

3H-7

SQL質問に対応する自然言語表現の生成

天野 浩文 上林 弥彦

(九州大学工学部)

1. まえがき

データベースを利用する際の質問の作成は一般利用者にとって負担となることがあるため、これまでに自然言語で書かれた質問を処理できるシステムの研究が数多くなされてきている。しかし、この自然言語インタフェースは、自然言語処理にともなう困難な問題を避けることができない。

われわれは、このような自然言語処理の困難さを避け、与えられた質問の意味を表す英文を簡単な方法で生成する質問作成支援システムの研究を行っている。そして、質問の意味を表す英文を、あらかじめ人間が作成した英文集合に対する単純な変換で生成する方法を得た¹⁾。

本稿では、前述の方法を関係データベース言語の標準となりつつあるSQL言語²⁾で記述された質問に応用することを考え、特に、SQLの特色の1つである入れ子質問の意味に対応する英文の生成法について考察する。

2. 基本的事項

属性集合 X の値の組合せに対し、属性集合 Y の値が一意に決まるとき、 X から Y への関数従属性(functional dependency, FD)が成立するといひ、 $X \rightarrow Y$ と表す。関係 R が $R[X_1]*R[X_2]*\dots*R[X_n]$ ($X_i \in R$, R は R に対応する属性集合、 $*$ は自然結合を表す)と常に等しいとき、結合従属性(join dependency, JD)が成立するといひ、 $*[X_1, \dots, X_n]$ で表す。

非キー属性が" \rightarrow "の右辺に現れるようなすべての自明でないFDの左辺がキーもしくはキーを含む属性集合になっており、自明でなくFDでも表せないJDを含まない関係を第五正規形であるといひ。

質問言語SQLの記述の単位は質問ブロックと呼ばれ、次のような形をしている。

```
select 属性名(のならば)
from   関係名(のならば)
where  条件式 (オプション)
```

この形でいくつかの関係を結合し、それによってできた関係から選択・射影して解を得るといふ質問を表す。

3. 標準表現群の作成

質問の意味に対応する英文を生成するもとなる文の集合は、次節以降で述べるような変換を容易にするため、

一定の制限を満たしていなければならない。ここで、例として以下のようなデータベーススキーマを考える。

(a) STUDENT(SNAME, SECT, AGE)

(b) PROFESSOR(PNAME, SECT, AGE, SAL)

第五正規形の関係のキーから個々の非キー属性へのFDに、それらの属性間の意味的なつながりを表す1つの単文を対応させ、これを標準表現と呼ぶ。

(a') Student {SNAME}: belongs to the section {SECT}
: is {AGE} years old(b') Prof. {PNAME}: belongs to the section {SECT}
: is {AGE} years old

: earns a salary of {SAL}

これらの標準表現群に対し文献1)のような変換を加えることによって、SPJ質問の意味に対応する英文が得られるが、場合によっては機械的変換に適さないこともあり、データベース管理者は事前に表現の検査・修正を行う必要がある³⁾。

4. SQLの質問ブロックに対応する英文の生成

SQLの単一の質問ブロックについては、SPJ質問との対応から、次のような変換によって、その意味に対応する英文を生成できる。

(1) from節で指定された関係に対する標準表現群だけを選ぶ。

(2) where節の条件式のうち、選択の条件を表すものについて、等式の場合は、対応する属性の記述の部分を変数で置換する。不等式の場合は、その部分のうしろに修飾語句を置く。

(3) 結合条件を表す式については、結合される2つの関係に対する表現の間に、その条件式に対応する表現を置く。

(4) 出力属性は[]でくくって示す。

(5) 質問に無関係な記述はなるべく削除する。ただし、キー属性およびデータベース管理者が文法的にのみで削除すべきでないと判断した属性については、"some ..."で置換する。

例1. SQL質問

```
select SNAME, PNAME
from   STUDENT, PROFESSOR
```

```
where STUDENT.SECT = PROFESSOR.SECT
and SAL >= 400000
```

の意味に対応する英文は次のようになる。ただし、破線部は標準表現に依存しない文字列である。

Student [SNAME] belongs to the section {SECT}.
Here, {SECT} is the same one that Prof.[PNAME] belongs to.
Moreover, Prof. [PNAME] earns a salary of {SAL} greater than 400000.

5. 入れ子のあるSQL質問に対応する変換

SQLでは、where節の中にさらに質問ブロックが任意の深さで入れ子になった形の質問が許されている。深さが任意であるために、入れ子質問全体の意味に対応する英文の一括生成はやや困難であり、また、質問作成支援システムとしての性格上、もとのSQL質問の構造を反映した表現が望ましい。そのため本稿では、内外の質問ブロックに対応する意味表現を前述の方法で生成し、それらを標準表現に依存しない文でつなぐ方法を考える。

以下では、中間に置くべき表現について述べる。この文の一般的な形は次のようになる。

{属性₁} is equal to the average of values(属性₂) that satisfy the following condition:

ただし、破線部は比較演算子や、集約関数の有無および種類により異なる。

例2. (集約関数を含む場合)

```
select SNAME
from STUDENT
where AGE = ( select min(AGE)
              from PROFESSOR )
```

この質問は教授の年齢の最小値に等しい年齢の学生の名前を求めるものである。これに対して生成される英文は次のようになる。

Student [SNAME] is {AGE} years old.
Here, {AGE} is equal to the minimum of values(AGE) that satisfy the following condition:

A professor is {AGE} years old.

例3. (内側のブロックが値の集合を決定し、属性値がそれに含まれるかどうかで検索する場合)

```
select PNAME
from PROFESSOR
where SECT is in (select SECT
                  from STUDENT)
```

Prof.[PNAME] belongs to the section {SECT}.
Here, {SECT} is equal to some {SECT} that satisfies the following condition:

A student belongs to the section {SECT}.

例2, 3の内側のブロックのwhere節において、外側のブロックの関係を参照するような条件式も許されてい

る。また、“is in”等の代わりに“contains”, “does not contain”が用いられる場合、内側のブロックの少なくとも1つは外側の関係を参照しなければ無意味である⁴⁾。

このような場合、その参照を陽に記述する表現を付加し、指示される語句が離れていてあいまいさを生じることのないよう、記号で明示しておく。

```
例4. select SNAME
      from STUDENT
      where AGE = (select min(AGE)
                  from PROFESSOR
                  where SECT = STUDENT.SECT)
```

これは、同じ講座の最年少の教授の中に自分と同じ年齢の人がいるような学生の名前を求める質問である。

Student [SNAME] is {AGE} years old, and belongs to the section {*SECT}.
Here, {AGE} is equal to the minimum of values(AGE) that satisfy the following condition:

A professor is {AGE} years old.
{SECT} is identical with {*SECT}.

6. むすび

本稿では、SQL言語で記述された質問の意味に対応する自然言語表現を求めるため、特に入れ子質問の意味表現の生成法について考察を行った。本稿で述べた方式により、元のSQL質問の構造を反映した意味表現が得られる。

内側のブロックが外側のブロックの関係を参照する場合、実際に入れ子を形成する部分と、その参照の部分とを簡単な記述で表現できず、若干わかりにくい文が生成される可能性がある。これは、関係データベース質問に対する自然言語の記述能力が十分ではないことを示唆している。

謝辞

日頃より熱心に御助言・御討論いただく古川哲也氏、Mohamed Elsharkawi氏をはじめとする研究室の皆様へ感謝いたします。

参考文献

- 1) 上林, 天野: 自然言語表現に対するデータベース操作, 情報処理学会自然言語処理研究会資料, 85-NL-52, 昭和60年11月.
- 2) Ullman, J. D.: Principles of Database Systems, Computer Science Press.
- 3) 上林, 天野: 意味表現のデータベース操作による変換, 情報処理学会第32回全国大会講演論文集, pp. 865-866, 昭和61年3月.
- 4) Kim, W.: On Optimizing an SQL-like Nested Query, ACM Trans. Database Syst., Vol. 7, No. 3, pp. 443-469, September 1982.