

## データベースアシスタントにおける知識の動的獲得

川染 俊樹† 平川 正人‡, 市川 忠男‡

‡広島大学大学院工学研究科

†広島大学工学部

自然言語によるデータベースアシスタントでは、ユーザからの多様な表現による問い合わせが予測される。しかし、それら全ての問い合わせに対応する事ができる知識ベースをあらかじめ構築しておく事は、知識ベース構築者にとって困難である。その為、ユーザのデータベース検索過程において動的に知識を獲得していく事が必要となる。

本稿では、データベース問い合わせ文中に未知語などの処理不可能な要素が含まれていた場合に、それらに対処するための質問列をユーザに提供し、動的に知識を獲得していく手法を提案する。

## Dynamic Acquisition of Knowledge for Database Assistant

Toshiki Kawasome† Masahito Hirakawa‡ Tadao Ichikawa‡

†Graduate School of Hiroshima University

‡Faculty of Engineering, Hiroshima University

1-4-1 Kagamiyama Higashi-Hiroshima 724, Japan

On a database assistant with a natural language interface, it is expected that queries are given in a various forms of expressions by users. But, it is hard for a knowledge-base designer to construct a knowledge-base that can respond to all possible types of queries. Therefore, it is required that the system acquires user's demands in the process of operations.

In this paper, we provide a new tool for dynamic acquisition of knowledge where given an unknown terms in a query. The system generates a series of questions properly to integrate the acquired knowledge by gaining that term.

## 1 はじめに

データベースの応用分野が広がるにつれて、コンピュータに関する専門的な知識を持たない人達が直接データベースを操作する機会が増えてきている。しかし現状では、彼等にとって直接データベースシステムを操作する事は容易ではない。このため、データベースシステムやデータベース検索言語に関して学習する必要性が生じる。しかし、それらを学習する事は彼らにとって負担が大きく、かつ困難である。そこで、そのようなユーザにとってもデータベースから必要な情報を直接かつ簡単に検索できるように、自然言語インタフェースを用いたデータベースアシスタントに関する研究が行われている[1]。現在、我々もKDA(Knowledge-based Database Assistant)[5]-[6]の開発を行っている。

自然言語インタフェースを用いたデータベースアシスタントでは、ユーザの多様な表現による問い合わせに対応できなければならない。しかし、ユーザからの全ての表現に対応し得るだけの知識ベースをあらかじめ構築する事は不可能である。そのため、ユーザが問い合わせを行う過程で、順次知識を獲得していく事が必要となる。

知識の獲得については、ユーザの問い合わせと検索結果からルールなどを抽出し、効率の良いデータベース検索を行うための研究[2]がある。また、未知語などを含んだ問い合わせに対処する研究もある。例えばCOOP[3]では、検索文に含まれるあいまい性を含む語に対して、より詳細な問い合わせをユーザに要求する事によって処理を完了させる。FIFA90[4]は、問い合わせに含まれる未知語に対して文法的な解析から品詞を決定し、リスト形式などで未知語の候補となる情報をユーザに提供し選択させ、ユーザに対して意味のある検索結果を返す方法を提供している。

本研究では、KDAシステム上でのデータベース検索過程において、未知語などの検索不可能な要素が含まれていた場合に、適切な質問列をユーザに提供し、動的に知識を獲得していく方法を提案する。

以下、2章ではKDAシステムの概要と本研究のアプローチ、3章では新しく導入した知識の構成、4章では知識獲得の方法、5章では知識ベース管理について述べる。

## 2 研究の概要

### 2.1 KDAの知識

KDAで用いられる知識ベースは、S-Net[5]ならびにネットワーク変換知識と辞書から構成されている。S-Netは、以下に定義されるように、ユーザ概念 $S_u$ およびデータベース概念 $S_d$ の2層から構成される意味ネットワークモデルである。また、それらの間の変換ルール $T$ を持つ。

$$S := \langle S_u, S_d, T \rangle$$

$S_u$ : ユーザ概念 S-Net

$S_d$ : データベース概念 S-Net

$T$ : ネットワーク変換知識

- ユーザ概念 S-Net

ユーザ概念レベルの問い合わせを表現するユーザ概念 S-Net  $S_u$ は、

$$S_u := \langle C_u, L_u \rangle$$

$C_u$ : ユーザ概念の集合

$L_u$ :  $C_u$ に属する概念間の関係の集合

で表される。

- データベース概念 S-Net

関係スキーマの意味構造は、データベース概念 S-Net  $S_d$

$$S_d := \langle C_d, L_d \rangle$$

$C_d$ : データベース概念の集合

$L_d$ :  $C_d$ に属する概念間の関係の集合

で表される。

- ネットワーク変換知識

KDAでは、ユーザが入力する検索要求文を一旦ユーザ概念 S-Netに変換する。そして、これをデータベース概念 S-Netへと変換し、そしてデータベース検索言語 SQLへと変換するが、この際に用いられる知識がネットワーク変換知識 $T$ である。これは、ユーザ概念 S-Netのノードとリンクをデータベース概念 S-Netのノードとリンクに変換するためのノード変換知識ならびにリンク変換知識の2種類から成る。

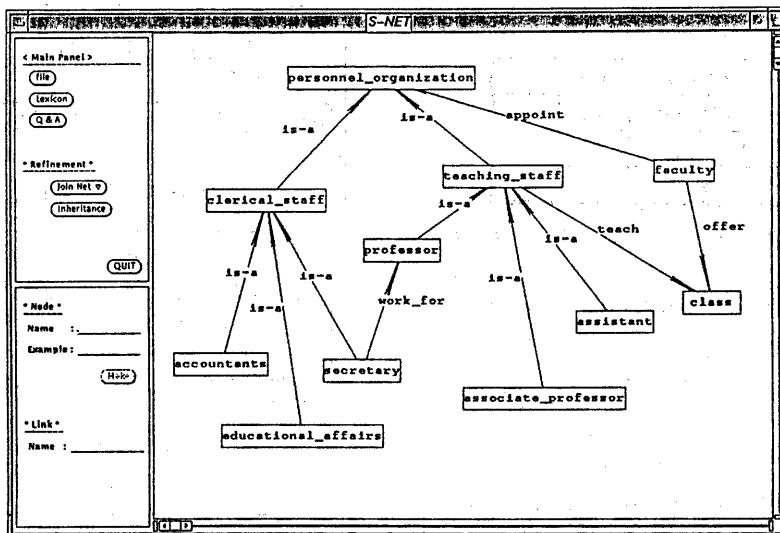


図 1: 知識ベース構築ツール

知識ベース構築者(以後、デザイナーと略称する)はあらかじめ専用のツール [6] を用いてこれらの知識を構築しておく(図 1)。

## 2.2 アプローチ

自然言語によるユーザからの問い合わせに対してシステムが正しく応答できない場合として、

- 1) データベース検索を行うための十分なデータがユーザの問い合わせに含まれていない
- 2) 知識として登録されていない単語(未知語)が使われている

などが考えられる。

1) の場合には、問い合わせ文を SQL に変換する際に必要な情報 (Relation, Attribute) が不足して SQL 文が生成できないか、あるいはまた、データベースの構造に適した SQL に変換できないなどの問題が生じる。

2) の場合には、知識として登録されていない単語が使われているために、品詞が分からず、従って構文解析ができないとか、品詞の推測が可能な場合においても単語とデータベースとの対応がとれないなどの問題が生じる。

これらの問い合わせが行われた時には、データベース検索実行に欠落している情報をユーザから獲

得しなければならない。

自然言語インタフェースの場合、ユーザからの知識獲得は、質問列を提供してそれらに対してユーザが答えを返すという会話形式で行われるのが一般的である。しかし、正確な知識を獲得しようとする、多くの質問列をユーザに提供し、これに回答してもらわなければならない。

本来、ユーザの目的は情報の獲得にあるので、逆にシステムからの質問に答え、システムに情報を提供する作業は大きな負担となる。そこで、ユーザに負担をかけずに知識を獲得する事を目標とし、各々の問い合わせに対しては、その時点における必要最小限の知識獲得にとどめる様にし、そしてそれらの知識を総合してより適格な知識へと変換していく方法を考える。

## 3 知識構成

### 3.1 知識獲得の対象となる問い合わせ

本研究では、知識獲得を伴うユーザの問い合わせとして、

- 1) データベースの Value がユーザの検索文の中で目的語となっている  
(以後、検索目的語と略称する)

#teaching\_staff

name	position	age	fno
Smith	professor	50	1
Brown	associate professor	37	1
Jone	assistant	29	2

図 2: 検索目的語が Value となるデータベース

2) 未知語 (名詞、動詞) が使用されている

の 2 つの場合を考える。

1) というのは、例えば図 2 のスキーマ構造の Relation に対して

User > Get the name of professor.

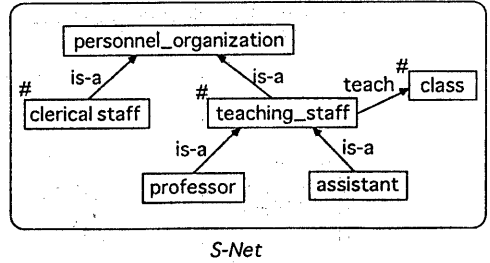
という問い合わせが行われた場合である。この問い合わせでは、単純には *professor* という Relation の *name* という Attribute を検索しようとする。しかしながら該当する Relation、Attribute は存在しないため、対応のとれた SQL 生成に失敗する。

2) の場合、まず未知語の品詞が何であるか、そしてそれがデータベースの Attribute、Value などの何に対応するか、または S-Net のどの Node や Link に対応する単語であるのかをユーザへの質問列を通じて限定しなければならない。

本システムでは、これらの問い合わせに対応する為の知識として Personal S-Net と未知語知識の 2 つを導入する。以下、それぞれについて説明する。

### 3.2 Personal S-Net

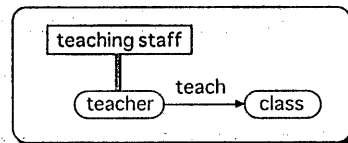
あるユーザのデータベース検索過程において問い合わせに未知語が存在した場合、それに対処するためにユーザから未知語と S-Net の対応に関する知識を獲得し、S-Net とのマッチングを行うが、そこで獲得した未知語と S-Net との対応の知識、ならびに、ユーザが使用した S-Net と対応している単語の関係を Personal S-Net として保存していく。例えば図 3 では、問い合わせ文中の *teach* と *class* が S-Net 中の同名の Link と Node に対応し、未知語 *teacher* は S-Net の *teaching\_staff* に対応していると予測されるので、*teacher* が *class* を *teach* するという知識と共に、*teacher* は *teaching\_staff* に対応しているという知識が Personal S-Net に保存される。



S-Net

User > Get the teacher  
who teaches the class of CS.  
KDA > Teacher is regarded as teaching staff.  
KDA > Smith.

検索例



Personal S-Net

図 3: Personal S-Net

Personal S-Net は、そのユーザがデータベース検索を行う間のみ存在する一時的な知識であり、そのユーザに対してはすでに正しさが確認された知識であるので、通常の S-Net と同様にその後の検索に使用される。

### 3.3 未知語知識

Personal S-Net として獲得された個々の知識の中で、未知語に関する全てのユーザの知識を未知語知識として総合して保存する (図 4)。未知語知識は全てのユーザの検索を対象として使用される知識とする。未知語知識に含まれる単語をユーザが使用した場合、その知識を用いて S-Net との対応を行う。

未知語知識  $UK$  は以下の様に定義する。

$UK := \langle U_n, U_a, U_l \rangle$

$U_n := \langle U_{n-own}, U_{n-up}, U_{n-new} \rangle$

$U_{n-own}$ : 1 つの Node それ自体に対応する未知語についての知識

$U_{n-up}$ : ある Node とその上位概念を含んで対応する未知語についての知識

$U_{n-new}$ : あるいくつかの Node の新しい上位概念として位置づけられる未知語についての知識

$U_a$ : Attribute に対応する未知語についての知識

$U_l$ : Link に対応する未知語についての知識

S-Net 上において、ある Relation  $R_0$  に対応している Node  $N_0$  とその上位概念の Node の集合を  $N$  とおく。ユーザの問い合わせにおいて、未知語  $X$  が検索目的語であり、 $R_0$  に関しての検索を要求していると考えられる場合、この要求を含む未知語  $X$  と S-Net 上の Node との対応は

$$X \in N$$

と表される。つまり、 $X$  が  $N$  のどの要素であるかを限定しなくても、 $R_0$  を検索する事によってユーザに対して必要な情報を提供する事ができる。

例えば図 5 において、

User > Get the teacher

who teaches the class of CS.

という問い合わせが行われ、teacher が未知語であったとする。その時、この問い合わせ文中の teach と class が S-Net とマッチするため、検索する Relation 同士の関係として、teaching\_staff に対応する Relation と class に対応する Relation の teach に相当する join 関係が唯一候補となる。そして、この関係についての問い合わせが含まれるのは、未知語 teacher の代わりに teaching\_staff 又は、その上位概念である personnel\_organization と person の内どれかが使われるという 3 つの場合である。しかし、これらの内のどれが使われたとしても、システムが生成する SQL の形は同じになる。そのため、これらをこの時点で限定する必要はないので、それら全てを未知語 teacher の候補とする。

このため、 $U_{n-up}$  という未知語知識が生じる。 $U_a$ ,  $U_l$ ,  $U_{n-own}$ ,  $U_{n-new}$  の 4 つの知識は、未知語と S-Net の Attribute, Link, Node がそれぞれ 1 対 1 で対応するが、 $U_{n-up}$  の場合は未知語と S-Net の Node との対応に範囲がある。したがって、その後のデータベース検索にこの未知語が使われた場合、ユーザ

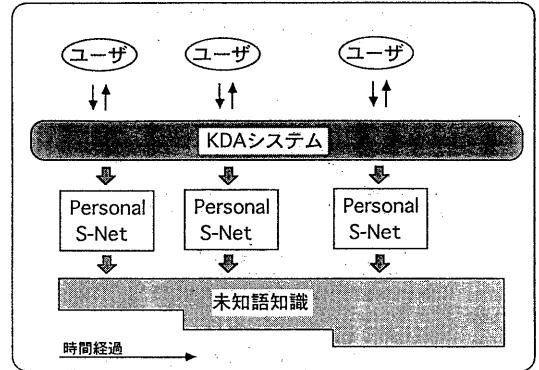
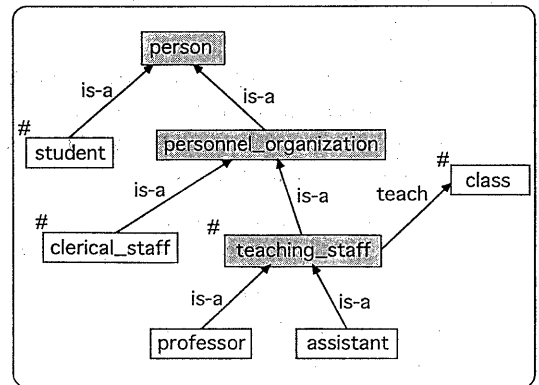


図 4: Personal S-Net と未知語知識との関係



... 未知語 teacher の候補

図 5: 未知語と S-Net との対応

の期待する応答ができる様に未知語候補の範囲をせばめていく質問列を提供し、より適格な知識へと変化させていく。

## 4 知識獲得

知識獲得を行う局面として、

- 検索目的語がデータベースの Value である場合
- また、未知語が使用されている場合においては、
- Personal S-Net に同名の単語が存在する場合
- 未知語知識に存在する場合
- 初めて使用された未知語の場合

という合計 4 つの場合が考えられる。以下、それぞれのケースごとの知識獲得処理について述べる。

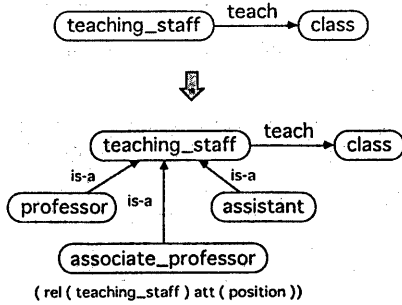


図 6: Personal S-Net への知識の追加

#### 4.1 検索目的語がデータベースの Value である場合

検索目的語に Value が使われる場合、問い合わせ中にはその Value が含まれる Relation に関しての情報は含まれないことが多い。例えば、図 2 のスキーマ構造の Relation に対しては、

User > Get the name of teaching staff  
whose position is professor.

という問い合わせ文に展開してはじめて検索が可能となる。しかし、teaching\_staff, position といった情報は問い合わせ文からは獲得できない。

KDA では、それぞれの Value について、Relation、Attribute の情報をあらかじめ辞書として獲得しているので、ユーザからこの様な問い合わせが行われた場合、その文に不足している知識を Personal S-Net に新しい Node として生成する (図 6)。そして Personal S-Net を用いてユーザの問い合わせの不足部分を補い、スキーマ構造に適した SQL を生成する。

#### 4.2 問い合わせに未知語が使用された場合

##### 4.2.1 Personal S-Net に存在する場合

ある未知語が Personal S-Net に存在する場合、その未知語は現在操作中のユーザが以前使用した事を意味する。その場合には、すでにユーザには未知語と S-Net との対応が過去の問い合わせが行われた時点で確認されているので、その時獲得した知識をそのまま使用する。

##### 4.2.2 未知語知識に存在する場合

ある未知語が未知語知識に存在する場合、その未知語は現在問い合わせしているユーザより以前に他のユーザが使用した事を意味する。その内で、未知語が  $U_{n-own}$ ,  $U_{n-new}$ ,  $U_a$ ,  $U_l$  の場合には、S-Net との対応が 1 対 1 に限定されているから、ユーザに対して対応関係の確認をとって使用する。

$U_{n-up}$  の場合、未知語と S-Net との対応に幅があるので、ユーザの要求を満たす範囲でその幅をせめていく事ができる様にする必要がある。

以下、その限定の手順を図 7 の例を用いながら説明する。なお、() 内は図 7 に示す例についての記述である。また前提として、未知語  $X$  に対する未知語知識  $U_{n-up}$ 、及びユーザからの問い合わせ  $Q$  を

$$U_{n-up}(X) \ni \{A, C, F, H\}$$

$$Q \ni \{I, s, X\}$$

とする。

**Step1:** ユーザが使用した未知語と同名の未知語知識が存在しており、かつ問い合わせから予想される S-Net 上での未知語の位置がその未知語知識の範囲に含まれるならば、その未知語知識を候補とする。

( $Q$  の要素、 $I, s$  から予想される  $X$  に対応する S-Net は  $C$  となり、これは  $U_{n-up}(X)$  に含まれるので、この未知語知識が候補となる)

**Step2:** 候補となった Node が  $U_{n-up}$  の最下位の Node であるなら、それは Relation と対応している Node なので、未知語がその Node と対応する事に決定する。

(最下位の Node は  $H$  なので、この時点では決定しない)

**Step3:** 候補の Node が持つ下位概念の内、 $U_{n-up}$  に含まれない Node について、ユーザに対して未知語に含まれるかどうかを問い合わせる。(  $D, E$  に対して、ユーザに問い合わせる)

そして全てが含まれるならば現在の候補を正式な対応とみなし ( $C$ )、一部が含まれるならば、その含まれる Node と  $U_{n-up}$  の下位概念 Node の新しい上位概念となる。

( $E$  が含まれる時、 $E$  と  $F$  の新しい上位概念)

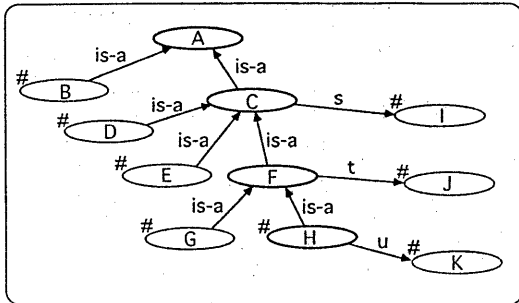


図 7:  $U_{n-up}$  の限定

もし、全てが含まれないならば、 $U_{n-up}$  の下位概念の Node ( $F$ ) を新しい候補とし、Step2, Step3 の処理を繰り返す。

#### 4.2.3 初めて使用された未知語の場合

ある未知語が初めて使用された場合、構文解析によって品詞を決定し、問い合わせに含まれる既知の単語と S-Net との関係から未知語と S-Net との対応を限定する。

##### ● 未知語が Node に対応する場合

未知語に対応する Node はユーザの問い合わせに含まれる Link に対応する単語から限定するが、意味ネットワークにおける Link の継承によって同名の Link が多数存在する場合がある。その場合には検索の候補となる Node が多くなる可能性があり、ユーザに提供する質問列が多くなってしまう。そこで、S-Net の中で未知語の候補が図 8 のモデルに位置する場合にはユーザに対して Yes/No 形式の質問列を、それ以外の場合には候補に含まれる Relation に対応している Node 全てをユーザに提示し選択させる方法により、未知語と Node との対応をとる。

##### ● 未知語が Attribute に対応する場合

この場合、未知語に対応する Attribute が属している Node について、その Node に属している Attribute の中で他の Relation の外部キーになっていない Attribute 全てを候補としてユーザに提示し、

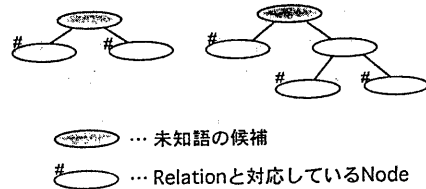


図 8: Node 決定のモデル

選択させる。

##### ● 未知語が Link に対応する場合

この場合、S-Net 上の 2 つの Node が未知語以外の単語から決定するので、その 2 つの Node を結ぶ Link に対しての同義語、あるいは反意語と考えられる。そこでその Link をユーザに提示し、同義語、反意語のどちらであるかを決定させる。

#### 4.3 データベース検索例

図 9 の S-Net が構築されていたとする。この時、あるユーザから

User > Get the name of a teacher  
who teaches the class of CS.

という問い合わせが行われた場合、teaching\_staff、personnel\_organization が未知語 teacher の候補となる。しかし、この場合はこれらを特定する必要がないので、

KDA > Teacher is regarded as a teaching\_staff.

というコメントと共に検索結果をユーザに返す。この時、未知語知識

$U_{n-up}(teacher) \ni$   
{ personnel\_organization, teaching\_staff }

が生成される。また、この状態で別のユーザから

User > Get the name of teacher  
who is appointed in Computer Science.

という問い合わせが行われた場合、 $U_{n-up}(teacher)$  が候補となり、範囲を限定する為の質問

KDA > Is teacher included in clerical staff ?

をユーザに対して行う。この場合 No. とユーザが答

えると、*teacher*の対応が *teaching\_staff*に限定される。その結果、未知語知識も限定され

$$U_{n-up}(teacher) \ni \{teaching\_staff\}$$

となる。

## 5 知識ベース管理

未知語知識は、ユーザの検索文から獲得した知識であるから、必ずしも正確であるとは言えない。ユーザ特有の言葉の使い方から獲得された知識である可能性もある。そこで、デザイナーの知識ベース管理時に、データ変更などの他に S-Net 上へ獲得した未知語知識を提供し、知識獲得の作業を行う。

作業内容としては、

- 検索目的語として使用された Value について、その使用が頻繁に行われるものであれば S-Net の Node として生成する
- 同義語についての確認
- ある2つ以上のノードの新しい上位概念に相当するノードの生成
- S-Net との対応が限定できなかった未知語に対しての限定を行う

を行い、正確な知識として獲得していく。

## 6 まとめ

自然言語によるデータベースアシスタントにおいて、ユーザの問い合わせに対してより柔軟に対応し、また検索過程において動的に知識を獲得していく手法について述べた。この手法では、個々の問い合わせに対しては、ユーザの要求に応えられる範囲での知識獲得を行い、それらを総合してより正しい知識へと限定していくため、個々のユーザに対しての負担を軽減する事ができる。今後の課題としては、獲得した知識のデザイナーへの提供方法、複数の未知語を含む検索文への対応などが挙げられる。

なお、KDA システムは、言語に Common Lisp、C 言語、X View、データベースに G-BASE を用いて Sparc Station 上で構築中である。

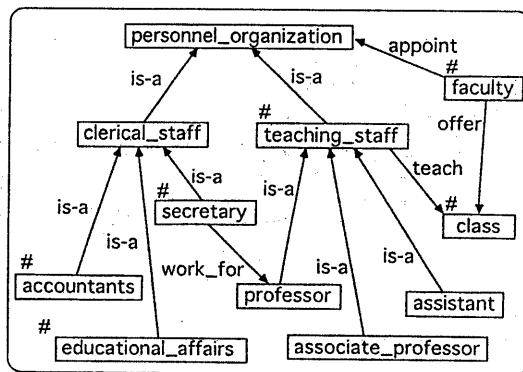


図 9: S-Net 例

## 参考文献

- [1] G. Jakobson, G. Lafond, E. Nyberg and G. Piatetsky-Shapiro, 1 "An Intelligent Database Assistant," IEEE EXPERT, vol.1, no.2, pp.65-78, 1986.
- [2] C. T. Yu and W. Sun, "Automatic Knowledge Acquisition and Maintenance for Semantic Query Optimization," IEEE Transactions on Knowledge and Data Engineering, vol.1, no.3, pp.362-375, 1989.
- [3] S. J. Kaplan, "Cooperative Responses from a Portable Natural Language Query System," Artificial Intelligence, 19, pp.165-187, 1982.
- [4] Z. Trabelsi and Y. Kotani, "自然言語データベースインタフェースにおける未知語の文脈的处理," 信学技報, NLC92-34, pp.23-30, 1992.
- [5] X. Wu and T. Ichikawa, "KDA: A Knowledge-Based Database Assistant with a Query Guiding Facility," IEEE Transactions on Knowledge and Data Engineering, vol.4, no.5, pp.443-453, 1992.
- [6] 田中, 川染, 平川, 市川, "データベースアシスタント構築支援ツール," 情報処理学会研究報告, 93-DBS-96, pp.45-54, 1993.