

あいまい性を含む問合せ処理能力を備えたデータベース・アシスタント

三石 淳† 吉高 淳夫‡ 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・システムにおいては、検索文中の表現が必ずしもデータベースの内部構造と対応するとは限らない。その場合、検索が失敗したことをユーザに示すだけでは不十分であり、検索者が次の方針を立てる際の手助けとなるような情報を別途提供する必要がある。

本稿では、あいまい性を含んだ表現に対して知識に基づく解釈を施し、検索文とデータベースとの対応を図る手法を提案する。また、解釈過程を提示することで、検索者への確認と検索方針の立案をアシストする。

A Database Assistant with Ability of Processing Ambiguous Queries

Jun Mitsuishi† Atsuo Yoshitaka‡ Masahito Hirakawa‡ Tadao Ichikawa‡

†Graduate School of Hiroshima University

‡Faculty of Engineering, Hiroshima University

1-4-1 Kagamiyama Higashi-Hiroshima 739, Japan

In a database assistant system with a natural language interface, an expression in a query doesn't always correspond to a database structure. Furthermore, the action of indicating the failure in retrieval doesn't support formation of the next plan of data retrieval.

In this paper, we propose a database assistant system which provides the ability of query processing based on knowledge for ambiguous term in a query. Furthermore, the system also assists confirmation and formation of a plan of retrieval by indicating the process of query interpretation.

1 はじめに

データベースの応用分野が広がるにつれて、データベース非専門家であるユーザ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、データベースシステムやデータベース検索言語についてほとんど知識を持たないユーザにとって、直接データベースシステムを操作することは容易ではない。このため、データベースシステム及び検索言語について学習する必要性が生じるが、それはユーザにとって負担が大きい。そこで、データベース検索を容易に行うための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]。これまで、我々も KDA (Knowledge-based Database Assistant) の開発を行ってきている [2]。

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・システムは、メニューなどを用いる場合と異なり、欲しい情報を直接引き出すことができ、かつ親和性に富み、多様な入力を可能とする。しかし、自然言語表現はあいまい性を含んでいるために、ユーザが要求を提示した際に、それが必ずしもシステムが保持している対象世界に関する知識と直接対応するとは限らない。またそれに伴って検索に失敗した場合、その旨を応答するだけではユーザが次に行う検索の方針を立てることに困難が生じる。よって、ユーザの要求に対する柔軟な解釈能力と応答能力を持つことが必要となる。

例えば FLEX [3] では、ユーザからの要求文に対して説明やアシストする事なしに空値を返さない手法が提案されている。COOP [4] においては、要求文に含まれるあいまい性を伴う語に対して、ユーザとの対話によって対処する手法が提案されている。

本研究では、KDA システムにおいて形容詞を使って表現されたあいまい性を含む検索要求に対して、それを処理する手法、ならびに処理の過程を検索結果に併せて提示することによって検索要求に対する確認と検索方針の立案をアシストする方法について提案する。

以下、2 章では KDA システムの概要と本研究のアプローチを、3 章では KDA システムにおける知識の拡張について述べる。4 章ではユーザからの検索要求の解釈方法について示し、5 章ではユーザへの応答方略について述べる。

2 研究の概要

2.1 KDA システムの概要

KDA ではユーザから自然言語の形で与えられた問合せに対して、構文解析、意味解析を施すことによって問合せを中間言語である S-Net (図 1) に変形し、それを用いてデータベース検索言語 SQL に従った質問列を生成する。それに基づいてデータベース検索を行い、ユーザに返すべき値を獲得する。KDA システムでは、自然言語表現と中間言語 S-Net との対応を図るための辞書、また中間言語 S-Net と SQL の対応を図るための S-Net から構成される知識ベースを備えている。以下、それぞれについて記す。

Q: Get the name of teacher who teaches class of CS.

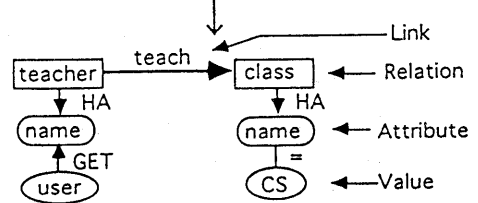


図 1: 中間言語 S-Net

- 辞書：各データベース・ドメインから独立した個々の単語に関する品詞、活用などの統語的情報、またドメインに依存した Relation, Attribute, Value に関する情報を備えている。
- S-Net：データベース・ドメインに属する概念 (Relation)、もしくはそれらが保持している特徴 (Attribute) をノードとし、また概念間の関係をリンクとして表現した意味ネットワークである。S-Net は以下のように構成されている。

$$S\text{-Net} ::= \langle S_u, S_d, T \rangle$$

$$S_u ::= \langle C_u, R_u \rangle \quad S_d ::= \langle C_d, R_d \rangle$$

S_u : ユーザ概念 S-Net

C_u : ユーザ概念でのノードの集合

R_u : ユーザ概念でのリンクの集合

S_d : データベース概念 S-Net

C_d : データベース概念でのノードの集合

R_d : データベース概念でのリンクの集合

T : S_u から S_d への変換知識

2.2 アプローチ

自然言語表現には特有のあいまい性が含まれる。ここで、言語学におけるあいまい性として、(a) 多品詞語 (1つの単語が単一の品詞ではなく、いくつかの品詞を持つ) によるあいまい性、(b) 構文上のあいまい性 (前置詞句の修飾関係によるあいまい性、もしくは並列句における範囲のあいまい性)、(c) 意味上のあいまい性 (単語の多義性、形容表現、省略表現といった単語の解釈に起因するあいまい性) が挙げられる。

KDA システムでは、(a) 及び (b) については構文解析時に辞書、及び構文規則によってあいまい性を解消する。しかし、(c) については単語の含み得る意味を考慮する必要があるため、辞書、S-Net といった従来の知識ではそれらの表現に対応することが出来ない。よって本研究ではこの点に注目し、意味上のあいまい性の解消についての考察を行う。

なお、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、形容表現や省略表現を含む要求に対してもデータベース構造との対応づけを行い、ユーザの要求に適った問合せを生成する必要がある。ただし、システムの解釈がユーザの要求とは必ずしも一致しないため、協調的応答を行うことが望ましい。

そこで、ユーザからのあいまい性を含んだ問合せに対処するために、本研究においては知識に基づく解釈を施し、データベースの内部構造との対応を明確にするとともに解釈を施した過程、及びそれによる検索結果をユーザに提示する。このような観点から検索をアシストする手法についての考察を行う。

3 知識の意味的拡張

ユーザからの問合せの解釈を促すために辞書の意味的拡張である意味辞書、S-Net の拡張である属性知識を用意する。以下にそれぞれについて述べる。

3.1 意味辞書

あいまい性を含む表現を明確にするためには、その表現がどのような意味を含んでいるのかを考える必要がある。そこで、単語を意味素 (Semantic-Primitive) の集合として捉える。

本来、意味素とは単語と単語が結合し得るために必要な意味的整合関係、すなわち選択制限を行うための情報である。しかし、本研究においては、自然言語表現とデータベース構造との抽象的レベルでの対応を図るための情報として扱う。この知識は形容表現 (4.1 節) を解釈するために用いられる。よって形容詞、並びにそれが修飾し得る名詞についての意味素を定義する。

名詞：問合せ文中に用いられる名詞は概念 (現実世界の対象、及び現象)、もしくは概念の特徴 (概念が保持する特性) である。それらを中間言語 S-Net と対応させると、ノードに相当することになる (図 2)。ここで、Relation、Attribute として用いられる語義は異なっているため、名詞の意味素 (SP_{Noun}) を

$$SP_{Noun} := \{SP_{Rel} \mid SP_{Att}\}$$

SP_{Rel} : 概念 (Relation) に関する意味素

SP_{Att} : 特徴 (attribute) に関する意味素

と定義する。

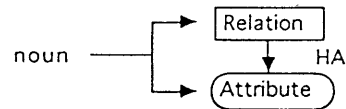


図 2: 名詞と中間言語 S-Net との対応

SP_{Noun} は図 3 のような階層構造で表現され、下位の意味素は上位の意味を継承する。なお、これらは文献 [7] における意味素性に準ずる。

形容詞：形容詞は、文献 [6] において、修飾される対象の属性を表す形容詞としての属性形容詞、その語を用いる主体の感情を表現する形容詞である感情形容詞に分類される。本研究では属性形容詞を対象とする。

ここで、修飾される対象は Relation、属性は Attribute を指すと捉えることができる。また、形容詞はものごとの性質、状態を示す語であることを考慮すると、中間言語 S-Net との対応は図 4 の斜線部となる。

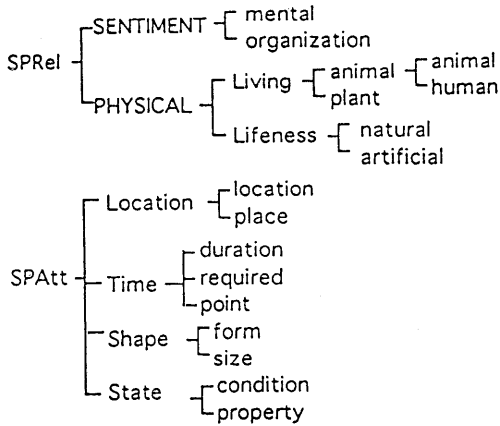


図 3: 名詞の意味素階層

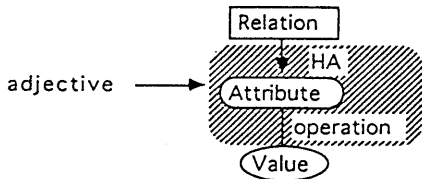


図 4: 形容詞と中間言語 S-Net との対応

従って、意味素 (SP_{Adj}) を以下のように定義する。

$$\begin{aligned}
 SP_{Adj} &:- \{SP'_{Rel}, SP'_{Att}, SP_{Ope}\} \\
 SP'_{Rel} &: SP_{Rel} \text{の集合} \\
 SP'_{Att} &: SP_{Att} \text{の集合} \\
 SP_{Ope} &:- \{\text{more} \mid \text{less} \mid \text{between} \mid \text{equal}\}
 \end{aligned}$$

3.2 属性知識

対象となるデータベースドメインでのユーザ概念 S-Net においては、異なる Relation 間で共通の概念で、かつ共通の特徴をもつ Attribute, すなわち意味辞書における SP_{Rel} , SP_{Att} が同一であるような Attribute を統括し、これを属性知識として設ける (図 5)。属性知識はあいまいな表現の解釈過程における対象となる Attribute の決定、また検索過程の説明文生成に用いられる。

属性知識は Attribute の集合であり、以下のように定義する。

属性知識 :- $\{SP_{Rel}, SP_{Att}, L_{Att}\}$
 L_{Att} : 統括される Attribute のリスト

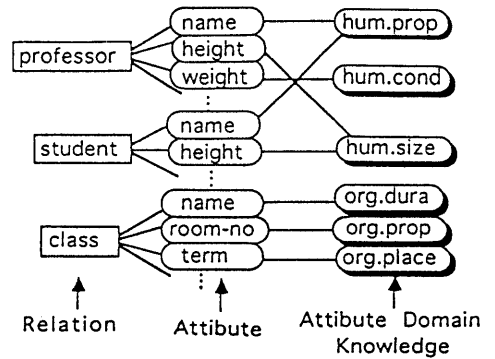


図 5: 属性知識

この知識は、データベースと知識ベースを同時に作成する際に意味辞書を参照して自動的に生成される [9]。

4 検索要求の解釈

4.1 形容表現

4.1.1 表現の解釈

ユーザからの問合せに対して形態素解析を行った結果、名詞句 (NP) に含まれる表現について以下のルールに従って解釈を施す。以下、修飾される名詞を target-noun と呼ぶ。

(a) target-noun を修飾する語が 1 語の場合

$$\begin{aligned}
 NP &:- \{[\text{target-noun}] \langle \text{relative-pronoun} \rangle \\
 &\quad \langle \text{be-verb} \rangle \langle \text{adjective} \rangle\} \\
 NP &:- \{\langle \text{adjective} \rangle [\text{target-noun}]\}
 \end{aligned}$$

ここで、 $\langle \text{adjective} \rangle$ は $[\text{target-noun}]$ が保持している特徴に対し何らかの条件を課す表現である。すなわち、 $\langle \text{adjective} \rangle$ は SQL における WHERE 節に対応するものと解釈する。よって、それを構成する

Attribute (T_{Att}), Operation (T_{Ope}), Value (T_{Val}) が検索文中に明示されていなければならないことを考慮し,

NP:- {[target-noun] whose (T_{Att})
is (T_{Ope})(T_{Val})}...[R1]

という文型に変形する。そして <adjective> と T_{Att} , T_{Ope} , T_{Val} との対応を図る。なお下線部分の変数 (T_{Att} , T_{Ope} , T_{Val}) をつなぐ自然言語の定型パターンである。

(b)target-noun を修飾する語が 2 語の場合

- (1) NP:- {<cardinal><noun>[target-noun]}
- (2) NP:- {<adjective><adjective>[target-noun]}
- (3) NP:- {<ordinal><adjective>[target-noun]}

(1) においては

NP:- {[target-noun] whose <noun>
is <cardinal>}...[R2]

とし、<noun> は [target-noun] の Attribute として取り扱う。また、データベース構造と対応がとれていない場合には、<noun> は Attribute に対する未知語 [10] としての処理を行う。

また (2), および (3) に関しては [R1] を適用し、<adjective> についての解釈を施す。

4.1.2 解釈の手順

ユーザの検索要求を解釈する手順を以下に示す。
Step1: 構文解析時に形容表現を検出した場合には、それらを削除した表現により中間言語 S-Net を生成する。

例えばユーザからの検索要求文として

USER > Get the names of masters who are taller
than Prof. Smith.

が与えられた場合、構文解析によって taller が tall の比較級、またその比較対象 (T_{Val}) が Smith であるこ

とが検出される。よって、taller 以外の表現を S-Net に変換すると図 6 のようになる。

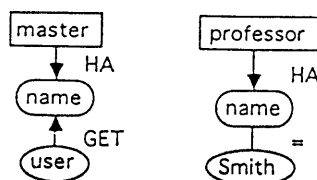


図 6: 形容表現を削除した中間言語 S-Net

Step2: 形容表現に対して意味辞書を参照し、その表現を意味素表現に変換する。

Step1 で検出された tall とそれが修飾し得る master についての意味素表現は

tall: (SP_{Rel} (PHYSICAL) SP_{Att} (size) SP_{Ope} (MORE))
master: (SP_{Rel} (human))

である。よって、tall は human が保持する特徴 size に制約 MORE をかける語として解釈される。ここで MORE は T_{Ope} に対応する。

Step3: 意味素表現によって属性知識を決定し、Step1 で生成された中間言語 S-Net と対応している Attribute を選出する。

図 5 において、属性知識 human.size が保持している master, professor に関する Attribute である height の補完を行う。また、 T_{Att} の候補となる Attribute が複数存在する場合は、ユーザに問い合わせることにより T_{Att} を決定する。

Step4: T_{Val} が問合せ中に明示されていない場合には、ユーザへの質問列を生成し、その応答により T_{Val} を決定する (4.1.3 節)。

Step5: 決定された T_{Att} , T_{Ope} , T_{Val} を S-Net に変形し、Step1 で生成された中間言語 S-Net に結合する。

以上の処理を施すことにより、図7のような中間言語 S-Net が生成される。

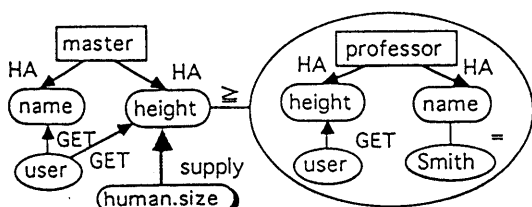


図7: 解釈を施した中間言語 S-Net

4.1.3 基準値 (T_{Val}) の決定

T_{Val} は形容詞の活用形に大きく依存する。比較級、及び最上級の場合、 T_{Val} は検索文中に明示されているか、もしくは暗に示されている。しかし、

USER > Get the names of tall masters.

のように原級が用いられる場合においては必ずしも示されるとはいえない。この場合、ユーザとの対話により T_{Val} を決定する。ここで、Boolean-Logic の問題点 [8] を考慮し、ユーザが納得できる値の獲得を行うために以下の手順で T_{Val} を決定する。

Step1: ユーザへ基準値に関する質問を提示し、応答された値を T_{Val} として検索を行う。

Step2: 検索された検索結果の件数をユーザに示すと共に、件数の多少、もしくは検索された値の表示を選択させる質問を提示する。ユーザが前者を選んだ際には Step3 へ、後者の場合は Step4 へ。

Step3: 基準値の変更を行う。それをユーザに提示し、Step2 へ。

Step4: ユーザに基準値を提示すると共に、検索結果を提示する。

4.2 省略表現

省略表現は省略語、指示語として問合せ文中に表れる。それらは、ユーザが文中の表現の重複を避け

るために用いていると考えられる。よって、それらの表現に対し、意味を一意に定める必要が生じる。そこで、それらの表現がデータベース構造の Relation, Attribute, もしくは Value のいずれに対応するのかを S-Net と属性知識から決定し、表現の補完、もしくは置換を行う。

例えば、ユーザからの問合せとして

USER > Get the name of secretary whose salary is more than that of Jone.

が与えられた場合、検索要求に対する中間言語 S-Net は図8のようになる。ここで、*that* は secretary の salary を統括している属性知識によって assistant への置換がなされる。

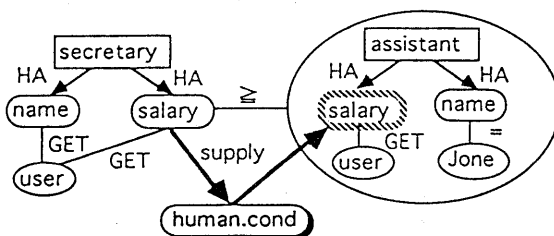


図8: 省略表現に対する中間言語 S-Net

5 応答方略

ユーザに対する応答として、問合せを解釈した説明文とその検索を行った過程を提示する。これにより、ユーザに対する確認を図るとともに、次の検索方針の立案を促すことを方略とする。

5.1 解釈説明文の生成

知識に基づく解釈を施された中間言語 S-Net を用いて解釈説明文を生成する。この際、ユーザの誤解を避けるためにも、解釈説明文は正規化された問合せ文、すなわち直接検索可能な問合せ文の形で表現することが妥当である。

そこで、以下に示す4つのルールに従い、中間言語 S-Net から正規化された問合せを生成する。

Rule1: 属性知識により補完された Attribute は検索対象属性とする。これにより検索された値はユーザへ付加情報として与える

Rule2: 条件部が検索対象である Relation に属する Attribute についてのものであれば、関係代名詞 “whose” を用いて表現する

Rule3: 検索対象である Relation に対して Link の入もしくは出があれば、前者の場合は過去分詞、後者は現在分詞で表現する。また、Link の先にある Relation, Attribute は用いずに、その末端にある Value のみを表現する

Rule4: 検索対象である Relation に対して複数の条件が存在する場合は、Rule2, Rule3 を踏まえた上で接続詞 “and” を用いて文を接続する

これらのルールを図7に適用することにより

*Get the names of masters and their height
whose height is more than the height of Smith.*

という正規化された問合せ文が生成される。

5.2 検索過程の表示

検索過程をユーザに提示する際において、5.1の方法で生成された説明文をそのままユーザに提示しただけでは、それが複雑である場合、検索過程を提示したとは言いがたく、逆にそれによってユーザの理解を妨げる要因となり得る場合が生じる。それには以下の2つのケースが考えられ、それぞれについて対処する。

Case1: 属性知識により、複数の Relation に対しての Attribute の補完がなされた場合

補完がなされた Relation を対象とした問合せに分割し、検索順序に従って問合せとその検索結果の組を表示する。なお、検索は S-Net のリンクをたどり、ユーザが要求している Attribute を保持する Relation から距離の遠い Relation から順次行う。

4.1.2節の例において、ユーザが要求するのは master の name であるため、master についての検索は最後

に行う。まず、professor についての問合せ、

*Q1 > Get the height of professor
whose name is Smith.*

を生成し、その検索結果 (A1 とする) との組でユーザに提示する。次に、正規化された問合せにおいて *height of Smith* を A1 で置換し、master についての検索を行う。

Case2: ある Relation において、検索される Attribute が多数存在している場合

それぞれの検索対象 Attribute についての検索を行う問合せに分割し、各問合せとその検索結果との組として表示する。

Case1 において検索された A1 を用いて、master の name, height に関する問合せに分割し、それぞれについて検索を行なった結果との組でユーザに提示する (図9)。

6 まとめ

自然言語インタフェースを用いた知的データベースアシスタントにおいて、柔軟な解釈能力と応答能力を兼ね備えることでユーザからのあいまい性を含む検索要求に対応する手法について述べた。この手法では、言語特有のあいまい性を含む表現の意味についての解釈を施すことにより、対象世界に適した要求文を生成することができる。さらに、その解釈過程を提示することでユーザの理解を促進することが可能である。今後の課題としては、意味素の妥当性の実験的検証、また、属性知識による補完がなされない場合における検索文代替案の生成などが挙げられる。

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。

参考文献

- [1] G. Jacobson, G. Lafond, E. Nyberg, and G. Piatetsky-Shapiro, “An Intelligent Database

```

Q & A
QUIT
Question:
USR > Get the names of masters who are taller than Prof. Smith.
KDA > THE KDA SYSTEM REGARDS YOUR QUERY AS BELOW.
KDA > (GET THE NAMES OF MASTERS WHO ARE TALLER THAN SMITH)
KDA > ->
KDA > (GET THE NAMES OF MASTERS AND THEIR HEIGHTS WHOSE HEIGHT IS MORE
      THAN THE HEIGHT OF SMITH)
KDA > GET THE HEIGHT OF PROFESSOR WHOSE NAME IS SMITH.
A-1 > 170.2.
KDA > GET THE HEIGHTS OF MASTERS WHOSE HEIGHT IS MORE THAN 170.2.
KDA > (GET THE NAMES OF MASTERS WHOSE HEIGHT IS MORE THAN 170.2 )
KDA >
A-2 > 190.5(EDWARD) 178.0(YAMAI) 175.5(TAKAKA) 175.5(MITSUISHI).

* SQL Query *
open database '/demo/fire/ISL';
SELECT PROFESSOR.PHEIGHT
FROM PROFESSOR
WHERE PROFESSOR.PNAME = 'SMITH' ;
SELECT MASTER.MHEIGHT
FROM MASTER
WHERE MASTER.MHEIGHT >= 170.2
ORDER BY MASTER.MHEIGHT DESC;

```

図9: 解釈過程の表示

- Assistant," IEEE EXPERT, Vol. 1, No. 2, pp. 65-78, 1987.
- [2] X. Wu and T. Ichikawa, "KDA: A knowledge-based database assistant with a query guiding facility," IEEE Trans. on Knowledge and Data Engineering, Vol. 4, No. 5, pp. 443-453, 1992.
- [3] A. Motro, "FLEX: A Tolerant and Cooperative User Interface to Databases," IEEE Trans. on Knowledge and Data Engineering, Vol. 2, No. 2, pp. 231-246, 1990.
- [4] S. J. Kaplan, "Cooperative Responses from a Portable Natural Language Query System," Artificial Intelligence, Vol. 19, pp. 165-187, 1982.
- [5] 間瀬, 木山, 絹川, "DB 検索用自然語インタフェースにおける解釈結果確認文生成方式の開発とその評価," 情報処理学会論文誌, Vol. 35, No. 8, pp. 1579-1590, 1994.
- [6] 内海, 堀, 大須賀, "自然言語処理のための形容詞の意味表現," 人工知能学会誌, Vol. 8, No. 2, pp. 192-200, 1993.
- [7] 長尾, "言語工学," 昭晃堂, 1983.
- [8] P. Bosc and M. Galibourg, "Flexible Selection Among Objects: A Framework Based on Fuzzy Sets," Proc. 11TH Int'l. Conf. on Research & Development in Information Retrieval, pp. 433-449, 1988.
- [9] 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," 情報処理学会データベースシステム研究会報告, 93-DBS-96, pp. 45-54, 1993.
- [10] 川染, 平川, 市川, "データベースアシスタントにおける知識の動的獲得," 情報処理学会データベースシステム研究会報告, 95-DBS-101, pp. 81-88, 1995.