

マルチデータベース日本語インタフェースの試作 — 問合せの生成・実行 —

2D-2

真鍋俊彦, 高橋一重, 住田一男
(株) 東芝 研究開発センター

1. はじめに

データや情報の分散化に伴い、データの位置、形式を意識させないで、手軽に使えるデータベースインタフェースが求められている。自然言語インタフェースもその一つであるが、従来はデータベース管理システムの種類に依存する形で開発されてきた。これに対して、著者らは、関係型データベースや文書データベース、PC/EWS上の各種アプリケーションのデータを同時に扱えるシステムを開発した [1]。

自然言語インタフェースは、入力解析部と、その解析結果から問合せを生成する問合せ生成部の二つに分かれる [2]。問合せ生成部がデータベース管理システムの種類に大きく依存していたため、異種データベースを同時に扱うことはできなかった。

著者らのシステムの問合せ生成部は、各データベース管理システムの検索機能に基づき論理スキーマレベルの問合せを生成する。SQLのような強力な検索機能をもつ言語から、アプリケーションの単純な検索コマンドによる問合せまでを同じ処理で生成する。

2. 問合せ生成部

問合せ生成部の処理の流れを図1に図示する。

概念辞書には、入力解析部の生成した中間コマンド中の語彙と各データベースの論理スキーマ上の表現（属性名、式、検索条件など）との対応関係を記述しておく。この対応関係はデータベース毎に記述する。このことにより、同じ単語でもデータベースによって、その管理システムに合わせた問合せ表現に展開できるようにする。

データ辞書には、データベースとその管理システムとの対応関係と、各管理システムでの機能、例えば、結合演算可能か、集約関数を利用できるかなどを記述しておく。このデータ辞書を用いて、データベース管理システムやアプリケーションの能力に応じた単位に問合せを分解する。

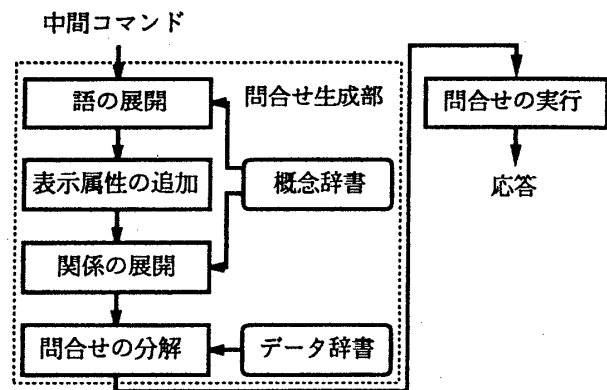


図1: 問合せ生成部の処理の流れ

「鈴木さんの今期の残業時間は」というユーザの質問からは次の中間コマンドが生成される。このコマンドの処理例に基づき、各処理内容を説明する。

```
(a) select 残業時間
      from 残業実績
      where 氏名('鈴木') and 今期;
```

2.1 語の展開

概念辞書を用いて、コマンドを実体・関係 (ER) モデル上の表現に変換する。例えば、コマンド (a) は、次のようになる。

```
(b) select Overtime_Hour
      from Overtime
      where emp->Name = '鈴木'
      and Date in ('1995.04', ..., '1995.09');
```

“emp->Name” は、関係 emp を辿って Overtime 以外のデータの属性 Name を参照することを表現している。その内容を図2に図示する。

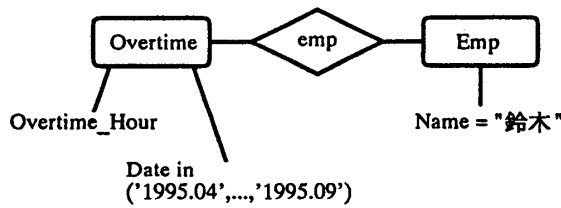


図 2: コマンド (b) でのデータ間の関係

ところで、上記の例では、“今期”を“Date in ('1995.04', ..., '1995.09')”に置き換えている。もし、「鈴木さんの今期提出した報告書のタイトル」という質問がなされていたならば、この場合の“今期”は、報告書を検索するための表現、例えば、“YYMM>='199504'”に置き換える。

2.2 表示属性の追加

コマンド (b) の問合せ結果は、単に 6 つの残業時間であり、これをそのまま表示するだけでは、それぞれが何年何月の残業時間かわからない。このような場合に対応するため、検索条件中で値の不定な属性を select 項目に追加して表示の便宜を図る。例えば、コマンド (b) に対しては、属性 Date の値が一意に決まらないので、それを追加する。

2.3 関係の展開

概念辞書中の関係とデータベース間の結合条件の対応関係に基づいて、コマンド中の関係を結合条件に置き換える。コマンド (b) については、関係 emp を Overtime と Emp の結合条件に置き換え、表示属性の追加と合わせて次のように変換する。

```
(c) select Date, Overtime_Hour
      from Overtime, Emp
      where Name = '鈴木'
            and Date in ('1995.04', ...,
                        '1995.09')
            and Overtime.account =
              Emp.account;
```

2.4 問合せの分解

データ辞書を用いて、実行可能な単位にコマンドを分解する。例えば、コマンド (c) で Overtime と Emp が異なる管理システムに分かれていたら、Emp を検索する部分と Overtime を検索する部分の二つ

に分解する。

また、文献情報が発行年によって異なるデータベースに格納されている場合など、データベースに仮想的な属性 (仮想属性) を定義し、検索条件中のその属性値の指定によってデータベースを選択するようにしている。例えば、コマンド (c) で Date が仮想属性の場合、Overtime を検索する部分を、さらに、'1995.04' から '1995.09' までの値のそれぞれについて別のデータベースを検索するように分解する。

3. 問合せの実行

問合せの実行はデータベースへのアクセスを仲介するゲートウェイプログラムにより行う。ゲートウェイプログラムは独立したプロセスであり、システム本体から前節のコマンドの形式で問合せを受ける。

ゲートウェイプログラムの機能は次のとおり。

- セッションの開始、終了の制御
- データベースの選択の制御 (open, close)
- データベース固有の検索コマンドへの変換、実行

システム本体はこれらの機能を RPC で呼び出して、各データベースにアクセスする。データベース間の結合演算などはシステム本体側で行う。

4. おわりに

データベース管理システムやアプリケーションの検索機能に基づき、各管理システムで実行可能な単位の間合せを生成することで、異種データベースを一括して扱える問合せ生成処理を実現した。このことによりユーザには分散したデータや情報にアクセスしていることを意識させないシステムを構築できた。今後は、問合せの実行時間の予測や、意味的な問合せの最適化機能を検討、組み込んでいきたい。

参考文献

- [1] 住田他: マルチデータベース日本語インタフェースの試作—日本語入力解析処理, 第 51 回情全大 2D-1, 1995.
- [2] Androutsopoulos, I. et al.: Natural Language Interfaces to Databases - An Introduction, Research Paper 709, Dept. of Artificial Intelligent, Edinburgh, 1994.