

## 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

間瀬久雄<sup>†</sup> 木山忠博<sup>†</sup> 絹川博之<sup>†</sup>

自然言語でデータベース (DB) 検索システムを効率的に操作するには、入力文の解釈結果を自然言語文で利用者に報知することが効果的である。しかもこの自然言語文は、省略情報を補い、曖昧性を解消し、また、利用者の馴染みやすい表現であることが肝要である。本論文では、これらの要件を満たした、解釈結果確認文を生成する方式を提案している。すなわち、本方式は、(1)特定の二属性項目間の関係を定義した自然言語表現を入力文に応じて適宜連結することにより任意の二属性項目間の関係を自然で滑らかに表現し、(2)入力文の記述順序に則った文を生成し、(3)属性項目名の省略された属性データに対して属性項目名に相当する自然言語を補完し、(4)検索対象が明示されていない入力文に対して検索対象情報を補完する、という四つの特徴を有している。具体的な DB を対象として本方式を評価した結果、条件構成の違いにより 6 種類に分類できる検索指示文のそれぞれに対して、入力文に対応した確認文の生成が可能となり、(A)解釈結果内容を自然言語文で把握できることを確認した。また、(B)確認文を入力文として流用することによって曖昧性のない入力文を作成できることも確認した。以上から、解釈結果内容を自然言語文で報知することが有効である見通しを得た。

### Confirmation Sentence Generation for Natural Language Interface in Database Retrieval Domain and its Evaluation

HISAO MASE,<sup>†</sup> TADAHIRO KIYAMA<sup>†</sup> and HIROSHI KINUKAWA<sup>†</sup>

This paper presents a sentence generation method for confirming user's intension, expressed in natural language sentence, in the database retrieval domain. Basic premise under this research is that confirmation sentence facilitates users to understand system's ability and to acquire with expressing acceptable retrieval sentence. Using the natural language expression defining relation among database columns, the method generates non-ambiguous confirmation sentence, including the target column name not expressed explicitly in retrieval sentence. Another features of the method is that the confirmation sentence has the description order of phrases expressed in a retrieval condition as the input retrieval sentence. Using six types of retrieval sentence, the paper illustrates that implemented system satisfies user requirement and that generated sentence can be also analysed as non-ambiguous retrieval sentence.

#### 1. はじめに

社会の情報化に伴い、計算機サービス事業が発展し、さまざまなソフトウェアが普及している。しかし、これらの操作は容易でないため、どんな利用者でも容易に操作できるためのインタフェース (IF) が求められている。この要求に応えるため、人と計算機との円滑なコミュニケーションを目指すヒューマン IF の研究が進んでおり、なかでも自然語 IF 研究はその中心となっている<sup>1)-13)</sup>。

筆者らはリレーショナルデータベース (RDB) 検索用自然語 IF の研究を進めており<sup>2),3),8)-12)</sup>、プロトタイプシステム HITNICE-Q (Hitachi Natural Language Interface-for Query) を開発した。これは、検索指示を表す日本語文を解析し、検索指示言語 SQL (Structured Query Language) に変換する。

自然言語で DB 検索システムを効率良く操作するためには、入力文の解釈結果を利用者に報知し、確認・修正を可能とすることが効果的である<sup>11),12)</sup>。報知方法としては、コマンドや中間表現、表、自然言語文などがある。HITNICE-Q では、利用者との親和性が最も高く、内容理解が容易な自然言語文 (以下、確認文と呼ぶ) で報知する方式を採用し、実現した。

確認文の利用効果は、(A)解釈結果を確認でき、

<sup>†</sup> (株)日立製作所 システム 開発研究所 関西 システムラボ  
ラトリ  
Kansai Systems Laboratory, Systems Development  
Laboratory, Hitachi Ltd.

(B)システムの受理可能な語彙や構文を習得でき、(C)確認文を入力文として流用でき、(D)確認文を合成音声や点字などに変換して出力できることである<sup>10)</sup>。本論文では、入力文の解釈結果を表す確認文を生成する方式を提案し、評価している。

DB 検索用の日本語 IF システムに関しては笠<sup>6)</sup>や谷<sup>7)</sup>らの報告がある。しかし、これらの中では確認文の生成方式については述べられていない。

また、形式言語から確認文を生成する研究としては上林らの報告がある<sup>13)</sup>。上林らは、制限された英文の集合を定義し、質問内容に応じて変換・連結といった文字列操作を施すことにより質問の解となるデータの意味に対応する英文を生成している。

しかし、上林らの方式を日本語文生成に適用する場合には、語順(係り受け)や、助詞・助動詞の使い分けなど、日本語特有の表現への対処が必要である。これらの対処は、自然で滑らかな日本語文を生成するためには不可欠な課題である。上林らの報告では、フレーズのおのおのを自然言語によって自然で滑らかに表現する方式については提案しているが、上記の特有表現への対処を含んだ日本語フレーズ同士の連結方式(日本語文の組み立て方式)については言及していない。

また、上林らの方式は、利用者が形式言語を入力することを想定しており、生成される英文の構造は形式言語の構造を反映している。すなわち、利用者が DB 構造を理解しているという前提に立っている。しかし、自然言語文を入力とする DB 検索では、利用者が DB の構造を熟知していないことを想定する必要がある。したがって、利用者が確認文を理解し、システムの解釈結果が自分の意図に合致しているかを確認しやすくするためには、DB 構造を意識しない、入力日本語文の構造(検索条件の記述順序など)を反映させた確認文を生成する必要がある。その一方で、解釈結果を正確に利用者に報知するために、入力文において省略された情報を補填し、曖昧性のない確認文を生成する必要がある。

これに対して、本論文では、自然で滑らかな日本語文を生成する方式を提案しており、日本語フレーズの表現方法および日本語フレーズ同士の連結方式の両方について言及している。具体的には、属性項目間の関係の記述方法(3.1.1項)や、検索条件句の記述順序の決定方法(3.1.2項, 3.2.4項)、検索対象の決定方法(3.2.1項)、検索条件句の組み立て方法(3.2.6項)

などに日本語の特徴を反映させている。また、DB 構造を熟知していない利用者でも容易に理解し、確認できるように、入力日本語文の構造(記述順序)を反映させた確認文を生成している(3.2.1項, 3.2.4項)。

すなわち、本論文では、(1)特定の二属性項目間の関係を定義した自然言語表現を必要に応じて適宜連結することにより、DB の構造を利用者に意識させることなく任意の二属性項目間の関係を自然で滑らかに表現し、(2)入力文の記述順序に則った文を生成することにより、入力文との対応付けがしやすい確認文を生成する。また、曖昧性のない確認文を生成するために、(3)属性項目名の省略された属性データに対して属性項目名に相当する自然言語を補完し、(4)検索対象が明示されていない入力文に対して検索対象情報を補完するという四つの特徴を持つ確認文生成方式を提案している。

以下、2章では HITNICE-Q の概要を、3章では確認文生成アルゴリズムを、4章では本方式の評価実験および評価結果を述べる。

## 2. HITNICE-Q の概要

HITNICE-Q では、入力日本語文と RDB の構成要素との対応関係を解析し、SQL を生成する。

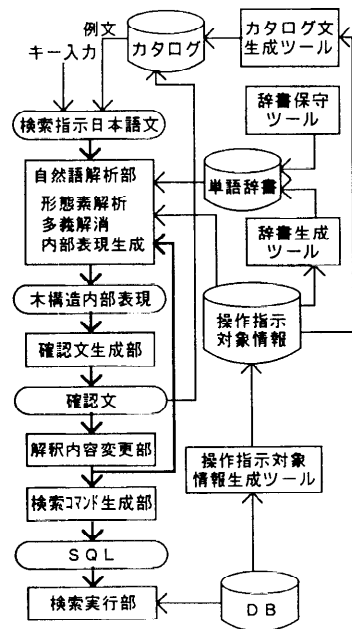


図1 自然語インタフェース HITNICE-Q の構成  
Fig. 1 System configuration of HITNICE-Q.

2.1 システム構成

HITNICE-Q のシステム構成を図1に示す。

入力文はキーボードから入力するか、カタログから文を選択することにより入力する。カタログには、例文や過去の確認文を蓄積する。例文は、カタログ文生成ツールを用いて作成することができる。

自然語解析部ではまず、形態素解析により入力文を単語に分割し、単語辞書から単語情報を取得する。次に、多義解消により、近接する単語の持つ情報を基に、各単語と RDB の構成要素を 1 対 1 に対応付けることによって、語彙の曖昧性を解消する<sup>9)</sup>。さらに、内部表現生成により単語間の関係を解析し、係り受けの曖昧性を解消し、内部表現を生成する。内部表現は検索コマンドに 1 対 1 に対応する木構造の中間表現である。すなわち、入力文に含まれる曖昧性は内部表現生成までで解消される。

確認文生成部で内部表現を入力として確認文を生成し、利用者に報知する。必要に応じて解釈内容変更部で修正エディタにより解釈結果を修正する<sup>12)</sup>。

検索コマンド生成部で内部表現を入力として SQL を生成し、検索実行部で SQL を実行し検索結果を利用者に報知する。

HITNICE-Q では、対象 DB 世界を表階層モデル<sup>2)</sup>で表す。表階層モデルは、検索指示を表す自然語文によって DB 検索を行う自然語インタフェースを実現するためのモデルである。すなわち、自然語と RDB の構成要素との対応関係を計算機に理解させるためのモデルである。表階層モデルは、RDB であれば適用することができる。

表階層モデルは、図2に示すように、関係表を構成する属性項目とその下位の関係表との階層的連結情報を用いて、複数の関係表の間の意味的階層関係を表現する。そして、この階層関係を参照して、入力自然語文をボトムアップに解析することによって、入力文と RDB の構成要素との対応関係を認定している。

表階層モデルを適用する際には、表1に示す操作指示対象情報、および、表2に示す階層的連結情報を定義する(表1、表2は図2に対応する)。これらの情報は、操作指示対象情報生成ツールによりシステム管理者が定義する。

単語辞書は、各単語の言語情報・DB 情報を格納する。これらは、辞書作成・保守ツールを用いて作成・保守する。

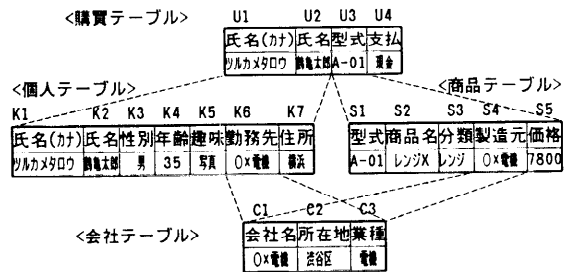


図2 表階層モデルの概念図  
Fig. 2 The concept of hierarchical-tree structure model.

表1 操作指示対象情報  
Table 1 The information of DB world.

意味コード	正規語	テ-タイプ	フォーマット 検索対象	テーブルの名前	関係表名称
U 1	氏名(カナ)	文字	O F F	購買テーブル	購買者
U 2	氏名	文字	O N	購買テーブル	購買者
U 3	型式	文字	O F F	購買テーブル	購買者
U 4	支払	文字	O F F	購買テーブル	購買者
S 1	型式	文字	O F F	商品テーブル	商品
S 2	商品名	文字	O N	商品テーブル	商品
S 3	分類	文字	O F F	商品テーブル	商品
S 4	製造元	文字			

表2 階層的連結情報  
Table 2 The information of hierarchical-tree structure model.

上位テーブル名	下位テーブル名	上位テーブル 接 続 ID	下位テーブル 接 続 ID
購買テーブル	個人テーブル	U1	K1
購買テーブル	個人テーブル	U2	K2
購買テーブル	商品テーブル	U3	S1
個人テーブル	会社テーブル	K6	C1
商品テーブル	会社テーブル	S4	C1

接続 ID は、表1の意味コードに対応。

2.2 木構造内部表現

2.2.1 基本構造

内部表現は木構造で、表階層モデルに対応した階層構造をなし(図3)、射影、選択、表結合、検索結果編集、副照会、論理演算などを表現できる。

検索結果編集とは、検索した結果に対する加工処理であり、平均や件数の計算、ソーティングなどがある。

副照会<sup>14)</sup>とは、ある検索結果を用いて別の検索を行うときの、最初の検索のことである。副照会に対して、後の検索を主照会と呼ぶ(2.2.2 項参照)。

内部表現は、次の10種類のノードで構成する。

- (1) 文ノード(S)

内部表現の最上位ノード、

(2) 検索単位ノード (unit)

一検索単位に相当する部分木の最上位ノード、

(3) 検索ノード (rtv)

検索条件と検索対象を表す部分木の最上位ノード、

(4) 検索結果編集ノード (proc)

検索結果編集を表す部分木の最上位ノード、

(5) 表ノード (tbl)

一つの関係表に関する検索条件および検索対象を表す部分木の最上位ノード、

(6) 属性項目名ノード (attr)

関係表の属性項目名を表すノード、

(7) 属性データノード (data)

関係表の属性データを表すノード、

(8) 範囲ノード (btwn)

属性データの範囲を規定するノード、

(9) 論理演算ノード (ope)

複数の属性データの論理演算関係 (and/or) を規定するノード、

(10) 検索結果編集分類コード (kind)

検索結果編集の種類を表すノード、

図3に示すように、表結合は属性項目名ノードの下に表ノードを子ノードとして連結して表現する。関係表の上下関係は表階層モデルに対応する。

副照会は、主照会の属性項目名ノードの下に副照会の部分木を子ノードとして連結して表現する。

検索結果編集は、検索結果編集分類ノードの下に編集対象となる属性項目名ノードを子ノードとして連結

して表現する。

属性項目名ノードおよび属性データノードは、それぞれ次の4種類のデータを参照できる。

(1) 意味コード

各ノードに対応する属性項目を表すコード(図2, 表1参照)。関係表を識別するアルファベットと関係表の属性項目を識別する数字からなる。

複数の意味コードを持つ(語彙的曖昧性を持つ)単語もある。この場合、多義解消で一つに決定する。

(2) 正規語

DBの構成要素に1対1に対応する自然言語表現。正規語を用いて確認文を表現することにより、確認文における語彙的曖昧性を解消する。意味コードと正規語は単語辞書に格納される。

(3) 文節番号

入力文の文頭から何番目の文節に対応するかを表す。

(4) 関係指示情報

大小関係(〜以上)や集合関係(〜以外)を表す。

### 2.2.2 生成例

図2に対する次の例文を考える。なお、本論文で用いる例文は、すべて図2のRDBを対象としている。HITNICE-Qの処理系は、DBの内容に依存しない、汎用性を保っている。

(例1) 一番高いビデオの値段以上のテレビで、東京か神奈川の人が現金で買ったテレビの平均価格は?

この例は、「一番高いビデオの値段」を検索する副照会と、「その値段以上のテレビで、東京か神奈川の人現金で買ったテレビの価格」を検索する主照会からなる。HITNICE-Qでは、入力文における「属性項目名+関係指示情報」「数値でない属性データ+関係指示情報」などの表現に着目し、その表現よりも前方の部分副照会に関する記述とし、後方の部分主照会に関する記述として区別することによって、内部表現を生成している。

例1に対する内部表現を図4に示す。各単語の持つ意味コードを参照することによって、「東京」と「神奈川」は「個人の住所」を表す属性データノードに対応し、「人」は属性項目名ノード「氏名」に対応すると解釈し、これらのノードを表ノード(個人)の下に連結する。同様に、「現金」「買った」「テレビ」「価格」についても、それぞれ対応する表ノード(購買, 商品)の下に連結する。次に、表ノードを最上位ノードとす

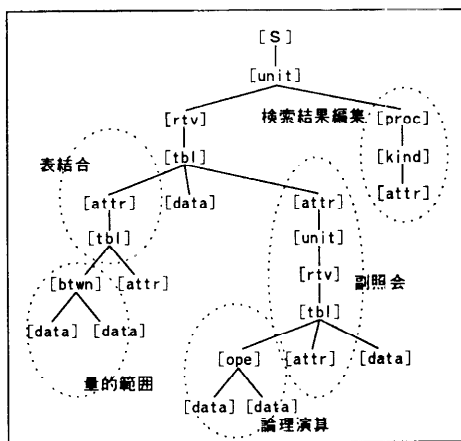


図3 木構造内部表現の構造

Fig. 3 Intermediate expression based on tree-structure.

るまともり同士を表2の階層的連結情報に基づいて連結する。

一方、副照会「一番高いビデオの値段」については、対応する部分木を属性項目ノード<14>「価格」の下に連結する。また、「一番高い」、「平均」は検索結果編集を表すと解釈し、編集対象の属性項目名ノード「価格」を、検索結果編集分類ノードとともに検索結果編集ノードの下に連結する。

### 3. 確認文生成方式

#### 3.1 確認文の統語構造

確認文は、利用者の入力した検索指示文をシステムがどう解釈したかを利用者へ報知するためのものである。したがって、確認文は曖昧性がなく、読みやすい文であることが肝要である。そこで、これらの具備要件を満たした以下の四つの特徴を備えた確認文生成方式を提案する。

- (1) 任意の属性項目間の構文関係の定義
- (2) 入力文に則った検索条件の組み立て
- (3) 省略された属性項目名の補完
- (4) 検索対象の補完

#### 3.1.1 任意の属性項目間の構文関係の定義

属性項目間の構文関係を自然で滑らかに表現するこ

とにより、読みやすい確認文を生成できる。例えば、図2のDBで、属性項目「商品名(S2)」と属性項目「業種(C3)」との間の関係を、「商品を製造した会社の業種」あるいは「業種がXである会社が製造した商品」などと表現すれば、DBの構造や検索コマンドを熟知していない利用者でも容易に理解できる。

DBの任意の二つの属性項目の間の構文関係を効率良く定義するために、(1)同一表における属性項目と関係表名称の間の構文関係と、(2)表結合する関係表名称間の構文関係を定義し、これらを組み合わせて任意の属性項目間の構文関係を表現する。ここで関係表名称とは、一つの関係表の内容を総括して表す単語であり、次の類型

“[属性項目名]が[属性データ]である[X]”を想定した場合の変数Xに当てはまる単語である。例えば図2の商品テーブルでは「商品」、会社テーブルでは「会社」といった関係表名称を定義できる。

(1)の「同一表における属性項目と関係表名称の間の関係」は、用言を介して定義する。定義例を表3に示す。例えば、購買者テーブルの属性項目「支払」については表3より、「現金で支払った購買者」などと表現できる。表3に定義されない属性項目については

表3 属性項目と関係表名称間の構文関係の定義

Table 3 Syntax between attribute and table name.

属性項目名	属性項目データ(例)	構文関係	関係表名称
支払	現金	で支払った	購買者
型式	A-01	を購入した	購買者
製造元	○×電機	が製造した	商品
趣味	釣り	が好きな	人
勤務先	○×電機	に勤務する	人
住所	東京	に住む	人
所在地	東京	にある	会社

表4 関係表名称間の構文関係の定義

Table 4 Syntax between table names.

関係表名称1	構文関係	関係表名称2
購買者	が購入した	商品
商品	を購入した	購買者
購買者	(定義せず)	人
商品	(定義せず)	購買者
会社	を製造した	会社
会社	が製造した	商品
人	が勤務する	会社
会社	に勤務する	人

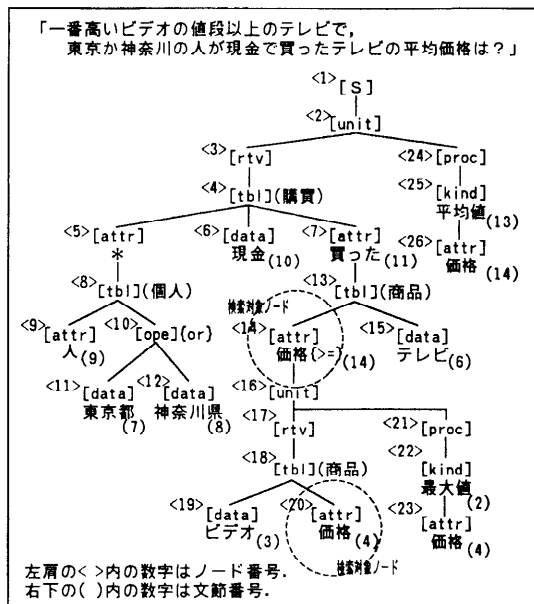


図4 (例1)に対する内部表現  
Fig. 4 The example of intermediate expression.

次のデフォルトの構文関係を用いて表現する。

“[属性項目名]が[属性データ]である[関係表名称]”

例えば、商品テーブルの属性項目「価格」については、「価格が5000円である商品」などと表現する。

(2)の「表結合する関係表名称間の関係」は、用言を介して双方向に定義する。定義例を表4に示す。ただし、図2の購買テーブル(関係表名称「購買者」と個人テーブル(関係表名称「人」)の間に見られるように、互いに同格関係にある場合は、用言を介さなくても関係表名称の間の関係を表現できるので定義しなくてもよい(あえて定義するならば、「人である購買者」「購買者である人」となる)。

表3および表4の構文関係を組み合わせることにより、任意の二つの属性項目間の関係を表現できる。例えば図2の属性項目「住所」と属性項目「価格」の間の関係として、「Xに住む人(である購買者)が購入した商品の価格」、あるいは、「Xに住む人が勤務する会社が製造した商品の価格」と表現できる(どちらの表現を用いるかは入力文による)。

なお、これらの関係表名称および構文関係のデータはシステム管理者があらかじめ定義する。

### 3.1.2 入力文に則った検索条件の組み立て

利用者が、入力文と確認文との対応付けを容易に行えるようにするために、確認文の中で記述すべき検索条件の記述順序を入力文に対応させる。例えば、

(例2) 釣り好きな30才の東京の人は?

における3種類の検索条件(釣りが好き、年齢が30才、東京都に住む)の記述順序を重視し、確認文においても「釣りが好きで、年齢が30才で、東京都に住む」の順で検索条件を記述する。

### 3.1.3 省略された属性項目名の補完

利用者の入力する検索指示文においては、属性データのみを記述してその属性データの所属する属性項目名を省略することがしばしばある。例えば、

(例3) 東京の人の氏名は?

では、属性データ「東京」に対応する属性項目名が明示されていない。したがって、システムが「東京」を図2の「個人の住所」に対応する属性データとして解釈したのか、「会社の所在地」に対応する属性データとして解釈したのかを確認文の中で明確に表現する必要がある。そこで、内部表現に基づいて、属性データに対応する属性項目名を補完することにより、解釈結果を明確に表現する。すなわち、3.1.1項で述べたよ

うに、あらかじめ定義された用言(表3)を補完するか、属性項目名を助詞「が」を介して補完する。例えば、例3の「東京」を「個人の住所」として解釈した場合、確認文では、「東京都に住む」または「住所が東京都である」と表現する。属性データはそれぞれ意味コードを持つので、対応する属性項目名を操作指示対象情報(表1)から容易に取得することができる。

#### 3.1.4 検索対象の補完

利用者の入力する検索指示文においては、検索対象を省略することがしばしばある。例えば、

(例4) 東京に住む男性は?

では、「東京に住む」「男性」はともに検索条件であり、検索対象に関する記述が省略されている。このように検索対象が省略された場合、HITNICE-Qでは内部表現生成時に、システム管理者があらかじめ指定した「デフォルト検索対象」(表1のデフォルト検索対象の値がONになっている属性項目)を検索対象として補完している。確認文生成では、この補完した属性項目名を検索対象とする。例4では、個人テーブルのデフォルト検索対象である属性項目「氏名」を検索対象とする。

## 3.2 確認文生成アルゴリズム

確認文の統語構造をバックス標準形で図5に示す。また、確認文生成アルゴリズムを図6に示す。本節では、前述の例1を用いてアルゴリズムを説明する。

### 3.2.1 検索対象の決定

ここでは、内部表現および入力文から検索対象(SQLにおいて、“select”の直後に記述する属性項目名)を決定する。

ここでは、検索指示を表す日本語文では、文末部分に検索対象を記述することが多いという日本語の特徴に着目している。また、検索対象は属性データを伴わない属性項目名であり、属性データが検索対象となったり、属性データを伴う属性項目名が検索対象となったりすることはない。したがって、内部表現中の属性項目名ノードを対象として、対応する入力文の単語の記述順序を調べることにより検索対象を決定することができる。ただし、例外として、

(例6) 誰がテレビを買ったのか?

のように、疑問代名詞「誰(氏名)」を検索対象とする場合がある。

以上をふまえ、本方式では、次の手順で検索対象を決定する。

(1) 入力文の中に疑問代名詞(「誰」「どこ」「何」

など)がある場合、これに対応する属性項目名を検索対象とし、この検索対象に対応する内部表現中の属性項目名ノードを検索対象ノードとする。

- (2) 内部表現を構成する属性項目名ノードに対応する入力文中の単語のうち、文末に最も近い単語およびそれと並列関係にある単語を検索対象とし、この検索対象に対応する内部表現中の属性項目名ノードを検索対象ノードとする。

なお、副照会を含む場合は、上記手順に従って、主照会とは独立に検索対象を決定する。

例1では、上記(2)により文末の属性項目名「価格」を検索対象とし、図4の属性項目名ノード<14>を検索対象ノードとする。また、副照会「一番高いビデオの値段(ノード<16>以下)」についても、上記(2)により末尾の属性項目名「値段(正規語は価格)」を副照会の検索対象とし、図4の属性項目名ノード<20>を副照会の検索対象ノードとする。

### 3.2.2 検索対象部の生成

確認文は大きく検索条件部と検索対象部と検索結果編集部からなる。検索対象部は、前項で決定した検索対象に関する記述部分を指す。

検索対象部は、前項で決定した検索対象に基づいて次の鋳型を用いて生成する。なお、角括弧[ ]で囲んだ部分は、さらに展開可能な部分であることを示す。

- (1) 主照会で検索結果編集を含む場合  
 “[関係表名称]の[検索対象の正規語]を検索し、”
- (2) 主照会で検索結果編集を含まない場合  
 “[関係表名称]の[検索対象の正規語]を検索せよ。”
- (3) 副照会で検索結果編集を含む場合  
 “[関係表名称]の[検索対象の正規語]の”
- (4) 副照会で検索結果編集を含まない場合  
 “[関係表名称]の[検索対象の正規語]”

検索対象を表す属性項目名が複数ある場合、それらの正規語を助詞「と」を介して連結する。

例1において、主照会の検索対象部は検索結果編集「平均」を含むので、上記(1)により、

“商品の価格を検索し、” (3.1)

と表現する。一方、副照会も検索結果編集「一番高い

```

<確認文> ::= <検索条件部> <検索対象部> <検索結果編集部>
<検索条件部> ::= (<条件句>)
<条件句> ::= <条件句1> | <条件句2> | <条件句3> |
             <条件句4> | <条件句5>
<条件句1> ::= <属性項目名> <属性データ> <関係指示語>
             (<論理演算> <属性データ> <関係指示語>) <付属語>
<条件句2> ::= <属性データ> (<論理演算> <属性データ>)
             <構文関係> <関係表名称> <付属語>
<条件句3> ::= <属性項目名> <属性データ1> から
             <属性データ2> まで <付属語>
<条件句4> ::= <検索条件部> <関係表名称> <構文関係> |
             <検索条件部> <関係表名称> |
             <構文関係> <関係表名称>
<条件句5> ::= <確認文> <関係指示語> である
             <属性項目名> <付属語>
<関係指示語> ::= 以上 | 以下 | より上 | 未満 | 以外
<論理演算> ::= かつ | または
<付属語> ::= で、 | である
<検索対象部> ::=
<関係表名称> の (<属性項目名>) <属性項目名> を検索せよ。 |
<関係表名称> の (<属性項目名>) <属性項目名> を検索し、 |
<関係表名称> の (<属性項目名>) <属性項目名> |
<関係表名称> の (<属性項目名>) <属性項目名> の
<検索結果編集部> ::= <属性項目名> の <編集分類1> を求めよ。 |
<属性項目名> を <編集分類2> ソートせよ。 |
<編集分類1> ::= 平均値 | 最大値 | 最小値 | 件数
<編集分類2> ::= 昇順に | 降順に
    
```

\* : システム管理者による定義情報を利用

図5 確認文の構成

Fig. 5 The syntax of sentence for confirmation.

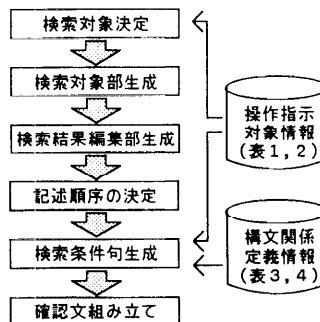


図6 確認文生成手順

Fig. 6 Algorithm of generation of sentence for confirmation.

(最大値)」を含むので、上記(3)により、

“商品の価格の” (3.2)

と表現する。ここで、語彙的曖昧性をなくすために、「値段」は正規語「価格」に置き換えて表現する。

### 3.2.3 検索結果編集部の生成

検索結果編集部は、検索結果に対して行われる加工処理(検索結果編集)に関する記述部分を指す。演算の内容としては、平均値、件数、最大値、最小値、ソートなどがある。

検索結果編集を含む場合、その分類に応じ、次の鋳

型を用いて検索結果編集部を生成する。

(1) 平均値, 件数, 最大値, 最小値の場合  
主照会の場合, 次のように記述する。

“[属性項目名]の[編集の分類]を求めよ。”  
(3.3)

副照会の場合, 次のように記述する。

“[編集の分類]”  
(2) ソートの場合  
“[属性項目名]を[ソートの種類]ソートせよ。”  
(3.4)

例1では, 主照会に価格の平均を計算する検索結果編集ノードがあるので, 上記(1)により, 次のように記述する。

“価格の平均値を求めよ。” (3.5)

また, 副照会に価格の最大値を計算する検索結果編集ノードがあるので, (1)により, 次のように記述する。

“最大値” (3.6)

### 3.2.4 検索条件句の記述順序の決定

検索条件句は, 検索条件部を構成する単位である(図5参照)。ここでは, 検索条件句の記述順序を決定する。この決定は, 検索対象を決定した後に行う。というのは, 検索対象が何であるかによって, 検索条件句の記述順序が変動するからである。

入力文における記述順序に則った, 自然で滑らかな確認文を生成するために, 検索条件句の記述順序を次の手順で決定する。なお, 以下では, あるノードおよびそのノード以下に連結されたノードからなる集合を「群」と呼ぶ。

(1) 最上位の表ノードにおける検索対象ノードを含まない子供群と, 検索対象ノードの兄弟群と, 検索対象ノードの子供群を対象として, 入力文における記述順序の早い順に群を順序付ける。

(2) 最上位の表ノード以外の表ノードの子供群を対象として, 入力文における記述順序の早い順に順序付ける。検索対象ノードを含む群は最後に順序付ける。

なお, 副照会を含む場合, 上記手順に従って主照会とは独立に順序付ける。

例1に対して, 上記(1)を適用する。図4で, 上記(1)を満たす群は, 群5, 群6, 群15, 群16の四つである。文節番号を参照してこれらを記述順序の早い順に順序付けると,

“[群16][群5][群6][群15]” (3.7)

の順となる。また, 上記(2)で, 群5に含まれる表ノード<8>についてその子供群を対象として順序付けると次のようになる。

“[群5]”→“[群10][群9]” (3.8)

一方, 副照会についても同様に順序付ける。上記(1)を満たす群は群19であるので, 副照会の検索条件部は次のようになる。

“[群19]” (3.9)

そして副照会の検索対象部(3.2)および検索結果編集部(3.6)を(3.9)に連結して, 副照会に関する次の確認文を得る。

“[群19]商品の価格の最大値” (3.10)

さらに, ノード<14>の持つ関係指示情報(>=)を参照して副照会を表す群16と主照会の属性項目名ノード<14>の関係を記述することにより, 次の展開式を得る。

“[群16]”→“{[群19]商品の価格の最大値}以上の価格” (3.11)

ここで, 副照会に関する記述の範囲を確認文の中で明確にするために, 副照会に関する記述部分を中括弧{ }で囲むことにしている。

### 3.2.5 検索条件句の生成

ここでは, 記述順序の決定した検索条件句のそれぞれについて, 対応する自然言語表現を生成する。検索条件句の種類に応じて次のように生成する。

#### (1) 属性データノード

属性データノードの属する属性項目について構文関係が定義されている場合(表3), この構文関係を用いて対応する項目属性名を連結する。定義されていない場合, 3.1.1項で述べたデフォルト関係(助詞「が」を介する)により項目属性名を連結する。

例1では, ノード<19>, ノード<6>, ノード<15>が属性データを表すので, 表3の構文関係およびデフォルト関係を用いることにより, それぞれ次のように検索条件句を生成する。

“[群19]”→“分類がビデオである” (3.12)

“[群6]”→“現金で支払った” (3.13)

“[群15]”→“分類がテレビである” (3.14)

ただし, これらの末尾表現については, 検索条件句を組み立てる際に変動しうる(3.2.6項参照)。

#### (2) 範囲ノード

助詞「から」および「まで」を介して属性データを連結する。さらに, 上記(1)と同様に属性項目名を連結する。すなわち, 例えば, 「価格が5000円から1万



円まで」などと表現する。

### (3) 論理演算ノード

接続詞「かつ」「または」を介して属性データを連結する。さらに、(1)と同様に属性項目名を連結する。例1では、論理和を表す論理演算ノード〈10〉があるので、表3により次のように表現する。

“[群10]”→“東京都または神奈川県に住む”  
(3.15)

### (4) 属性項目名ノード

内部表現における、検索対象ノード以外の属性項目名ノードは、関係表名称を表す単語に対応するノードであり、検索条件にも検索対象にも該当しないので、確認文生成では無視する。例1では、関係表名称「人」に対応する属性項目名ノード(群9)が無視される。

“[群9]”→”” (3.16)

(3.8) および (3.11) から (3.16) までの展開式に基づき (3.7) をトップダウンに展開すると次のようになる。

“{分類がビデオである商品の価格の最大値}  
以上の価格東京都または神奈川県に住む現金で支払った分類がテレビである” (3.17)

### (5) 表結合

ある関係表に関する検索条件句(のまとまり)と別の関係表に関する検索条件句(のまとまり)を滑らかに連結するために、両者の関係表の間に成立する構文関係を表す自然言語表現を挿入する(表4)。表4に示すように、関係表間の構文関係は双方向的であるので、どちらの方向の構文関係を適用するのが適切であるのかを判断する必要がある。ここでは、検索対象ノードから距離の遠い関係表ノードから近い関係表ノードへ向かう方向に関する構文関係を用いて記述する。ここで、距離とは、あるノードからあるノードへ行くのに、最短で何本の枝を通ればたどりつけるかを表す。

例1では、表ノード〈8〉とノード〈4〉、表ノード〈4〉とノード〈13〉が表結合している。表ノード〈8〉の方が表ノード〈4〉より検索対象ノード〈14〉からの距離が大きいので、表ノード〈8〉から表ノード〈4〉への構文関係を表す自然言語表現を挿入する。表4に構文関係が定義されていないので、この場合はどちらかの表ノードの関係表名称(「人」または「購買者」)を挿入する。挿入場所は、個人テーブルに関する記述「東京都または神奈川県に住む」と、購買テーブルに関する記述「現金で支払った」との間である。同様

に、表ノード〈4〉から表ノード〈13〉への構文関係を表した表現「購買者が購入した商品」を、購買テーブルに関する記述「現金で支払った」と商品テーブルに関する記述「分類がテレビである」との間に挿入する。その結果、次のようになる。

“{分類がビデオである商品の価格の最大値}  
以上の価格東京都または神奈川県に住む人現金で支払った購買者が購入した商品分類がテレビである” (3.18)

### 3.2.6 確認文の組み立て

ここまでで生成された検索条件句(のまとまり)を連体修飾表現(付属語「である」や連体形用言)あるいは、連用修飾表現(付属語「で、」や連用形用言)で連結することにより、自然で滑らかな検索条件部を生成する。そして、検索条件部、検索対象部、検索結果編集部の順で組み立てて確認文とする。

検索条件句の連結方法について以下に述べる。

まず、各検索条件句(のまとまり)がどの関係表に関するかを内部表現から認定する。ある関係表に関する検索条件句とは、

- (1) その関係表に対応する表ノードの子ノード群のうち、検索対象を含まないもの
- (2) その関係表に対応する表ノードの子ノードが検索対象ノードであり、かつ、その検索対象ノードが子ノード群を持つ場合、その子ノード群
- (3) その関係表に対応する表ノードの上位に結合する表ノードの子ノード群の集まり

のどれか一つを満たすものである。ある関係表に関する検索条件句(のまとまり)が複数存在する場合、入力文における記述順序の最も遅い検索条件句の末尾表現を連体修飾とし、その他の検索条件句の末尾表現は連用修飾とする。検索条件句の末尾が用言である場合、用言を連体形(または連用形)に変換する。それ以外の場合は、付属語「である」(または「で、」)に変換する。

例1について説明する。図4で、購買テーブル(表ノード〈4〉)に関する検索条件句は、上記(1)より、群5(「東京都または神奈川県に住む人」と群6(「現金で支払った」)があるが、群6のほうが記述順序が遅いので、群6を「現金で支払った」と連体修飾で表現し、群5を「東京都または神奈川県に住む人で、」と連用修飾で表現する。

個人テーブル(表ノード〈8〉)については、上記(1)より、群10、群9があり、群9のほうが記述順

序が遅いが、群9は検索対象でない属性項目名ノードであり、(3.16)に示すように無視するので、群10を「東京都または神奈川県に住む」と連体修飾で表現する。

商品テーブル(表ノード<13>)については、上記(2)より群16(「分類がビデオである商品の価格の最大値」以上の価格)、上記(1)より群15(「分類がテレビである」)、そして、上記(3)より購買テーブルの子ノード群の集まり(「東京都または神奈川県に住む人で、現金で支払った購買者が購入した商品」)がある。この中で、最も記述順序が遅い群15を「分類がテレビである」と連体修飾で表現し、他の二つを「{分類がビデオである商品の価格の最大値}以上の価格で、」「東京都または神奈川県に住む人で、現金で支払った購買者が購入した商品」と連用修飾で表現する。

以上により、(3.18)から、次の検索条件部を得る。

“{分類がビデオである商品の価格の最大値}

以上の価格で、東京都または神奈川県に住む人で、現金で支払った購買者が購入した商品で、分類がテレビである” (3.19)

そして、これに検索対象部(3.1)、検索結果編集部(3.5)を連結し、最終的に次の確認文を得る。

“{分類がビデオである商品の価格の最大値}

以上の価格で、東京都または神奈川県に住む人で、現金で支払った購買者が購入した商品で、分類がテレビである商品の価格を検索し、価格の平均値を求めよ。” (3.20)

## 4. 評価実験

### 4.1 実験方法

本方式に対する次の評価実験を行った。

#### (1) 対象 DB

図2の商品購買に関するRDB。

#### (2) 入力

本研究とは直接関係のない5人の研究者から収集した43例文を用いた。

HITNICE-Qは、条件構成の違いによる次の6種類の表現の組み合わせからなる入力自然語文をSQLに変換することを実現している。

- ①射影：(例)「商品名と価格は」
- ②選択：(例)「横浜に住む人は」
- ③検索結果編集：(例)「社員の平均年齢は」
- ④副照会：(例)「A社の洗濯機より安い製品は」
- ⑤表結合：(例)「神奈川県に住んでいる人でA社の

製品を買った人は」

- ⑥論理積和：(例)「20才から25才までの男または女で横浜に住む人は」

本実験では、HITNICE-Qが解析をサポートしていない表現を含む入力文についてはあらかじめ除外した。すなわち、本実験で用いる43例文は、すべて上記6種類の表現が組み合わさってできている日本語文である。

- (3) DB 定義情報：表1, 表2, 表3, 表4に従う。本実験では、次の三つの観点に基づいて評価する。

- ①正確性：属性項目間の関係を明確にし、省略情報を補完できたか否かで判定。
- ②忠実性：記述順序が対応しているかで判定。
- ③再現性：確認文を入力文とした場合に、その確認文が入力文と一致するかで判定。

ここで、再現性について述べる。1章で述べたように、確認文の利用効果の一つに、確認文を入力文として再利用できるという効果がある。この効果により、入力速度の向上と、解釈結果の修正効率の向上が期待できる。確認文に曖昧性がなければ、確認文を入力としたときにシステムは一通りに解釈することができ、生成される確認文は入力文と同一の文となるはずである。以上のことから、本実験の評価項目の一つに再現性を加え、入力文として再利用できるかどうかを評価した。

### 4.2 実験結果および考察

実験で用いた入力文およびその確認文の生成結果の一部を図7に示し、その評価結果を表5に示す。

#### 4.2.1 正確性

43例文中42例文について、属性項目間の関係を明確にし、属性データに対応する属性項目名や検索対象情報を補完した確認文を生成できた。

確認文を正しく生成できなかった、次の例文を考える(図7参照)。

(例7) スポーツが好きな人の趣味は。

この例では、検索対象は属性項目「趣味」となるはずである。しかし、「スポーツ」が属性項目「趣味」に属する属性データであるため、内部表現生成において、属性項目「趣味」が属性データ「スポーツ」と結合して属性データノードを形成し、項目名ノード「趣味」が存在しなくなり、「趣味」を検索対象として決定することができなかった。属性項目名を表す単語が検索条件句の一部であるのか検索対象であるのかを正しく認定するための処理方式を今後さらに検討する必

要がある。

4.2.2 忠実性

正確な確認文が生成された 42 例文中、41 例文について、入力文の記述順序に対応した確認文を生成できた。対応しなかった次の例文を考える (図 7 参照)。

(例 8) 川崎にある会社の 100,000 円以下の電気製品を知りたい。

商品テーブルには、各商品が電気に関する商品であるか否かを判定するための属性項目がないので、「電気製品」という入力は「業種が電気である会社が製造した商品」と解釈される。しかし、検索対象が「製品 (正規語は商品名)」に決定され、会社テーブルに関する記述が文頭に存在することから、会社テーブルに関する検索条件句を商品テーブルよりも先に記述することになり、記述順序が逆転してしまった。

4.2.3 再現性

43 例文に対する確認文を入力文とした結果、生成された確認文はすべて入力文と一致した。したがって、本論文で述べた方式で生成した確認文を入力文として再利用することが有効であるという良いと考える。

4.2.4 実験のまとめ

実験結果から、射影、選択、検索結果編集、副照会、表結合、論理積和を表す検索指示文に対する確認文を生成する本方式が有効であるという見通しを得た。

また、確認文を入力文として再利用することが有効であるという見通しも得た。

今後は、否定文 (全否定、部分否定) や、「少なくとも〜」「すべての〜」「〜と同じ」等の表現を含む検索指示文に対する解析能力 (内部表現生成能力) を向上させ、これに対応する確認文の生成能力をも向上していく必要がある。

5. おわりに

自然言語で DB 検索システムを効率的に操作するために、入力文の解釈結果を表す自然言語文 (確認文) を生成する方式を提案し、HITNICE-Q 上で実現した。省略情報を補い、曖昧性を解消し、また、利用者の馴染みやすい確認文を生成するために、(1)特定の二属性項目間の関係を定義した自然言語表現を入力文に応じて適宜連結することにより任意の二属性項目間の関係を表現し、(2)入力文の記述順序に則った文を生成し、(3)属性項目名の省略された属性データに対

(1) 日立の社員の平均年齢は	(検索結果編集)
→日立製作所に勤務する人の年齢を検索し、年齢の平均値を求めよ。	
(2) 日立の洗濯機よりも安い製品を買ったことのある人は	(副照会)
→ (日立製作所が製造した商品で、分類が洗濯機である商品の価格) 以下の価格である商品を購入した購買者の氏名を検索せよ。	
(3) 日立に勤務する人が買った製品は	(表結合)
→日立製作所に勤務する人が購入した商品の商品名を検索せよ。	
(4) 20才から25才までの男または女で横浜に住む人は	(論理演算)
→年齢が20才から25才までで、性別が男または女で、横浜市に住む人の氏名を検索せよ。	
(5) 川崎にある会社の10万円以下の電気製品を知りたい	
→川崎市にあり、業種が電気である会社が製造した商品で、価格が10万円以下である商品の商品名を検索せよ。	
(検索条件句の記述順序が入力文と異なる)	
(6) スポーツが好きなの人の趣味は	
→スポーツが好きなの人の氏名を検索せよ。(意味が異なる)	

図 7 評価実験で用いた検索指示例文の一部およびその確認文 Fig. 7 Sample sentences and sentences for confirming analysis result.

表 5 確認文生成結果 Table 5 Result of generating sentences for confirmation.

意味的に正しい確認文を生成した例文		正しい確認文を生成できなかった例文
入力文の記述順序に対応したもの	入力文の記述順序に対応しないもの	
41 文 (96%)	1 文 (2%)	1 文 (2%)

して属性項目名に相当する自然言語を補完し、(4)検索対象が明示されていない入力文に対して検索対象情報を補完するという四つの特徴を持つ確認文生成方式を提案した。

また、一般利用者から収集した 43 例文を基に、正確性、忠実性、再現性の三つの観点から本方式を評価し、(1)条件構成の違いにより 6 種類に分類できる検索指示文 (射影、選択、表結合、検索結果編集、副照会、論理積和表現) のそれぞれに対して、入力文の記述順序に対応した確認文を生成でき、また、(2)確認文を入力文として流用できることを確認した。

今後の課題としては、自然言語解析能力 (内部表現の記述能力) を向上させ、より柔軟な自然言語入力に対応する確認文を生成可能にする必要がある。具体的には、否定文 (全否定、部分否定) や、「少なくとも〜」「すべての〜」「〜と同じ」等の表現を含む検索指示文に対する確認文を生成可能とすることにより、確認文の表現能力を向上していく。また、DB 検索以外の操作指示文に対しても確認文を生成可能とすべく、本方式を汎用化していく。

謝辞 本研究の機会を与えて下さった (株)日立製作所システム開発研究所前所長の堂免信義技師長、本研究を進めるにあたり御協力をいただいた、同所主任研

究員の辻洋博士をはじめとする研究室の方々に深く感謝致します。

### 参考文献

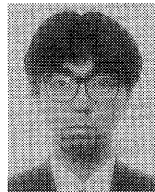
- 1) 藤崎哲之助, ほか: データベース照会システム「ヤチマタ」と名詞句データ模型, 情報処理学会論文誌, Vol. 20, No. 1, pp. 77-84 (1979).
- 2) 絹川博之: 表階層モデルに基づく自然語インタフェース処理方式, 情報処理学会論文誌, Vol. 27, No. 5, pp. 499-509 (1986).
- 3) 難波康晴, ほか: 機能連鎖構造に基づく自然語インタフェース構築ツール, 情報処理学会論文誌, Vol. 32, No. 9, pp. 1180-1189 (1991).
- 4) 中川 優, ほか: 日本語データベース検索システムにおける意味理解方式, 情報処理学会論文誌, Vol. 27, No. 11, pp. 1069-1076 (1986).
- 5) 牧之内顕文, ほか: 移行性のあるデータベース自然言語インタフェース, 情報処理学会論文誌, Vol. 29, No. 8, pp. 749-759 (1988).
- 6) 笠 晃一, ほか: 自然言語問合せ文の意味表現方法とその応用, 情報処理学会論文誌, Vol. 34, No. 5, pp. 925-933 (1993).
- 7) 谷 幹也, ほか: 自然言語インタフェース構築キット: IF-Kit, 電子情報通信学会研究報告, NLC91-62, pp. 25-32 (1992).
- 8) 間瀬久雄, ほか: 自然語インタフェースにおけるインタラクティブ型多義解消方式の開発, 第44回情報処理学会全国大会論文集(3), 7P-5, pp. 235-236 (1992).
- 9) 木山忠博, ほか: 自然語インタフェースにおける語句学習方式, 第44回情報処理学会全国大会論文集(3), 7P-6, pp. 237-238 (1992).
- 10) 森木紀恵, ほか: 自然語インタフェースのための音声ガイダンス, 第44回情報処理学会全国大会論文集(3), 7P-7, pp. 239-240 (1992).
- 11) 間瀬久雄, ほか: 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, 第45回情報処理学会全国大会論文集(3), 2F-7, pp. 135-136 (1992).
- 12) 木山忠博, ほか: 自然語インタフェースにおける対話型解釈内容変更方式, 第45回情報処理学会全国大会論文集(3), 2F-6, pp. 133-134 (1992).

13) 上林弥彦, ほか: 基本関係データベース演算による自然言語表現の変換, 情報処理学会論文誌, Vol. 30, No. 10, pp. 1316-1323 (1989).

14) 平尾隆行: SQL 入門, オーム社 (1992).

(平成5年3月22日受付)

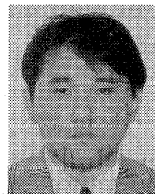
(平成6年4月21日採録)



間瀬 久雄 (正会員)

1965年生. 1988年名古屋大学工学部電気工学科卒業. 1990年同大学院工学研究科情報工学専攻修士課程修了. 同年(株)日立製作所入社. 以来, 自然言語処理の研究に従事.

現在, 同社システム開発研究所関西システムラボラトリに所属. 人工知能学会会員.



木山 忠博 (正会員)

1966年生. 1984年長崎工業高等学校情報技術科卒業. 同年(株)日立製作所入社. 1986年より1987年まで, 同社日立京浜工業専門学院ソフトウェア工学科に在学. 以来, 同社

システム開発研究所において自然言語処理, 自動インデクシングなどの研究開発に従事. 現在, 同社コンピュータ事業本部公共情報事業部に所属.



絹川 博之 (正会員)

1947年生. 1970年東京大学理学部数学科卒業. 同年(株)日立製作所入社. 以来, 漢字・日本語情報処理システム, 仮名漢字変換, 自動インデクシング, 日本語文書処理, 自然

言語処理などの研究開発に従事. 1986年度情報処理学会論文賞受賞. 理学博士. 現在同社システム開発研究所関西システムラボラトリ長. 電子情報通信学会, 計量国語学会, ACL 各会員.