

擬自然言語データベースインタフェースにおける 質問作成補助システム

上林 弥彦 澤田 直 天野 浩文

九州大学工学部情報工学科

データベースシステムの普及に伴って使いやすい利用者インタフェースが重要となってきた。我々は、自然言語の分かりやすさと機械処理の容易さという2つの特色を持つ擬自然言語インタフェースの研究を行ってきた。本稿では、このシステムをさらに使いやすくするための質問作成補助システムについて述べる。本システムでは、利用者にシステム内の関係名やその意味についての知識があまりなくても使えるように、システムから提示される属性名のうち、質問に関係するものを利用者が選択する。選ばれた属性名に対応する意味表現が提示され、その編集操作によって質問を構成する。内部アルゴリズムとして超グラフの処理が必要となるが、それについては一般のグラフ上のアルゴリズムへの変換法を示した。このシステムは質問作成をメニュー選択と文字列代入という形で行うため、質問文入力の手間を大幅に省くことができる。

A Query Preparation Support System for Pseudo-Natural Language Database Interface

Yahiko KAMBAYASHI, Sunao SAWADA and Hirofumi AMANO

Department of Computer Science and Communication Engineering
Faculty of Engineering
Kyushu University
6-10-1, Hakozaki, Higashi, Fukuoka, 812 Japan

A user friendly interface is getting very important for database systems. We have been developing a pseudo-natural language database interface which has the following two advantages; easy to understand by using natural language-based expressions, and easy to be processed by computers. In this paper we will discuss a query preparation support system in order to make the system easier to use. First, names of attributes are displayed to the user, since he may not know the relations and their meanings stored in the system. By selecting appropriate names, sentences related to these names are displayed. A query is formed by editing these sentences. In order to realize such a system we need to develop algorithms to handle hypergraphs, we converted them into equivalent conventional linear graphs and developed algorithms for them. This system requires only menu selection and word substitution, which will reduce input operations to generate a query.

1. まえがき

近年のワークステーションなどの計算機技術の急速な進歩により、専門家でない利用者が関係データベース^{1),2)}を利用する機会はますます増大する傾向にある。この傾向は今後も続いていくものと考えられ、このような利用者でも容易にデータベースを利用できるような環境を整えることが重要な研究課題となっている。本稿では、現在開発中の擬自然言語データベースインタフェースにおける質問作成補助システムについて述べる。

データベースに蓄えられた情報を検索するためには、自分の求める情報がどのようなものであるかを利用者自身が記述したもの(これを質問と呼ぶ)をシステムが処理できる形で与える必要がある。現在広く採用されている質問言語はSQL³⁾などの形式的な言語が主流であるが、それらによる質問の作成はデータベースに関する専門知識を持たない一般の利用者にとって必ずしも容易ではなく、作成した質問が利用者の意図するものと一致しないこともある。

近年、この問題を解決するために多くの研究者が、利用者インタフェースの改善について研究を行ってきた。それには主に以下の3つの方法がある。

- ① 表形式のインタフェース
- ② 自然言語によるインタフェース
- ③ アイコン、ポインティングデバイス等を用いたオブジェクト指向型のインタフェース

これらのうち①の代表的な例としてはQBE⁴⁾がある。QBEは画面上の表に選択条件等を書き込むことにより、システムに質問を与えるものであり、覚えやすく使いやすい言語として知られている。しかしそのようにして作成された質問には、構文的には正しいが利用者の意図どおりでない場合が27%程含まれるという調査結果も報告されている。⁵⁾

②の方法は、利用者が質問を作成するための特別な言語を学ぶ必要がなく、自分の作成した質問の意味を確認することができるといった利点を持っている。しかし、自然言語処理に伴う各種問題のためその利点を十分に生かしたインタフェースはまだ得られておらず、システムの複雑さや、システムの個別化の困難さ、あいまいさの発生といった問題が残されている。また、この方法で利用者が求めたい情報の意味を表す文章を毎回作成し、キーボードから入力するのであれば、かえって利用者の負担を増大させることにもなりかねない。

③の方法は視覚的で分かりやすく、また、入力も簡単であるため、あまりコンピュータに馴染みの無い利用者に対して有効な方法であると思われる。しかし、利用者によってアイコンの意味を様々に解釈しうるといったあいまいさが生じる可能性がある。

これまで我々は、自然言語の利点(可読性)を取り入れながら、その欠点(システムの複雑さなど)を回避するような擬自然言語インタフェースの研究開発を行ってきた。^{6),7),8)}

これは、利用者の与えたSQL質問の意味を表す表現(擬自然言語表現)をシステムが自動的に作成し、利用者はその表現で自分の与えた質問の確認を行うとともに、その表現に各種の編集操作を施して、元の質問を修正、変更するといったインタフェースである。

擬自然言語表現とは、データの意味を表す表現群(あらかじめ管理者が作成しておいたもの)に単純な操作(値の挿入、語句の挿入/削除など)を施すことによって生成されるものである。ここで、あらかじめ用意しておく表現群に構文的制限を設けておくと、擬自然言語表現を生成する操作を単純で機械的なものにすることができるため、小規模な計算機(ワークステーションなど)上でも実用に耐えるシステムを実現することができる。

しかし、これまで述べてきた方法ではSQLなどの形式的言語による入力を前提としており、形式的言語を利用者の目から完全に隠すことはできない。

そこで、本稿では、利用者に形式的な言語を見せないような質問作成補助システムについて述べる。このシステムにおいて、利用者が質問に関与する属性をメニューから選択すると、システム内でそれらを含むような質問の原形の擬自然言語表現を作成するので、その表現を利用者の意図にあうように編集する、といったものである。また、これらの操作はマウスなどのポインティングデバイスを用いて行われるため、自然言語入力に伴う煩わしさを解消することができる。

2. 基本的事項

関係データベースは関係の集まりであり、関係は1枚の表に対応させることができる。表に見立てた関係の列はそれぞれ1つの属性に対応し、同じ表の行は組(関係の要素)と呼ばれる。関係Rを構成する属性集合を関係スキーマと呼び、 $\underline{R}$ で表現する。

質問は関係代数で表現される。関係代数の基本的な演算には次のようなものがある。

[射影(projection)] 関係の列のうちから、必要なものだけを取り出す操作を射影という。属性集合 $X \subset \underline{R}$ に対し、 $\underline{R}$ に属する組 r の X の部分のみを集めたものを $r[X]$ で表すとき、 $\underline{R}$ の X への射影は次の式で定義される:

$$R[X] = \{r[X] \mid r \in R\}.$$

[選択(selection)] 関係の中から、ある条件を満たす組だけを取り出す操作を選択という。 θ を比較演算子 $=, \neq, <, \leq, >, \geq$ のいずれかとするとき、関係 $\underline{R}$ の属性Aに関する θ 選択を次のように定義する:

$$R[A \theta 'c'] = \{r \mid r \in R \wedge (r[A] \theta 'c')\}.$$

ただし、Aの定義域とcは比較可能でなければならない。

[結合(join)] 2つの関係の組をつなぎあわせて得られる新たな組を集め、属性集合が2つの関係の属性集合の直和となるような新たな関係を得る操作を結合という。2つの組を結び付けるのは、それぞれの組の間に与えられた条件が成立する場合に限定する。この条件を結合条件と呼び、2つの組の属性値の比較で表す。同じ関係同士の結合を特に自己結合と呼ぶ。属性A($\in \underline{R}$)と属性B($\in \underline{S}$)が比較可能であるとき、 $\underline{R}$ と $\underline{S}$ の θ 結合を次のように定義する:

$$R[A \theta B]S = \{t \mid t[R] \in R \wedge t[S] \in S \wedge (t[A] \theta t[B])\}.$$

ここで、Tを結合後の関係とすると、 $\underline{T}$ は $\underline{R}$ と $\underline{S}$ の直和である。

選択や結合は複数個の属性についても適用できる。2つの関係に共通する全属性について等結合(θ が $=$ のとき)した場合、同じ値になる列の対ができる。こ

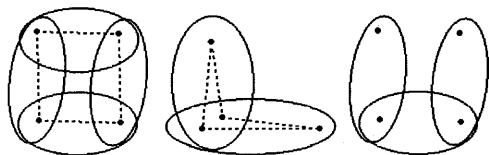
のような値の重複の除去まで行う演算を自然結合といい、 $R * S$ で表す。なお、本稿では自己結合を含まない等結合の場合のみを考える。

選択・射影・結合の他に、関係除算や和集合、共通集合、直積などの集合演算がある。関係除算は他の演算で表現できることが知られている。選択・射影・結合の3演算のみで表現できる質問を選択・射影・結合(selection/projection/join, SPJ)質問という。SPJ質問は処理効率を無視すると、必要な関係を全て結合し、それに対して選択と射影をこの順で実行することによって処理できることが知られているので、本稿ではそのような場合のみを考える。

超グラフ(hypergraph)は、節点集合 N と超枝集合 E の対 (N, E) で定義され、枝は節点の空でない部分集合で表される。

$x_1, \dots, x_m$ が相異なる節点、 $S_1, \dots, S_m$ が相異なる枝であるとき、次の4つの条件を満たす経路 $(S_1, x_1, S_2, x_2, \dots, S_m, x_m, S_{m+1})$ をBerge-閉路と呼ぶ。

- (1) $m \geq 2$
- (2) $\forall i (1 \leq i \leq m); x_i \in S_i \wedge x_i \in S_{i+1}$
- (3) $x_i \neq x_j$
- (4) $S_1 = S_{m+1}$



(a) (b) (c)

図1 Berge-巡回, 非巡回超グラフ

超グラフがBerge-閉路を持つとき(図1 (a), (b)において破線で示す)、Berge-巡回であるといい、そうでないとき(図1 (c))、Berge-非巡回であるという。

3. オブジェクトと擬自然言語表現

本節では、本研究におけるオブジェクトと用意しておく表現群、質問の分類とそれぞれに対応する擬自然言語における表現方法について述べる。

以下次のようなデータベーススキーマを例として用いる。

STUDENT (SNAME, AGE, DEPT)
PROFESSOR (PNAME, AGE, DEPT, SAL)
LECTURE (SNAME, SUBJECT, PNAME)
EXAM (SNAME, SUBJECT, MARK)

これは、大学における学生と教授、それにその講義と試験に関する情報を記録するためのデータベースである。関係として、STUDENT(学生)、PROFESSOR(教授)、LECTURE(講義)、EXAM(試験)があり、属性として、SNAME(学生名)、PNAME(教授名)、AGE(年齢)、DEPT(所属学科)、SAL(給与)、SUBJECT(科目)、MARK(得点)がある。

3.1 オブジェクト(対象)を用いた文生成

関係データベースにおける正規化の理論を用いると、関係はその中にキーからの関数従属性と自明な結合従属性しか含まない関係の集合に分解される。これを、FD-JD正規形と呼ぶ。FD-JD正規形

の関係とオブジェクトとの関連は次のようになる。関係のキーがその関係に含まれる全属性の場合は、それを一つのオブジェクトに対応させる。そうでない場合は、その関係は自明でない関数従属性 $X \rightarrow Y$ を満足することになる。Yが属性 $A, B, \dots$ の集合であるとする、この関係に対応するオブジェクトは $X A, X B, \dots$ となる。与えられたFD集合とJD集合の性質によってはFD-JD正規形の関係とならない場合もありうるが、それらの関係についてもほぼ同様の考えでオブジェクトを求める。普遍関係仮定を考えた場合、全属性のキーに相当する属性集合のみを含む関係を加える必要があるが、詳細は設計論の問題であるため省略する。

ただし、システム内での関係自体はオブジェクト単位には分解されず、もとの関係スキーマを保存している。

すなわち、関係 STUDENT は2つのオブジェクト {SNAME, AGE}, {SNAME, DEPT} に分解されるが、システム内では、STUDENT (SNAME, AGE, DEPT) といった関係スキーマを保存している。

このそれぞれのオブジェクトの意味は、そこに含まれる全ての属性の名を含む自然言語の1文で記述することができる。これらの属性の名を含む文は、そのオブジェクトに対応する組の集合の意味を表すテンプレートと考えることができる。

それぞれのオブジェクトに対し、各属性を節点、オブジェクトを超枝とする超グラフを考える。たとえば、先の例の関係スキーマのオブジェクト構造は図2のように表される。

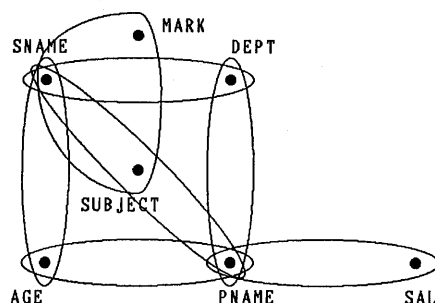


図2 関係のオブジェクト構造

オブジェクトに対応する英文は出力文の材料となるもので、あらかじめ管理者が作成しておく。この際、そのオブジェクトを表す1つの文(標準表現)と、他のオブジェクトと結合されうる属性についてそれを先行詞とする関係詞節や前置詞句(副節)を用意しておけば、それらの組合せにより質問に対応する文章を作成することができる。すなわち、1つのオブジェクトの標準表現に、そのオブジェクトと共通部分を持つオブジェクトの副節を埋め込んでいくことにより、各属性名が1回ずつしか現われないような形式で質問に対応する文章を作成することができる。

例えば、オブジェクト {SNAME, SUBJECT, PNAME} に対しては次のような標準表現及び副節を作成しておけばよい。ただし、2番目に示した副節は標準表現より容易に導出できるので、必ずしもあらかじめ作成しておく必要はない。

Student {SNAME}

: takes the lecture {SUBJECT}
given by Prof. {PNAME}.
Student {SNAME}
: who takes the lecture {SUBJECT}
given by Prof. {PNAME}
Prof. {PNAME}
: who gives the lecture {SUBJECT}
taken by student {SNAME}

他の関係の標準表現及び副節については、付録を参照されたい。このような表現は、意味ネットワーク等を用いて自動的に導出できることが望ましいが、現段階では管理者が用意するものと仮定しておく。

質問をオブジェクトの組合せで表すために質問に関係するオブジェクト集合の超グラフを考える。ただし、結合条件で比較される属性の節点は異なるものであっても同一とみなし、超枝を重ね合わせ、逆に結合に関与しない属性の節点はそれぞれ別のものとみなし、超枝を重ね合わせない。後者の例としては、年齢を教授と学生とで分けて考えるような場合がある。

この超グラフは、表現の骨格を表すと同時に質問の骨格(関係の結合の構造)も間接的に表している。このような超グラフに、選択条件、結合条件、出力属性の集合といった情報を付加した超グラフを質問超グラフと呼ぶことにする。なお、本稿では質問超グラフの構造が Berge-巡回になっているとき、その質問のことを巡回質問と呼ぶ。

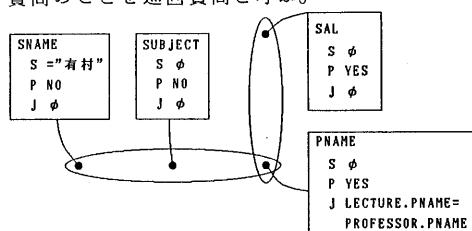


図3 質問超グラフの例

3.2 質問の分類とそれぞれに対応する表現

質問超グラフによって表される英文は、選択・射影・結合の各演算に対応する構文的特徴を有しており、SQL質問と1対1に対応している。本節では、この対応関係をもとに、関係データベース質問と英語の構文との関連について検討する。

選択・射影・結合(自己結合を除く)演算で記述される関係データベース質問を、その質問超グラフの構造から次のような3つのクラスに分類する。

- (第1種) 質問超グラフの構造が Berge-非巡回であり、かつ、その任意の節点が相異なる関係スキーマに含まれる3個以上の枝に共有されない質問(図4(a))
- (第2種) 質問超グラフの構造は Berge-非巡回であるが、ある節点が相異なる関係スキーマに含まれる3個以上の枝に共有される質問(図4(b))
- (第3種) 質問超グラフの構造が Berge-巡回である質問(図1(a), (b))

第1,第2種の質問は、Berge-非巡回であるから、後方の副節の属性名が副節のはるか前方にもあると

いった事態が起きない。したがって質問の意味を表す文章を1文で表すことができる。

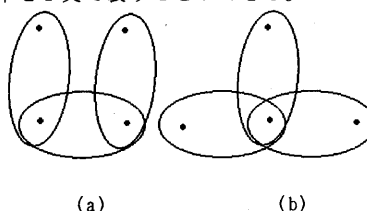


図4 第1,第2種質問の対象構造の例

このうち第1種の質問は、1つの先行詞に複数の副節がつながることがなく、あいまいな表現にはならない。しかし、第2種の質問は1つの先行詞に複数の副節が連結される可能性があり、副節が直前の先行詞を修飾できず、あいまいさが生じるおそれがある。そこで、そのあいまいさを除くような表示方法をとらなければならない。

第3種の質問は Berge-巡回であるので、後方の副節の属性名が副節のはるか前方にもあるといった事態が生じ、あいまいな表現になるおそれがある。また、2つのオブジェクトが複数の属性を共有するような場合(図1(b))、(連続した位置にないかもしれない)2つの先行詞を同時に修飾しう副節は一般には存在しないため、これもあいまいな表現となってしまう。

いままでは、第1種の質問だけを扱ってきたが、本稿では第2,第3種の質問に対しても対応できるように以下の方法を用いて擬自然言語表現を作成する。

ここで、選択に関する属性集合をS、射影に関する属性集合をP、結合に関する属性集合をJとする。

[第1,第2種質問の擬自然言語表現の作成法]

- ① S U P U J を被覆する連結なオブジェクト集合を元の質問の結合に関する部分を参照して決定する。その中で射影属性を最も多く含むオブジェクトをレベル1とし、一般にレベルiのオブジェクトと共通属性を持つオブジェクトをレベルi+1とする。
- ② レベル1のオブジェクトに対応する標準表現をレベル2のオブジェクトによって修飾される語句まで表示した後、次の行にレベル2の対応する副節を字下げして表示する。レベル1の文の残りの部分はその次の行に元の字下げで表示する。同様のことをレベルiとレベルi+1の副節の間でも行う。
- ③ Sに対応する語句の後に選択条件を表す修飾語句を埋め込む。
- ④ 出力属性集合Pを強調するため、表現中のPの要素の部分を[]でくくる。

[第3種質問の擬自然言語表現の作成法]

- ① それを取り除くことによりすべての巡回構造を解消できるようなオブジェクト集合のうち最小のものを求める。これが複数ある場合には射影属性をできるだけ多く含むものを選ぶ。
- ② ①で求めたオブジェクト集合に対応する擬自然言語表現と、そのオブジェクト集合を取り去った残りのオブジェクト集合に対応する擬自然言語表現を通し番号をつけて並列に表示する。

- ③ それに続き、各属性がどの擬自然言語表現に含まれるかを調べ、複数の表現に含まれる属性について、

{属性名} in No. {番号}, No. {番号}, ... are all equal to each other.

とコメント文を表示する。

このようにして生成される擬自然言語表現は、質問の解となる組の集合の意味を表すテンプレートになっている。

3.3 擬自然言語表現上での S P J 質問の編集

利用者が、質問の意味として出力された文と自分の意図との相違に気付いた場合には、出力文を修正することにより質問を修正する。このとき、利用者の出力文に対する自由な入力を許すと、複雑な構文解析を行う必要が生じるため、操作を限定することによってシステムを単純化する。

本システムでは、利用者の編集操作として以下の6種類の操作を用意してある。ただし、④の操作は巡回質問の場合以外では行えないようにする。

[擬自然言語表現上での編集操作]

- ① 選択条件の変更, 挿入, 削除
- ② 副節の挿入, 削除
- ③ 出力属性の設定, 解除
- ④ 同じ属性名でも結合させないものを指定し、結合を解除する
- ⑤ 異なる属性名でも結合させるものを指定し、結合を追加する
- ⑥ or による条件の結合を含む場合、どの条件とどの条件を or で結ぶかを指定する

先に述べたように、質問文は質問超グラフで表せる。この質問超グラフから対応する質問を求めることは容易であり、上記の操作が表現に加えられたとき、対応する質問超グラフ上の操作を求めることができる。

①, ③, ⑥の操作は、対応する質問超グラフの超グラフとしての形状を変化させず、選択条件の定数あるいは比較演算子、結合条件を変更するだけの操作であるから、元の質問超グラフの上で、変更箇所に対応する部分を修正するだけでよい。

②, ④, ⑤の操作は、質問に対応する質問超グラフの形状を変化させる。②の操作に対しては、挿入(または削除)された副節に対応する超枝を加える(または削除する)だけでよい。ただし、この操作において質問超グラフが連結でなくなる場合は、その操作を受け付けないようにしておく。④の操作に対しては、指定された結合部分を質問超グラフ上で切り放してやればよい。⑤の操作に対しては、指定された属性に対応する部分を質問超グラフ上で重ね合わせればよい。

なお、これらの利用者の操作は数値などの入力を除いて画面上のポップアップメニューをポインティングデバイスで指定することにより行われる。

4. 質問作成補助システム

本節では、一般の利用者の質問の作成を補助するため、質問に関与する属性を選択すると、それらを結合する質問の擬自然言語表現を作成し、その表現

を利用者に意図に沿うように編集させる、といった質問作成補助システムの概要について述べる。

4.1 S P J 質問の作成

このシステムにおいて利用者は、画面上で以下の操作を行い、擬自然言語表現を得ると共に S P J 質問を作成する。

[画面上での操作]

- ① 画面上に表示された、データベースの属性に対応する名詞句のうち、質問に関係するものを選択する。
- ② ①で指定した属性に対応する名詞句のうち、射影するものを指定する。
- ③ 画面上に①の操作で選択された属性を含む擬自然言語表現が質問の原形として表示されるので、その表現に各種編集操作を施し、自分の意図にあったものにすする。

この一連の画面上での操作に対し、システム内では、S Q L 質問が作成される。

例えば、先の例において利用者が“有村という学生が受けている講義を教えている教授の名前とその給料”を知りたいとする。

まず、段階①②では

student	{SNAME}
professor	{PNAME}
age	{AGE}
department	{DEPT}
salary	{SAL}
subject	{SUBJECT}
mark	{MARK}

のように表示された属性に対応する名詞句のうち必要なものを選択する。この質問に対しては student, [professor], subject, [salary] を選択する。ただし、[] で囲まれているものは射影する属性である。すると、段階③では以下の文章が表示される。

```
(1) Prof. [PNAME]
    who gives the lecture {SUBJECT}
        taken by student {SNAME}
    earns the salary of [SAL].
(2) Student {SNAME} got the mark {MARK}
    in subject {SUBJECT}.
* Student {SNAME} in No. (1) and No. (2)
  are equal to each other.
* Lecture {SUBJECT} in (1) and Subject
  {SUBJECT} in (2) are equal to each
  other.
```

ここで(2)の文を削除し、{SNAME} のところに、{SNAME="有村"} と書き込むという編集操作を行い、以下の文章を得る。

```
Prof. [PNAME]
  who gives the lecture {SUBJECT}
    taken by student {有村}
  earns the salary of [SAL].
```

利用者がこの意味の質問でよいと確認をしたらシス

テム内で

```
SELECT PNAME, SAL
FROM PROFESSOR
WHERE PNAME = LECTURE.PNAME
AND LECTURE.SNAME = "有村"
```

といったSQL質問が作成される。

4.2 質問超グラフを決定するための前処理

上記「画面上での操作」①②で選ばれた属性の名を含む擬自然言語表現を「画面上での操作」③で表示するためには、その質問のオブジェクト構造を求めなくてはならない。そのために、まず、関係全体のオブジェクト超グラフに対し以下の処理を行って超グラフを一般のグラフに変換する。

〔超グラフから一般のグラフへの変換〕

- ① 1つのオブジェクトに対しオブジェクト節点を対応させ、属性には属性節点を対応させる。
- ② オブジェクト節点と、そのオブジェクトに含まれるすべての属性節点とを枝で結ぶ。

この変換により、超グラフは、属性節点とオブジェクト節点が交互に現れる一般のグラフに変換される。例えば、図2のオブジェクト構造を表した超グラフは、図5のような一般のグラフに変換される。ただし、この図において属性節点は●で、オブジェクト節点は▲で表される。

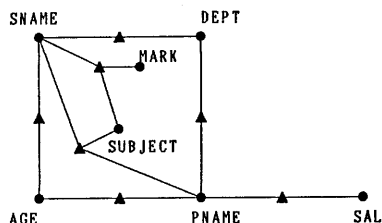


図5 超グラフから変換されたグラフの例

超グラフの任意の節点間の距離は、同じ枝内の異なる節点間の距離を1として推移的に定義されるが、この変換処理によって任意の属性節点間の距離は超グラフの時の2倍になる。したがって超グラフ上の距離関係はすべて保存されている。

次に、選ばれた属性節点を印付き節点に変換し、以下の処理を行ってグラフを簡単にするにより、質問のオブジェクト構造を得る。

〔グラフの簡単化〕

- ① オブジェクト節点の1つに注目する。
- ② そのオブジェクト節点に隣接する節点(属性節点及びこのステップで構成される印付き節点)のうち2つ以上が印付き節点であれば、そのオブジェクトを記録し、それに隣接するすべての節点とともに1つの印付き節点に変換する。
- ③ ①, ②の操作を、すべてのオブジェクト節点に対して繰り返す。

このグラフの簡単化の処理によって全体のグラフは、互いに1つのオブジェクト節点だけをはさずには隣接しない印付き節点と、属性節点、オブジェ

クト節点とによるグラフになる。

4.1節で挙げた例の質問について上記の処理を行うと、図5のグラフは図6のように印付き節点が1つにまとまったグラフになる。ここで印付き節点を○で示す。

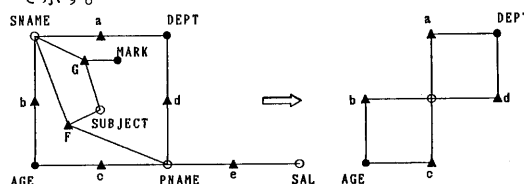


図6 簡単化を行ったグラフの例

4.3 あいまいさの要因

あいまいさの要因としては、条件の結合方式によるものと質問超グラフの構造によるものとがある。

条件の結合方式によるものについて述べる。

質問超グラフで表される質問において、利用者が特に指定しなければすべての条件は and で結合されているものとしている。これは、or による条件の結合を含むとすると、それを考慮して擬自然言語を表示する順番を定めなければならないが、利用者の与える属性集合にそのような情報は含まれないからである。

次に、質問超グラフの構造によるものについて述べる。

まず、印付き節点が1つにまとまった場合について考える。

この場合、「画面上での操作」①で選択された属性集合を、記録してあるオブジェクトだけで覆うことができるので、そのオブジェクト集合だけを用いて擬自然言語表現を作成できる。しかし、印付き節点の中身(オブジェクト集合)の構造が Berge-巡回になっている場合は、その質問は複数の意味を持つ可能性があるため、一意には定められない。

つまり、利用者の与えた属性集合だけを用いて「グラフの簡単化」の処理を行うと、不必要なオブジェクトを取り込んでしまったり、結合しなくてもよい属性を結合させてしまったりする可能性があるのである。

したがって、印付き節点の中身の構造が Berge-巡回になっている場合、以下の3つの場合について意味を考えなければならない。

- (a) 巡回質問である場合
- (b) 質問に不必要なオブジェクトが含まれている場合
- (c) グラフ上では結合しているが、実際は結合していない属性が存在する場合

次に、印付き節点が1つにまとまらない場合について考える。

関係のオブジェクト構造が図7のように複雑であるときは、利用者が「画面上での操作」①で選択した属性について「グラフの簡単化」の処理を行っただけでは、図8のように印付き節点が1つにまとまらずその節点間の経路が複数存在するような場合が起こり得る。これは、利用者が質問に関係する属性集合を選択したとき、それが十分ではなかったときに起こる。

その場合、次のような質問の候補が考えられる。

- (a) それぞれの印付き節点間のすべての経路で結合する場合、
- (b) いくつかの結合経路で結合する場合、
- (c) グラフ上では結合しているが、実際は結合していない属性が存在する場合

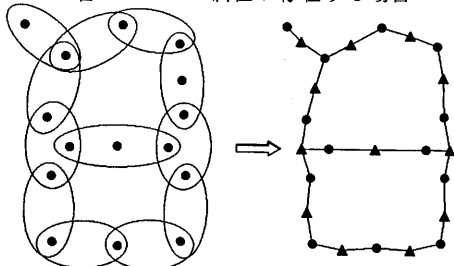


図7 複雑なオブジェクト構造の例

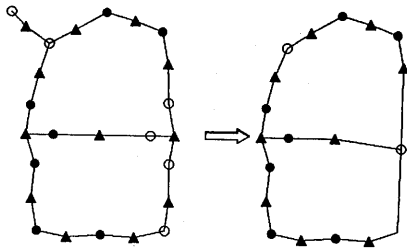


図8 印付き節点が1つにまとまらない場合の例

4.4 あいまいさの除去

前節で述べたようなあいまいさを除去するために、次のような擬自然言語表現を質問の原型として表示し、その表現に先に述べたような編集操作を施してあいまいさを除去する。

[質問の原型の擬自然言語表現]

- ① すべての条件は and によって結合されているものとする
- ② 印付き節点の中身が Berge-巡回になっている場合は、巡回質問であるものとする
- ③ 印付き節点が1つにまとまらず、かつ、その節点間の結合経路が一意に定まらない場合は、すべての結合経路で結合しているものとする

印付き節点が1つにまとまらず、その節点間の結合経路が一意に定まらない場合は、質問の原型の擬自然言語表現を得るため、[グラフの簡単化]の処理の後以下の操作を行い、印付き節点だけの重み付きグラフに変換する。

[重み付きグラフへの変換]

- ① 印付き節点のうち2つだけを結ぶすべての経路を、グラフ探索アルゴリズムを用いて求める。
- ② それぞれの経路に含まれるオブジェクトを記録し、そのオブジェクトの数を重みとする重み付きの枝で印付き節点同士を結ぶ。

図8で表される質問に上記の操作を行うと、図9のような重み付きグラフになる。

こうして求められた重み付きグラフで、まず、印

付き節点に対応するオブジェクト集合を擬自然言語表現で表し、次に、それらを結合する経路の候補として重み付き枝に対応するオブジェクト集合を重みの小さい順に擬自然言語表現で表示する。

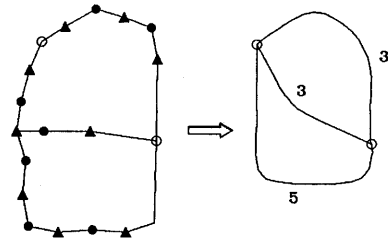


図9 変換された重み付きグラフ

4.5 S Q L 質問への変換

このようにしてあいまいさを除去した擬自然言語表現を S Q L 質問に変換する手法について述べる。

システム内において、関係、オブジェクト、属性は図10のようなネットワーク構造をとっている。

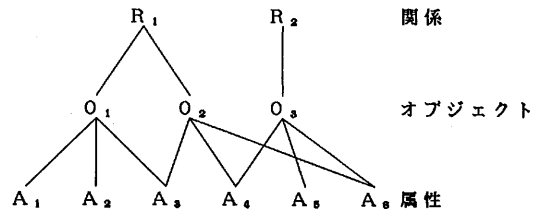


図10 システム内のネットワーク構造

このネットワーク上で以下の処理を行い、擬自然言語表現をもとに S Q L 質問を作成する。

[S Q L 質問の作成]

- ① 射影する属性が、それぞれどの関係に属しているかを調べる。
- ② 擬自然言語表現に含まれるオブジェクト集合が、それぞれどの関係に属しているかを調べる
- ③ 各関係がどの属性において結合しているかを調べる
- ④ 選択条件が付加されている属性が、それぞれどの関係に属しているかを調べる
- ⑤ S Q L 質問を
SELECT 射影属性の集合
FROM オブジェクトが属する関係の集合
WHERE 関係の結合条件
AND 選択条件
というように作成する

5. むすび

本稿では、利用者が形式的な言語を用いることなくデータベースを利用できるようにするための質問作成補助システムについて述べた。このシステムは、利用者が画面上での簡単な選択操作で質問の意味を表す英文(擬自然言語表現)を作成することにより、それに対応する S Q L 質問を内部で自動的に作成する。本システムでは自然言語による入力を直接に処理するわけではないが、利用者が自分の作成しつ

ある質問の意味を理解しながら作業できるという特色がある。

また、本稿では結合条件については等結合だけを用いている。これは、 θ 結合 (θ は $=, \neq, <, >, \leq, \geq$ のいずれか) にまで拡張すると、各オブジェクトに対してそれぞれの θ に対応する表現を用意しておき、それも利用者に選択させる方法を用いなければならないためであるが、そのような拡張も可能である。

なお本システムは、Sun 3/260C 上で開発中であり、マルチウインドウ上で擬自然言語表現の選択編集による SQL 質問の作成、及び、SQL 質問の作成に伴う擬自然言語表現を変更することによる質問の修正などを行う予定である。

謝辞

日頃より熱心に御助言御討論下さる本学情報工学科古川哲也助手、並びに大学院生掛下哲郎氏、Mohamed El-Sharkawi 氏をはじめとする研究室の諸氏に深謝いたします。

なお本研究は昭和63年度文部省科学研究費補助金特定研究「言語情報処理の高度化」のうちの「情報ドキュメンテーションのための言語の研究」の援助を受けている。

参考文献

- 1) Ullman, J. D.: Principles of Database Systems, 2nd Ed., Computer Science Press (1982).
- 2) Date, C. J.: An Introduction to Database Systems, Vol. 1, 4th Ed., Addison-Wesley (1986).
- 3) 芝野: データベース言語 SQL, 情報処理, Vol. 29, No3, pp. 208-214, 1988年3月.
- 4) Zloof, M. M.: Query by Example, Proc. AFIPS National Computer Conf., pp. 19-22, May 1975.
- 5) Thomas, J. C. Gould, J. D.: A Psychological Study of Query by Example, Proc. AFIPS National Computer Conf., pp. 439-445, May 1975.
- 6) 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), 1985年11月.
- 7) Kambayashi, Y.: An Overview of Natural Language-Assisted Database User Interface: ENLI, Proc. IFIP 10th World Computer Congress, pp. 1055-1060, September 1986.
- 8) 天野, 上林: 自然言語表現上の操作による関係データベース質問の作成, 情報処理学会データベースシステム研究会資料 66-7, 1988年7月.

付録

- (1) オブジェクト {SNAME, AGE} の標準表現及び副節
Student {SNAME}
: is {AGE} years old.
{AGE} years old
: that is as old as Student {SNAME}
- (2) オブジェクト {SNAME, DEPT} の標準表現及び副節
Student {SNAME}
: belongs to Department {DEPT}.
department {DEPT}

- : that Student {SNAME} belongs to
- (3) オブジェクト {SNAME, SUBJECT, MARK} の標準表現及び副節
Student {SNAME}
: gets {MARK} marks for Subject {SUBJECT}.
Subject {SUBJECT}
: for which Student {SNAME} gets {MARK} marks
 - (4) 関係 PROFESSOR (PNAME, AGE, DEPT, SAL) に含まれる各オブジェクトの標準表現及び副節
Prof. {PNAME}
: is {AGE} years old.
: belongs to the Department {DEPT}.
: earns the salary of {SAL}.
{AGE} years old
: that is as old as Prof. {PNAME}
Department {DEPT}
: that Prof. {PNAME} belongs to