって、語の区切りおよび語と語の相互関係の同定処理が行われ、その結果、語、カナ読み、品詞種別、意味カテゴリが出力される。ここで、意味カテゴリとは、世の中の事象や概念を抽象的概念からより具体的概念へ階層化した分類体系である。INDEXER では、分類語彙表[国語研 64]を基本に、それをさらに詳細化した 2715 種類の意味カテゴリを管理しており、同定処理の結果として各単語ごとに第一候補から第三候補までの三つの意味カテゴリの候補を出力する。INDEXER の出力結果を用語区切り判定法と用語種別の判定法に用いることにした。



付図3. INDEXERの処理概要

4.2 用語区切り判定方法

命名規則で用いる用語は、その一語で、事物、事象、値など、意味のあるものとしている。一方、INDEXER の出力である語には、名詞、動詞、形容詞など様々な品詞種別があり、接辞である「最」、「的」のように、他の品詞と結合して意味あるものになる語もある。すなわち、一つ以上の語の組合せで用語は成り立っていると考えられる。そこで、このような用語と語の関係を調べるため、先ず、予備的に社内の 12 の既存デ

データベースシステムのデータ項目名称を INDEXER で処理して語ごとに分解し、その結果から用語がどのような語から構成されるかを調べた。その結果、

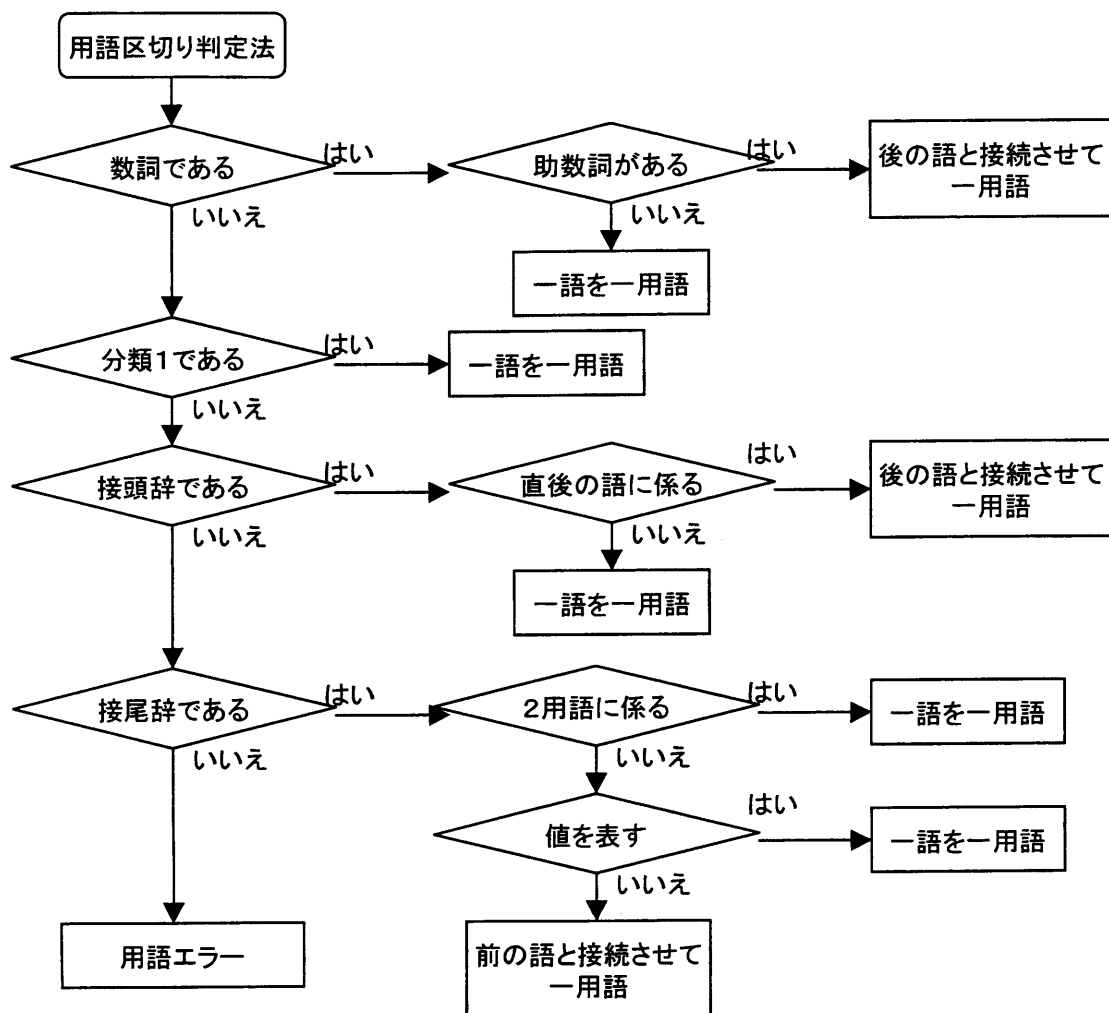
- ・ データ項目に使用される語の品詞種別は限定される。
- ・ 一つ以上の語からなる用語の品詞種別組合せパターンは限定される。

ことが分かった。その構成を付表2に示す。付表2に基づく、用語区切りの概略フローを付図4に示す。

付表2 用語の品詞構成と用語種別

用語種別	修主区	修主	修
用語(品詞構成)			
一般名詞	番号, 金額	回線, 電話	上位, 下位
サ変動詞型名詞	—	受注, 契約	—
形容動詞型名詞	—	—	固有, 可能
動詞転成型名詞	—	受付, 繰越	—
時詞	月, 日	休日	午後
数詞	—	—	複数, 400
固有名詞	—	—	NTT
他動詞	—	引込, 打込	—
副詞	—	—	以上, 随時
接頭辞	—	—	新, 旧
接尾辞	値, 数	—	別
接頭辞+一般名詞	—	上支線	最下位
接頭辞+サ変動詞型名詞	—	再送付	無応答
接頭辞+形容動詞型名詞	—	—	不完全
接頭辞+動詞動詞転成型名詞	—	再割付	未貸付
接頭辞+他動詞	—	再打込	未取消
一般名詞+接尾辞	—	事業部	自動的
サ変動詞型名詞+接尾辞	—	契約者	修正用
形容動詞型名詞+接尾辞	—	専用化	—
動詞転成型名詞+接尾辞	—	受付後	繰越用
他動詞+接尾辞	—	振込済	引込用
数詞+助数詞	—	—	1. 5M

(注) —印は生起しないケースを表す。

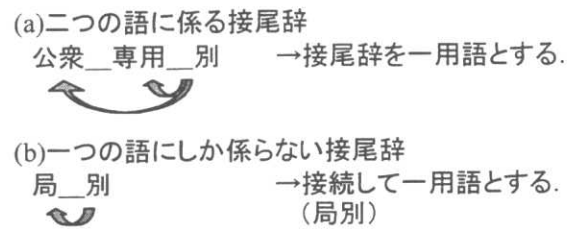


分類1:一般名詞, サ変動詞型名詞, 形容動詞型名詞, 動詞転成型名詞, 時詞, 固有名詞, 他動詞, 副詞

付図4 用語区切りの判定フロー

このフローは以下の考え方に基づく。

- (1) INDEXER により，一般名詞，サ変動詞型名詞，形容動詞型名詞，動詞転成型名詞，時詞，数詞，固有名詞，他動詞，副詞に分類される語は，用語として意味が明確であると考えられるので，一語で一用語とする。ここで，形容動詞型名詞は，「固有，可能」などのように語尾に「だ」や「な」という付属詞を付加すると形容動詞になる品詞種別であり，サ変動詞型名詞は「受注，契約」などのように語尾に「する」という付属語を接続して動詞となる品詞種別であり，動詞転成型名詞とは，「扱い」のように語尾を変化させるか，あるいは「受付」のように語尾に付属語「る」を接続することにより動詞に変化する品詞種別である。
- (2) 接頭辞に分類される語は，単独では意味が明確でないため，直後の語に接続させて一用語とする。直後に現われる語としては，一般名詞，サ変動詞型名詞，形容動詞型名詞，動詞転成型名詞，他動詞が許される。但し，次の場合は，例外とし，独立の用語とする。
 - (a) 「新電話番号」に示す接頭辞「新」は，「電話」に係るのではなく「番号」にかかっている。従って，「新」の直後にある語と接続させて用語「新電話」とすると意味が変化する。そこで，このような接頭辞は，独立な用語とする。
- (3) 接尾辞に分類される語は，単独では意味が明確でないため，直前の語に接続させて一用語とする。直前に現われる語としては，一般名詞，サ変動詞型名詞，形容動詞型名詞，動詞転成型名詞，他動詞が許される。但し，次の2つの場合は，例外とし，独立の用語とする。
 - (a) 「数」，「名」のように区分語になるものは，単独で用語とする。
 - (b) 二つ以上の語にかかる接尾辞（付図5）は，二つの語（公衆，専用）が並列におかれており，一方に接続させると全体の意味が異なるため，独立させることとする。
- (4) 数詞の後に来る助数詞は，それだけでは意味が明確でないため，数詞と助数詞とを接続させて一用語とする。
- (5) 判定不能となる語は，もともとデータ項目名称として使用できない品詞（助動詞，副助詞）であるので，エラーとしている。

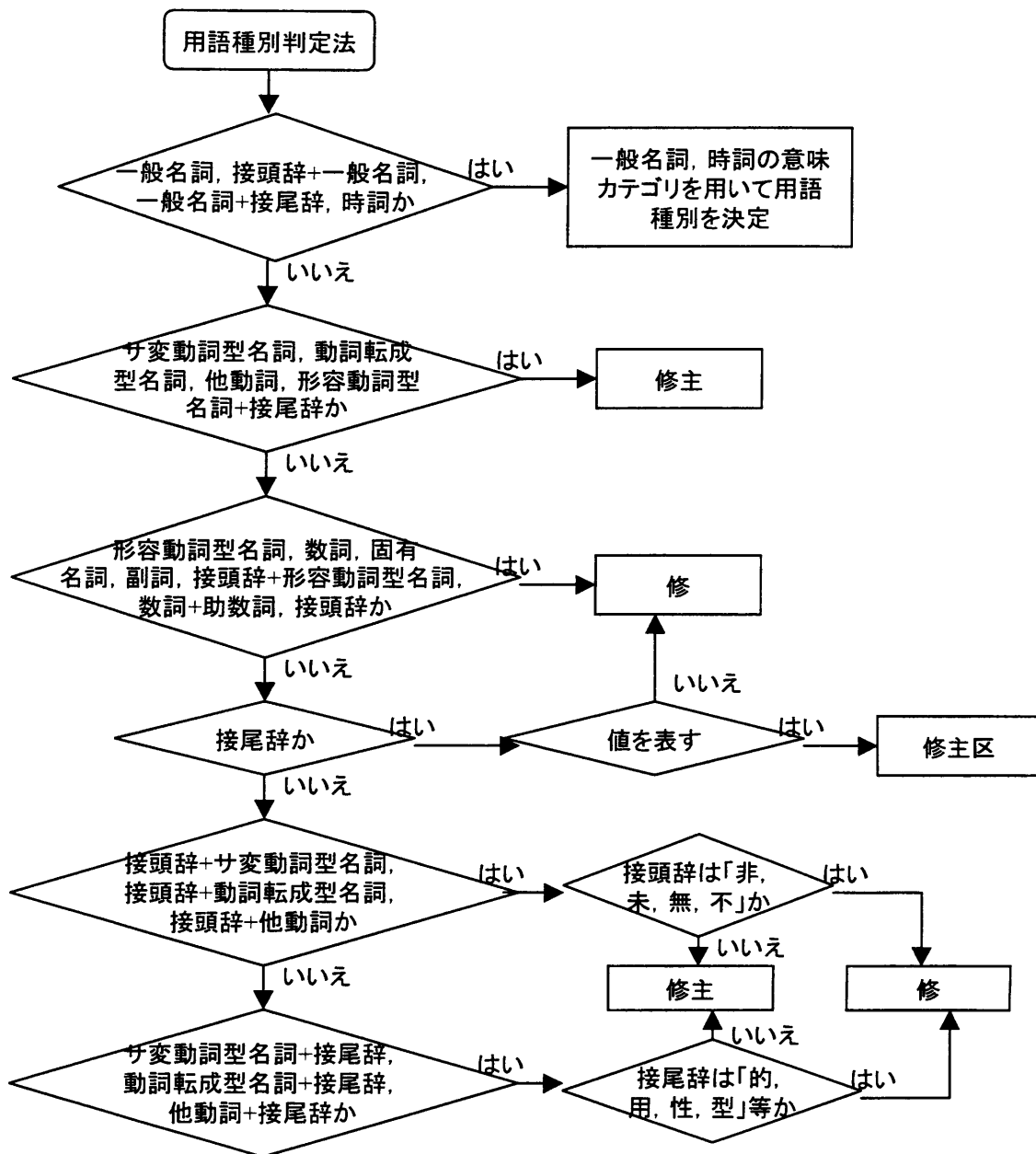


付図5. 接尾辞の係り受け

4.3 用語種別判定方法

文法上の役割によって分類したものが品詞種別であることから，用語種別の判定に品詞種別を用いることを検討した．社内既存データベースシステムの用語について，品詞種別と用語種別の関係を調べた結果，品詞種別だけで用語種別を判定できるものと，そうでないものがあることが分かった．

一般名詞と時詞，接尾辞を除く他の用語は，品詞種別だけで用語種別が決まることが分かった．品詞種別と用語種別の分類・整理結果を付表2に示す．この結果を用いた用語種別判定の概略フローを付図6に示す．



付図6 用語種別判定のフロー

一般名詞，時詞，接尾辞以外の用語種別決定方法の考え方を以下に示す。

- (1) INDEXER で分類される語で，サ変動詞型名詞，動詞転成型名詞，他動詞の品詞種別を持つものは，イベントの事象，または，状態を表すと考えられ，これらはデータベースにおいてはデータ管理の対象になるので，用語種別は「修主」とする。
- (2) 形容動詞型名詞，副詞，固有名詞，数詞，接頭辞，数詞+助数詞は，他の用語に接続してその意味を限定する働きがあると考えられるが，これらは単独でデータベースにおける管理の対象とならないので，用語種別は「修」とする。

なお，固有名詞がデータベースの管理対象となりえないのは以下の理由による。一般に，データベースの管理対象は，その管理対象に属するインスタンスが複数存在することを暗黙の前提とする。しかし固有名詞は，対象世界に唯一存在する管理対象を表すため，この範疇に入らない。例えば，固有名詞「NTT」を主要語とする「NTT 住所」，「NTT 番号」などは，いずれも唯一存在するデータであり，通常はデータベース中に管理されることはない。一方，「NTT」が修飾語になる「NTT 交換機番号」には，インスタンスが複数存在しうるので，このような名称が現われても問題ない。

- (3) 接頭辞，接尾辞が接続した用語の用語種別は，接続された用語の品詞種別，意味カテゴリを引き継いで決定する。但し，以下の接頭辞・接尾辞が付加されたものについては，例外として次のように用語種別を設定する。

(a) 接尾辞として「用」，「型」，「性」あるいは「的」などが接続した用語の用語種別は，その接尾辞が接続することにより形容動詞型名詞に品詞種別が変化するため，「修」とする。

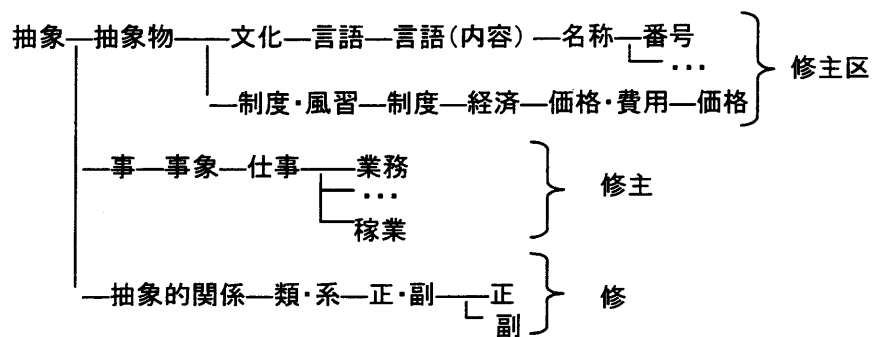
(b) 接頭辞として「未」，「不」，「非」あるいは「無」などが接続した用語の用語種別は，その接頭辞が接続することにより形容動詞型名詞に品詞種別が変化するため，「修」とする。

(c) 接尾辞として「化」，「視」を接続した用語は，「する」という付属語を接続すると動詞になるため，サ変動詞型名詞として扱い，用語種別は「修主」とする。

また，接尾辞は種類が少ないので区分語になりうるかを総当りで調べ，それを基に用語種別を決定した。

一方，一般名詞，時詞については，品詞種別だけでは用語種別の判定はできないので，INDEXER が出力する意味カテゴリを用いて，さらに判定を行うことにした。INDEXER が出力する意味カテゴリは 2715 種類あり，すべての意味カテゴリを 3 種類の用語種別に分類・整理した。意味カテゴリと用語種別の対応の例を付図 7 に示す。データの实

体と考えられるカテゴリ（業務，稼業など）は「修主」，値に置き換えることができるカテゴリ（番号，価格など）は「修主区」，抽象的な関係を表すカテゴリ（正，副など）は「修」とした．なお，INDEXER の処理の結果，意味カテゴリのうち，第一候補を用いて用語種別を決定した．



付図7 意味カテゴリと用語種別の関係(抜粋)

4.4 類義の標準用語のグルーピングの方法

用語辞書の構築者が、自らの経験（データベースシステムの設計・構築の経験、あるいは設計・構築の支援の経験）を元に、類義となりうる標準用語のグルーピングを手作業で行った。

グルーピングの作業において、類義語辞書等の外部の支援手段を用いることはなかった。

4.5 初期構築時の入力データ

用語辞書の初期構築においては、社内の基幹19システムのデータ項目名称を基に用語辞書を作成した。当初の用語辞書の用語数は1336用語であった。

5. 用語辞書の維持管理方法

提案し、また実際に社内において採用された、用語辞書の維持管理方法について説明する。

社内において、個別のデータベース設計プロジェクトとは別に、社内の情報資源を管理するデータ管理部門を設け、2者が以下のように役割分担することにより、用語辞書の維持管理を行う。

(1) 個別のデータベース設計プロジェクトの役割

予め、データ標準化支援ツールに、4章で構築された用語辞書（以下、共通用語辞書と呼ぶ）とは独立に、ローカルな用語辞書を管理する機能を設ける。

個別のデータベース設計プロジェクトにおいて、共通用語辞書にない用語が現われた場合は、対応する標準用語、用語種別、代表標準用語の選択を行い、ローカルな用語辞書に登録する。これらは、データ管理者がさらにチェックを行うために、データ管理部門のホストシステムに送信する。

(2) データ管理部門の役割

ローカルな用語辞書をチェックし、問題がある用語については、修正を行った上で共通用語辞書に登録する。また、新たな用語辞書の版をデータベース設計プロジェクトに再配布する。

ローカルな用語辞書のチェックの主な観点を以下に示す。

- ・ 既に同一の用語や標準用語が共通用語辞書に登録されていないか
- ・ 用語に対する標準用語の選択が適切か。また、共通用語辞書と矛盾がないか。
- ・ 標準用語に対する用語種別の選択が適切か。
- ・ 標準用語に対する代表標準用語の選択が適切か。また、共通用語辞書と矛盾がないか。

データ管理部門の担当者（データ管理者）が自らの経験に基づき、チェックおよび修正を行う。チェックおよび修正において、類義語辞書等の外部の支援手段を用いることはしない。