

日本語によるリレーショナルデータベース

2Q-1

問い合わせ文のSQL変換について

河村 一 樹

日本電子専門学校

1. はじめに

リレーショナルデータベース (RDB) の普及にともない、エンドユーザが直接データを操作したいという要求が生じつつある。そのためには、ユーザインターフェイスのよいRDB問い合わせシステムが必要となる。そこで、現在SQLを用いて、RDBを日本語で検索する「自然言語によるRDB検索のためのフロントエンドプロセッサシステム」について研究している。

本稿では、フロントエンドプロセッサにおけるSQL変換のための解析方法について取り上げる。その中で、日本語を形式化することによって、SQLへの変換が容易となることを提案する。

2. 本システムの構成

本システムの全体構成を、図1に示す。

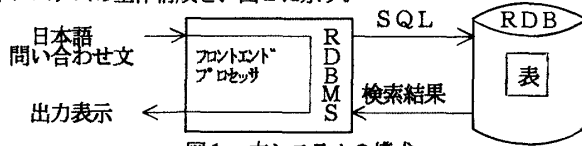


図1 本システムの構成

本システムが有する機能としては、以下の事項があげられる。

- ①. 日本語RDB問い合わせ文の入力。ただし、ユーザインターフェイスの向上を考慮した入力操作になっている必要がある。(→マルチウィンドウ、カーソル選択採用)
- ②. 日本語によるRDB問い合わせ文をSQL文へ変換。(→変換辞書採用)
- ③. ②で生成したSQL文を動的SQL機能によって実行。これによって、RDBの該当する表が検索可能。(→埋め込みSQL文によるモジュール言語採用)
- ④. 検索結果の出力表示。

本稿では、②のSQL変換の解析部について取り上げる。

なお、本システム作成上の前提条件は、以下の通りとする。

- ①. 検索対象となるRDBの表は、DDLで定義 (CREATE) 及び作成 (INSERT) 済みとする。
- ②. RDBの表名、列名、それらの属性といった情報は、データ検索の時に表示提供する。
- ③. 変換できるSQL文は、単一表の検索 (SELECT) のみとする。

3. 日本語問い合わせ文について

日本語によるRDB問い合わせ文をSQL文に変換する場合、2通りの方法がある。1つは、日本語に全く制約を与えない自然言語を用いる方法 (非形式化仕様) と、もう1つとして、形式化した自然言語を用いる方法 (形式化仕様) である。どちらを採用するかで、変換方式が異なる。

・非形式化仕様

自然言語文 → 形態素解析 → 構文解析 → 意味解析 → SQL変換 → SQL文

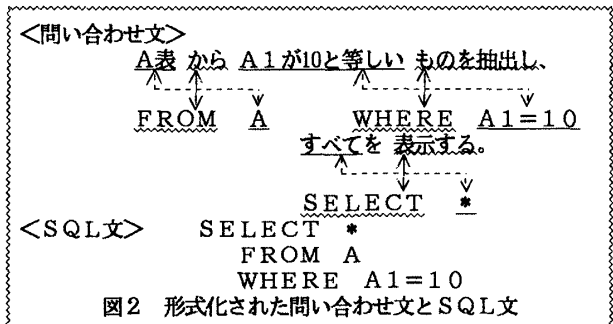
・形式化仕様

形式化された自然言語文 → 構文解析 → SQL変換 → SQL文

形式化仕様を採用することによって、形態素解析や意味解析はほとんど考慮しなくてよい。

形態素解析については、各ウィンドウから形式化された問い合わせ文を構成する個々の単語を、カーソル選択で入力する際、区切り情報を挿入することができる。このため、単語分解ができることになる。

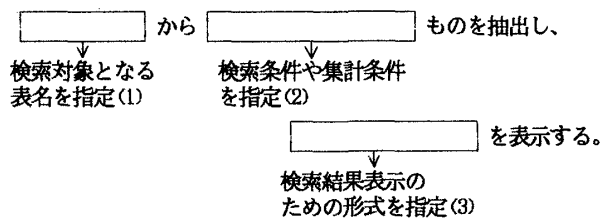
意味解析については、問い合わせ用語自身を形式化していることによって、用語のもつ意味を規定することができる。具体的には、SQL文のコマンドの意味に対応づけて、問い合わせ用語を表現すればよい。また、問い合わせ文の入力順や用語の組合せに制限を与えることによって、構文の意味は規定され、正確になるといえる。このため、意味解析はあえてやる必要がなくなる。例えば、図2のような形式化された問い合わせ文は、SELECT文にそのまま対応できる。



以上より、問い合わせ文に形式化仕様を導入することによって、構文解析の機構が全体的に単純化されることになる。

4. 問い合わせ文のSQL変換

問い合わせ文の構文パターンは、次のように規定する。



この問い合わせ文の各欄は、次のようにSELECT文の各コマンドに対応する。

- (1) ← FROM句
- (2) ← WHERE句、GROUP句 [, HAVING句]
- (3) ← SELECT句、ORDER句

この対応によって、形式化された日本語問い合わせ文を規則的なアルゴリズムによって、SQL文に変換することが可能となる。いくつかの変換事例を以下に示す。ただし、検索対象の表は、図3のものとする。

学生番号	氏名	クラス	国語	英語	数学 (Tuple)
87JN0001	阿部 伊材	A	90	70	80
{	}	}	}	}	}

図3 成績表のレイアウト

(1). 全体検索の場合

問い合わせ文	成績表からすべてのものを抽出し、すべてを表示する。
SQL文	SELECT * FROM 成績表

「すべての」があれば、WHERE句は作成せず。また、「すべて」で * に変換する。

(2). 単一条件検索の場合

問い合わせ文	成績表から学生番号 が 87JN0001 と等しいものを抽出し、国語、英語、数学を表示する。
SQL文	SELECT 国語, 英語, 数学 FROM 成績表 WHERE 学生番号=87JN0001

検索条件では、比較演算子 (=と等しい、!=と等しくない、>より大きい、<より小さい、>=以上の、<=以下の) の変換が中心となる。

(3). 複合条件検索の場合

問い合わせ文	成績表からクラス が A と等しく、かつ 国語 が 80点 以上のものを抽出し、氏名を表示する。
SQL文	SELECT 氏名 FROM 成績表 WHERE クラス=A AND 国語>=80

複合条件文は、すべて単一条件文の組合せとなる。単一条件文おし接続には、接続子を用いる。接続子 (ANDかつ、ORまたは) の変換が中心となる。

(4). 範囲条件検索の場合

問い合わせ文	成績表から国語 が 80点 から 100点 までのものを抽出し、学生番号を表示する。
SQL文	SELECT 学生番号 FROM 成績表 WHERE 国語 BETWEEN 80 AND 100

「～から～までの」によって、BETWEEN～AND～に変換できる。BETWEENは、ある値からある値までの範囲を検索するとき用いる。

(5). リスト探索条件検索の場合

問い合わせ文	成績表から国語 が 70点 か 80点 か 90点 のいずれかのものを抽出し、学生番号を表示する。
SQL文	SELECT 学生番号 FROM 成績表 WHERE 国語 IN (70, 80, 90)

「～か～のいずれかの」によって、IN (～, ～, …) に変換する。INは、値をカンマで区切って並べることによって、どれかの値と合致するものを検索するとき用いる。

(6). 一部探索条件検索の場合

問い合わせ文	成績表から学生番号 が 87JNで始まるものを抽出し、すべてを表示する。
SQL文	SELECT * FROM 成績表 WHERE 学生番号 MATCHES "87JN*"

一部探索とは、データ項目の一部分だけ合致するものを検索することである。「～で始まる」によって、MATCHES"～*"、"～で終わる"によってMATCHES"*～"、"～をもつ"によってMATCHES"*～*"にそれぞれ変換する。

(7). 集計関数検索の場合

問い合わせ文	成績表からすべてのものを抽出し、国語 の平均値、英語 の平均値、数学 の平均値を表示する。
SQL文	SELECT AVG (国語), AVG (英語), AVG (数学) FROM 成績表

集計関数とは、ある計算処理を関数に組み込んだものである。「～の最大値」によってMAX (～)、"～の最小値"によってMIN (～)、"～の平均値"によってAVG (～)、"～の合計値"によってSUM (～)、"その条件を満たす全件数"によってCOUNT (*) にそれぞれ変換する。

(8). グループ集計関数検索 (条件無し) の場合

問い合わせ文	成績表からクラス が 同一のものを抽出し、クラス、国語 の最大値、国語 の最小値を表示する。
SQL文	SELECT クラス, MAX (国語), MIN (国語) FROM 成績表 GROUP BY クラス

グループ集計とは、ある値でグループ化して、グループ毎の処理を行うことである。「同一の」によってGROUP BYに変換する。

(9). グループ集計関数検索 (条件付き) の場合

問い合わせ文	成績表からクラス が 同一で、国語 の平均点 が 60点 以上のものを抽出し、クラスを表示する。
SQL文	SELECT クラス FROM 成績表 GROUP BY クラス HAVING AVG (国語) >=60

「同一で、」によってGROUP BY句とHAVING句の2つへ変換する。「同一で」の前文によりGROUP句とし、「同一で」から「ものを抽出し」までの間の条件文をHAVING句とする。

(10). 分類検索の場合

問い合わせ文	成績表からクラス が A と等しいものを抽出し、国語 の昇順に 氏名を表示する。
SQL文	SELECT 氏名 FROM 成績表 WHERE クラス=A ORDER BY 国語

分類とは、あるデータ値の順番に並べ替えることである。「～の昇順に」によってORDER BY ～に、「～の降順に」によってORDER BY ～DESCに、それぞれ変換する。

以上みてきた中で、問い合わせ文の単語にアンダーラインがある部分は、すべてマルチウィンドウに表示されており、カーソル選択によって入力する。また波線がある部分は、キー入力となる。このように問い合わせ文のほとんどは、ウィンドウによってガイドされ、かつ、キー入力が少なくなる。これによってユーザーフレンドリーなRDBインターフェイスシステムとなる。

5. おわりに

以上、形式化された日本語によるRDB問い合わせ文を、SQL文に変換する手順について述べてきた。問い合わせ文を形式化仕様

に規定することによって、構文解析がやさしくなり、システム構築も容易となる。規則的なSQL変換のアルゴリズムによってシステム化を実現することができる。また、ウィンドウ上のガイドによる用語表示と、カーソルによる選択は、ユーザにとって使いやすさとわかりやすさが向上することになる。が、多少の問題として、複雑な検索記述を行うと問い合わせ文の表現に、冗長度が増してくる。いまいわしがどくなる傾向となる。

今後の課題としては、問い合わせ文の用語の精選と、機能拡張 (複数表検索、DB定義や操作など) をはかっていきたい。