

文献データベースにおける対話型問い合わせ理解とその評価

加納 康男 岸野 文郎

ATR 通信システム研究所

利用者との対話に基づき、あいまいな問い合わせを理解する文献データベースの知的インタフェースの構成とその評価について述べる。

本システムの特徴は、次の通りである。: (1) 利用者との対話によりあいまいな問い合わせを明確化するとともにその意味構造をユーザモデルに構築する (2) 同じ利用者の複数の検索要求を区別して理解する (3) ユーザモデルの構造を用いて論理式を合成する (4) 文献の再現率を改善するため、検索文献の中から利用者が選択した文献に付与されたキーワード情報をユーザモデルにフィードバックする。

対話とフィードバック処理によって効率的な問い合わせ理解ができることをキーワードおよび文献の検索効率によって評価した。

A Query Understanding with Dialogues for Document Database and Its Evaluation

Yasuo Kanou Fumio Kishino

ATR Communication Systems Research Laboratories

Sanpeidani, Inuidani, Seika-cho, Soraku-gun, Kyoto 612-02, Japan

An intelligent interface for a document database which understanding ambiguous queries based on dialogues and its evaluations are discussed.

This system's features are as follows: (1) A user model which represents semantic structures of the same user's simultaneous multiple queries is constructed through dialogues. (2) Boolean expressions are constructed using the semantic structure in the user model. (3) Keywords included in the correct documents are analyzed and important keywords not extracted by the dialogues are extracted to improve document recall.

Those query understanding mechanisms were shown to be effective by the evaluation with not only document retrieval efficiencies but also keyword efficiencies.

1 まえがき

近年、データベースやワークステーション等の普及によって、多くの利用者が関係データベースや、記事・科学文献の検索を行なえるようになってきた。しかしながらこれらのデータベースを検索するためには、その利用技術やデータベースに関する専門的な知識や仲介者が必要である。より多くの利用者が、データベース検索を行なうために、自然言語で問い合わせられた検索要求を理解するシステムが求められている。

しかしながら、特に文献検索を行なう場合、一般に利用者の検索要求はかなり漠然とした段階にある [4]。従って、文献検索システムに対して利用者が自然言語によって検索要求を行なう場合、検索に必要な概念を言葉で完全に表現することは難しい [1][2][3]。また利用者は、文献データベースにつけられたキーワードではなく自分自身の言葉で検索要求を行う。

このため、問い合わせ文において利用者の用いた用語だけでは、検索要求に適合したキーワードを得ることは困難である。文献検索システムの自然言語インタフェースには、このような不十分性やあいまい性を持った問い合わせ文を理解することが要求される。

我々はこれまで、問い合わせ文を理解して検索要求に相当する意味構造を表現し、その意味構造より論理式を合成して所望の文献を検索する文献検索システムを構築してきた [11]-[14],[16]-[19]。

ここでは、問い合わせ文を理解するためのアプローチとしてユーザモデルを導入している。ユーザモデルは、利用者ごとの検索要求を、概念名とその間の関係で表わした意味構造で表現するものである。システムは、利用者各々に対して、1つのユーザモデルを持つ。ユーザモデルを用いることによって、利用者独自の言葉を理解するとともに問い合わせ文に現れなかった検索要求を推論して補充することが可能となる。

本稿では、

(1) ユーザモデルの構築方法

ユーザモデルを、利用者との対話および利用者が選択した文献に付与されているキーワード情報（フィードバック情報）から構築する方法。

(2) 問い合わせ理解

ユーザモデルを用いて、利用者のあいまいな問い合わせ文から、これに対応する必要十分な検索概念（キーワード）を抽出する方法。

(3) 論理式合成方法

ユーザモデル内の意味構造を用いて、問い合わせ文に対応する最適な論理式を合成する方法。

を示す。

また、本文献検索システムに対しては、従来より文献検索の指標として用いられている文献検索効率だけではなく、問い合わせ理解に対する指標としてより直接的なキーワード検索効率を用いた評価を行ない、その有効性を示す。

本システムでは問い合わせ文に含まれるあいまい性のうち、次のものを対象とする。

- (1) 利用者独自の言葉、あるいは特有の言葉の使い方
利用者の問い合わせ文には、システムの知識にある言葉（キーワード）以外の独自の言葉を含むことがある。またキーワードであっても、利用者特有の意味で使用することがある。
- (2) 限定の不足
利用者が検索要求を問い合わせ文で十分に限定しない。
- (3) 過剰の限定
利用者が問い合わせ文で使った言葉の意味を広く解釈する方が望ましい場合がある。例えば、“Dempster Shafer 理論”と言ってもこれ以外の不確実性を表現するものを含むことがある。
- (4) 複数の検索要求を同時に行なう
同じ利用者であっても、異なる意味構造を持った複数の検索要求を同時に行なうことがある。

2 文献検索のモデル

あいまいな問い合わせ文を許す知的インタフェースを構成するために、我々は次のようなアプローチを採用した。図1に本文献検索システムの構成図を示す。

2.1 ユーザモデルの導入

ユーザモデルは、利用者の検索要求の意味構造および利用者独自の言葉の意味を表現する。同じ利用者の複数の問い合わせに対する複数の意味構造は、同一のユーザモデル内に互いに独立に存在する。

2.2 問い合わせ理解とユーザモデルの構築

問い合わせ文の理解は、ユーザモデルにおいて検索要求に対応する意味構造を同定または構築し、その意味構造に含まれるキーワードを抽出することとみなす。ユーザモデルにおける意味構造は、次の3つの手段によって構築される。

(1) 問い合わせ文からの意図理解

問い合わせ文に直接表明されている概念を用いて、ユーザモデル内にある複数の意味構造の中から対応する意味構造を同定する。

(2) 対話による意図理解

システムと利用者との対話を通して、検索要求を明確化すると共に、検索要求をユーザモデル内の意味構造に反映する。

(3) 検索文献からの意図理解

検索文献の中から利用者が選択した文献に付与されたキーワード情報をユーザモデルの構造にフィードバックする。

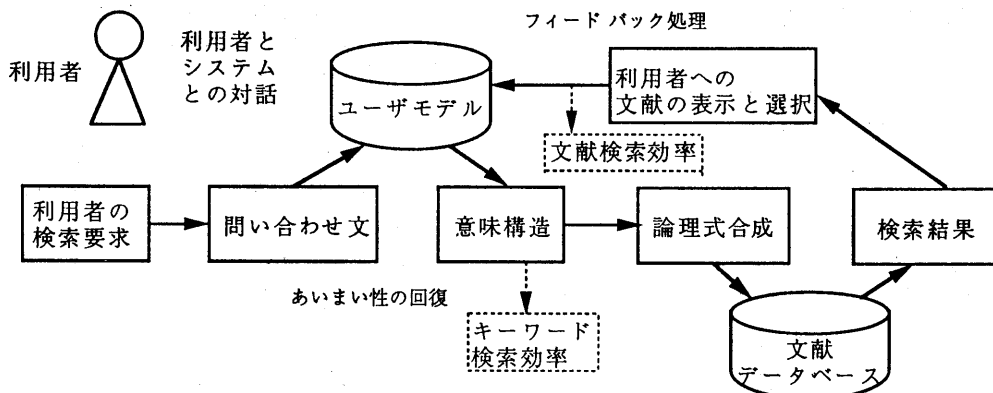


図 1: 文献検索システムの構成

表 1: 問い合わせ文中の名詞間の関係の分類 (例)

関係の種類	関係を表す言葉
階層関係	を含む などの としての 特に 例えば
対象限定	の での における への に対する
手段限定	による を用いた を使った を使用した
目的限定	のための に用いる
因果関係	による
関連関係 (併記)	と および や または あるいは

従来提案されているフィードバック処理では、検索結果を見て利用者自身が問い合わせ文(あるいは検索式)を修正するものである[7]。これに対し、この構成法では、検索文献からの自動的な意図抽出を狙っている。

2.3 ユーザモデルの構造を用いた論理式合成

ユーザモデル内の意味構造を利用して論理式を合成し、検索を実行する。

意味構造に含まれるキーワードおよびリンクの種別と構造を利用して得られる制約条件(AND条件)の最もきつい論理式を利用者の検索要求に最適なものとする。この論理式を用いてデータベースを検索した結果の文献集合が、検索適合率の最大となるべき文献集合である。このとき、条件の制約が大きく文献が検索できない場合は、意味構造を用いて論理式のAND条件を緩和した論理式を合成することができる。

3 ユーザモデルを用いた問い合わせ理解

3.1 問い合わせ文の構文解析

利用者の問い合わせ文は「(名詞句表現)に関する文献が欲しい」という形式で表現されるものとする。この名詞

句表現を、表1に従って、概念を表現している2つの名詞とその間の関係構造を抽出し、3項形式で表現する。例:「意味ネットワークによる対象世界モデルの表現」

(手段限定 (対象限定 表現 対象世界モデル) 意味ネットワーク)

3.2 ユーザモデルの表現方法

ユーザモデルは、名詞で表された概念ノードと概念ノード間の関係を表すリンクを持つ意味ネットワークで表現する。概念ノードには、検索キーワードに対応するキーワードノードと、独自語に対応する独自語ノードを設ける。

リンクは、次のように表現される。

$((ノード N_p) (ノード N_q) (関係の種類 R) (意図のレベル L) (意味構造識別子 I_i))$

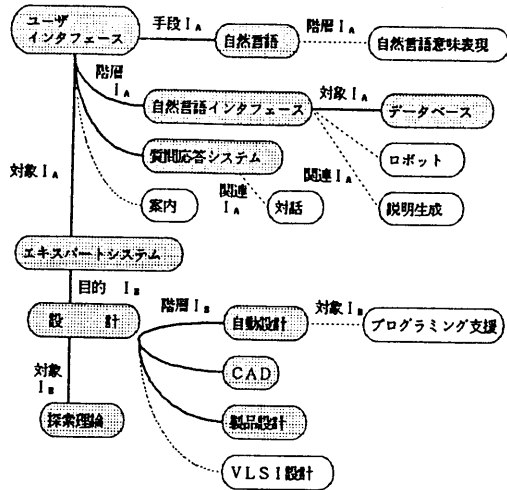
関係の種類Rには、階層関係、関連関係、同義関係、限定関係(対象限定、手段限定、目的限定、因果関係)の4種の関係の種類があり、これらは問い合わせ文中の助詞(表1)や対話の内容から決定される。

意図のレベルLは、2つの概念間での利用者の連想の強さを示し、「初期状態」、「意図する」、「意図しない」の3段階ある。重みのついたリンクは、問い合わせ文で現れなかった検索要求を推論するために用いる。

意味構造識別子 I_i は、同一のユーザモデル内にある複数の検索要求に対応する意味構造を識別するためのものである[13][14]。

複数の意味構造を識別するために、構文解析された意味構造に対応するユーザモデル上の構造(基本意味構造上)のノードも意味構造識別子 I_i をもつ。

ユーザモデルは、初期状態として、キーワードノード間を初期状態のリンクで結んだ典型的な対象世界モデル(シソーラス)として定義されている。ユーザモデルは、利用者の問い合わせ文と利用者との対話から初期シソーラスを修正することによって構築される。



問い合わせ文 Q_A 自然言語によるユーザーインタフェースについて
 問い合わせ文 Q_B 設計のためのエキスパートシステムについて

図 2: ユーザモデルの例 (一部のみ)

現在のシステムでは、人工知能分野を対象分野とし、典型的な対象世界モデルとして、文献 [8] のものを参考に使用した。これは、約 800 のキーワードノードからなる。図 2 に、構築されたユーザーモデルの例を示す。

3.3 隠れた検索要求の推論処理

ユーザーモデルを用いて、問い合わせ文に含まれない検索要求を推論するとともに、これに相当するキーワードを抽出する。問い合わせ文の関係構造に対応するユーザーモデル上の構造 (基本意味構造) から、対応する意味構造識別子を持ち、'意図する' レベルにあるリンクを次の推論規則によってトレースしてキーワードを抽出する。

- (1) 上位キーワード ⇒ 下位キーワード
- (2) 関連関係、同義関係、限定関係リンクは両方向に推論する
- (3) 独自語 ⇒ キーワード

ここでトレースされたリンクとノードの構造が、問い合わせ文に対応する意味構造である。

4 対話によるユーザーモデルの構築

問い合わせ文に含まれる名詞の意味や関連する概念を利用者に尋ねることにより検索要求を明確化し、これを意味構造としてユーザーモデルに反映する。この対話によってユーザーモデルが構築される。

4.1 意味構造の決定・構築

構文解析された問い合わせ文に対応するユーザーモデル上の構造 (基本意味構造) を同定もしくは構築し、対応する意味構造を構築する。

構文解析された問い合わせ文の関係構造と、既にユーザーモデル内に構築された意味構造の関係によって、意味構造の識別および構築は次の 3 つの場合に分けられる [14]。

Case 1: 基本意味構造が、既に意味構造識別子を付与された意味構造の部分集合で、識別子が一意の場合

基本意味構造内の識別子が、入力された問い合わせ文に対応する意味構造識別子と考える。基本意味構造から、この識別子を持ち、'意図する' レベルを持つリンクをトレースしてキーワード集合を抽出する。

Case 2: 問い合わせ文において、利用者が新しい概念や概念間の新しい関係を表明した場合

新しい検索要求として、新しい識別子をもった意味構造を構築する。このような、ユーザーモデル構築のための対話には次のようなものがある。

(1) あいまいな概念の意味の明確化

問い合わせ文中の名詞に対応する概念ノードに対して初期状態レベルのリンクで連結されている下位または関連、限定関係の概念ノードがあれば、これらを利用者に問い合わせ、その回答に応じてリンクの意図のレベルを '意図する' もしくは '意図しない' に修正する。これによって、一つのあいまいな概念に対する意味構造が規定される。

(2) 未知語に対する処理

問い合わせ文中の言葉が未知語の時は、利用者にその言葉の説明を求める。システムは、対話に基づいてこれを新たなリンク (意図するレベル、新しい識別子をもつ) と新たな独自語ノードとしてユーザーモデルに組み入れる。

(3) 関係構造の付加

利用者がユーザーモデルにある概念ノードを用いてそれらの間の新たな関係を表明したときには、その関係に基づいたリンク (意図するレベル、新しい識別子をもつ) を付加する。

Case 3: 基本意味構造内に、既に付与されている意味構造識別子が複数あり一意に決まらない場合

すでに構築された意味構造を利用して、次のように新しい意味構造識別子をもつ意味構造を構築する。

基本意味構造内に既に付与されている意味構造識別子が付与され、'意図する' レベルにあるリンクをトレースして候補となるキーワードを抽出し、これを利用者に表示する。利用者は、検索要求に対応するキーワードを選択する。利用者が選択したキーワードをトレースしたリンクには、新しい意味構造識別子を付加する。

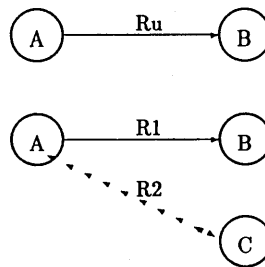
これによって、既に構築された意味構造を利用して、新しい検索要求に対する意味構造が簡単に構築できる。

4.2 利用者への確認と補足する要求の入力

対話の終りに当たって、抽出された全てのキーワードを利用者に表示し確認を求める。

利用者が、新たなキーワードを要求した場合は、そのキーワードと基本意味構造にある概念ノードの間に、‘意図する’レベルで現在の意味構造識別子を持つリンクを付与する。

利用者が、表示されたキーワード集合の中のキーワードを否定した場合には、そのキーワードをトレースしたリンクの意図のレベルを‘意図しない’に修正する。



5 ユーザモデルを用いた論理式の合成方法

対話の終了後、ユーザモデルにおける検索要求の意味構造を利用して文献データベースをアクセスするための論理式を合成する。基本的な考え方は次の2点である。

- (1) 意味構造におけるリンクの構造と、関係の種類および意図のレベルを利用して論理式を合成する。意味構造に含まれる全てのキーワードおよび関係構造を利用して作られた最も制約条件のきつい論理式を最適論理式とする。
- (2) 最適論理式で文献データベースをアクセスしたときに、対応する文献が得られなかった(0件であった)場合には、制約条件を緩和した論理式によって文献データベースを検索する。このようにして順に制約条件を緩和した論理式を合成して文献データベースを検索する。制約条件を緩和したことは論理式に与えられる評点に反映する。

5.1 最適論理式の合成

(1) 有効グラフの作成

問い合わせ文に対応する意味構造を、隠れた検索要求推論時での推論の方向で決定される方向性を持つリンクからなる有効グラフで表現する。

(2) 合成規則

以下の合成規則を用いて、意味構造の有効グラフをトップダウンに(基本意味構造を出発点として)、論理式に置き換えて行く。

規則1: 各リンクの関係の種類に対する基本変換規則は表2に示す。

規則2: 同一種類の複数の出力リンクをもつノードに対する変換規則(表3)

規則1によって得られた論理式のOR結合を与える。但し、意図しない階層関係のときは、意図しないノードのOR結合の否定とのAND結合とする。

規則3: 異なる種類の出力リンクをもつノードに対する

表2: 各リンク関係の種類に対する基本変換規則(規則1)

リンク R_u の種類	合成規則	論理式
意図する・階層関係	下位概念のみを与える	B
意図する・関連関係	ORで結合	$A+B$
意図する・限定関係	ANDで結合	$A \times B$
意図しない・階層関係	上位概念と、下位概念の否定をANDで結合	$A \times \bar{B}$
意図しない 関連、限定関係	意図しないキーワードは無視する	A

表3: 同一種類の複数の出力リンクをもつノードに対する変換規則($R_1=R_2$)(規則2)

関係 R_1 の種類	論理式
意図する・階層関係	$B+C$
意図する・関連関係	$A+B+C$
意図する・限定関係	$A \times (B+C)$
意図しない・階層関係	$A \times (B+C)$

変換規則(表4)

同一種類の出力リンクをグループ化する。

各グループの中で、規則2を適用して論理式を作る。各グループでできた論理式を、意図する階層関係、関連関係、限定関係、意図しない階層関係の順に、論理式に置き換えていく。

規則4: 多段のリンクがあるとき

- (1) 同一レベルにあるリンクについて規則3の処理を行なう。
- (2) さらにリンクがつながっているノードに対しては、その先のリンクに対して規則3の変換処理を行ない、この結果を、(1)の式のノード名と置き換える。

5.2 論理式の制約条件の緩和規則

上記で生成した論理式が、制約条件が最もきつい最適な論理式である。これをもとに条件を緩和した論理式を

表 4: 複数の出力リンクをもつノードに対する変換規則 (R1≠R2 のとき)(規則 3)

		R2			
		意図する・階層	意図する・関連	意図する・限定	意図しない・階層
R1	意図する・階層	$B+C$	$B+C$	$B \times C$	$B \times \bar{C}$
	意図する・関連	$C+B$	$A+B+C$	$(A+B) \times C$	$(A+B) \times \bar{C}$
	意図する・限定	$C \times B$	$(A+C) \times B$	$A \times B + A \times C$	$A \times B \times \bar{C}$
	意図しない・階層	$C \times \bar{B}$	$(A+C) \times \bar{B}$	$A \times C \times \bar{B}$	$A \times \bar{B} \times \bar{C}$

次のように合成する。以下の規則を、意味構造の最も外側から順に適用し、基本意味構造におけるリンクに対応したところで処理を終える。

規則 1: 意図しない階層リンクの変換規則を、上位のキーワードのみを与えるものとする。
即ち、基本変換規則を、 $A \times \bar{B} \rightarrow A$ に置き換える。

規則 2: 意図する階層リンクの変換規則を OR 結合とする。
即ち、基本変換規則を、 $B \rightarrow A+B$ に置き換える。

規則 3: 意図する限定リンクの変換規則を OR 結合とする。
即ち、基本変換規則を、 $A \times B \rightarrow A+B$ に置き換える。

これらの緩和規則が論理式に加わる毎に論理式の評点を減じる。

5.3 検索文献の表示と利用者による適合文献の選択

評点の高い、即ち制約条件の高い最適な論理式から、順に制約条件の弱い論理式を用いて文献データベースをアクセスする。

利用者には、評点の高い論理式で検索された文献から順に、各評点の論理式ごとに検索された文献集合を表示する。最終的に得られた文献数が少なくとも 20 件以上かつ多くとも 100 件になるまで条件の緩和された論理式を用いる。(表示された文献数が 20 件を越えた時点で終了)。

図 3 に、検索された文献が論理式の評点ごとに利用者に表示された例を示す。評点の低い論理式によって検索された文献集合には、評点の高い論理式によって検索された文献集合を含む。そこで、低い評点で示された表示文献では、評点の高い文献は除いてある。利用者はマウスクリックによって適合文献を選択する。ここで選択された文献から、後述するように、各論理式ごとの検索効率を計算するとともに、フィードバック情報が作成される。

The screenshot shows a search results window titled 'MAX-COARIT - 7'. It contains a list of search results with columns for document ID, title, and various attributes. The results are sorted by relevance score. The first few results are highlighted in blue. The attributes listed include '自然言語', 'データベース', '情報検索', '対話理解', 'エキスパートシステム', and '知識ベース'.

図 3: 検索文献表示例

6 検索文献からの意図理解

6.1 フィードバック手法の必要性

利用者が表面上あまり意識しておらず、従って対話により抽出できないキーワードであっても、検索された文献集合に高い確率で存在するものがある。

このようなキーワードは、利用者が陽に意識していないが、文献データベースの検索のために有効なものである。このキーワード集合を抽出し、論理式に含めて再び検索を行なうことによって、文献の検索効率を向上させることが期待できる。

本システムでは、検索文献のうち利用者が選択した文献の中から、対話処理だけでは抽出できなかった検索要求に適合する隠れたキーワードを発見し、これをユーザモデルの意味構造に反映して、再び検索を行なう。これをフィードバック処理と呼ぶ。新たに抽出されたキーワードをユーザモデルに反映しておくことにより、次の問い合わせ時にはこれを自動的に抽出できる。

6.2 検索結果からの検索意図の抽出方法

検索された文献のうち利用者が選択した文献に含まれるキーワードを解析することにより、利用者が意識していないが検索に有効なキーワードを抽出する。

即ち、利用者が選択した文献に付与されているキーワードのうち、

- 全文献中の出現頻度に比べて高い頻度で出現するキーワード
- 高い頻度で共起して出現するキーワード

を、利用者の検索要求を代表するキーワード考える。これらの検索要求を代表するキーワードのうち、対話処理で抽出できなかったものを抽出する。

ところで、対象とする分野において、データベース中50%以上の文献に付加されるようなその対象分野そのものを表したり、広い意味で与えられているようなキーワード(例えば人工知能分野では、“人工知能”、“エキスパートシステム”、“知識表現”といったキーワード)も上記の条件にあてはまる。しかし、このようなキーワードがフィードバックされて意味構造に含まれると、極めて広い範囲の文献が検索され再現率の低下を招くため、条件から省くことが必要である。

また、フィードバックされるキーワードは、対話によって抽出されたキーワードとの関連性を持つものに限ることが望まれる。

このような見地から、検索文献からの意図理解の結果をユーザモデルの意味構造にフィードバックするために、次の基準を満たすキーワードペアを抽出する。

- (1) 次の条件を満たすキーワード集合を抽出する。

$$\frac{S_n}{S} > \frac{N_n}{N} \times 2$$

S_n : 利用者が選択した文献中、そのキーワードを含む文献数

S : 利用者が選択した文献数

N_n : 全文献中、そのキーワードを含む文献数

N : 全文献数

- (2) (1)で得られたキーワード集合において、利用者が対話によって選んだキーワードと共起関係が最大のキーワードのペア。(但し、既に対話によって抽出されて意味構造に含まれていたキーワードとのペアは除く)

6.3 検索結果からのユーザモデルの構築方法

フィードバック処理により抽出したキーワードを検索要求を表すものとしてユーザモデルに取り込んでおけば、以降の問い合わせ文の処理において自動的に隠れた検索要求として推論することができる。

これは次のように、抽出されたキーワードペアをユーザモデルのノードとその間のリンクとして付加することによって行なう。

- ユーザモデル上の対応するキーワードノード間で、既に現在の意味構造識別子を持ったリンクが存在すれば、その意図のレベルを‘意図する’に修正する。
- ユーザモデル上の対応するキーワードノード間で、現在の意味構造識別子を持ったリンクが存在しないならば、それらのノード間に現在の意味構造識別子を持った関連関係の‘意図する’レベルのリンクを追加する。

7 検索システムの評価

7.1 評価指標について

文献検索システムにおける問い合わせ理解の性能を、キーワードの検索効率(適合率・再現率)および文献の検索効率指標で評価した。

- キーワード検索効率
キーワード検索効率(キーワード適合率 Kp_q およびキーワード再現率 Kr_q) は次のように定義される。

$$Kp_q = \frac{|Eq'|}{|Eq|}, \quad Kr_q = \frac{|Eq'|}{|Pq'|}$$

$|Eq|$: システムが抽出したキーワードの数

$|Eq'|$: システムが抽出したキーワードのうち利用者の検索意図に適合したキーワードの数

$|Pq'|$: 全キーワードのうち利用者の検索意図に適合したキーワードの数

- 文献検索効率
文献検索効率(文献適合率 Rp_q および文献再現率 Rr_q) は、上記キーワード検索効率の定義の“キーワード”を“文献”と読み換えたものとなる。

文献検索効率は順位付け出力を採用するシステムにおいては一般に再現率-適合率曲線で表現される[9][10]。

7.2 評価実験について

評価実験では各被験者に3つの問い合わせ文 Q_1 、 Q_2 、 Q_3 を作成してもらった。3つの問い合わせ文を順にシステムに与え、それぞれ対話処理とフィードバック処理を行なった。このセッションを3回繰り返す、このとき同一の検索要求であれば2、3回目の問い合わせ文での表現は変えてもよいとした。即ち、問い合わせ文は(1回目) $Q_1 \Rightarrow Q_2 \Rightarrow Q_3 \Rightarrow$ (2回目) $Q'_1 \Rightarrow Q'_2 \Rightarrow Q'_3 \Rightarrow$ (3回目) $Q''_1 \Rightarrow Q''_2 \Rightarrow Q''_3$ のように与えた。

また、被験者の検索意図にあったキーワードおよび文献は、被験者に全キーワード、全文献の中から選択してもらった。

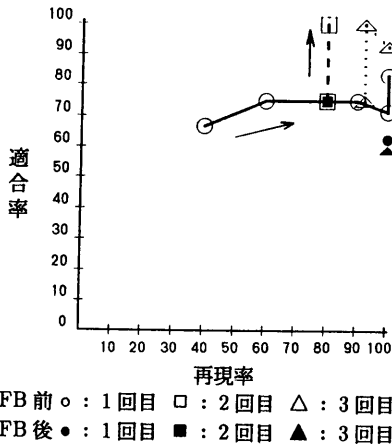


図 4: キーワード検索効率の推移 (被験者 A、Q₁)

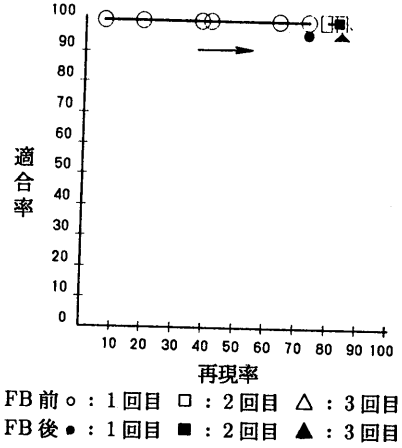


図 7: キーワード検索効率の推移 (被験者 B、Q₂)

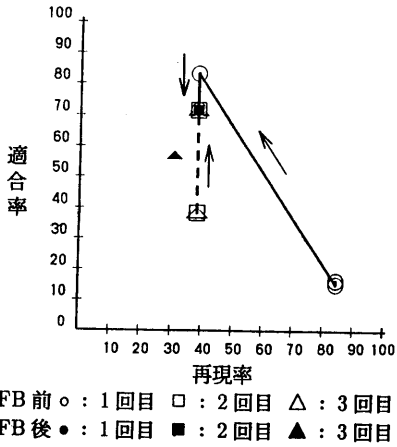


図 5: 文献検索効率の推移 (被験者 A、Q₁)

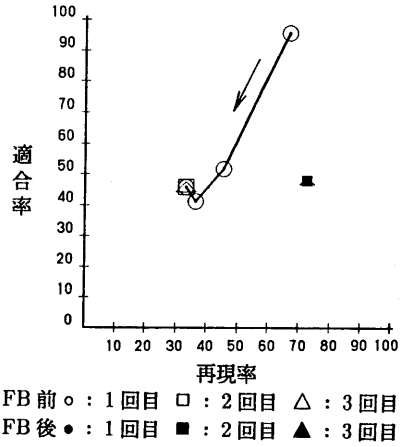


図 8: 文献検索効率の推移 (被験者 B、Q₂)

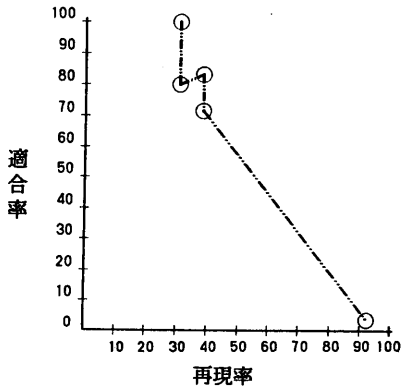


図 6: 1 回目の対話の最終の時点での各評点ごとの論理式で検索された文献集合の検索効率 (被験者 A、Q₁)

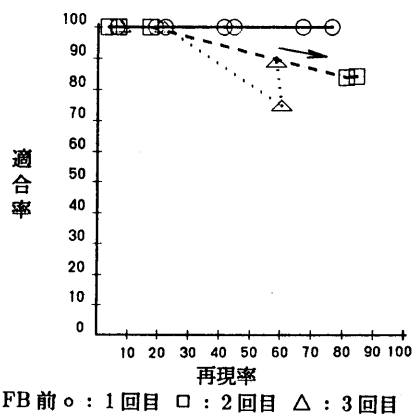


図 9: キーワード検索効率の推移 (被験者 B、Q₂、意味構造識別子を用いない場合)

7.3 評価結果

図4および図5は、被験者A、Q₁の対話の各時点のキーワード検索効率および文献検索効率の推移を、図6は、1回目の対話の最終の時点での各評点ごとの論理式で検索された文献集合の検索効率を示す。図7および図8は、被験者B、Q₂の対話の各時点のキーワード検索効率および文献検索効率の推移を示す。図中の矢印は対話が進んだ方向を示す。ここで文献検索効率は、5.3節で示した条件で利用者に表示された文献集合に対する検索効率を表す。

表5は、各対話の最終点において、フィードバック前に得られたキーワード集合およびフィードバック後に得られたキーワードを付加したキーワード集合に対するキーワード検索効率と、フィードバック前に被験者に表示された文献、フィードバック後に検索できた文献のキーワード検索効率を示す。フィードバック後に検索できた文献の検索効率は、フィードバック前と後に得られた文献の和に対する検索効率を示している。

図4、図7のキーワード検索効率の推移より、1回目の対話でユーザーモデルが構築され、2回目以降の対話ではこれを使った効率のよい対話がなされ、少ない対話数で高いキーワード検索効率を得られることがわかる。

図6の各評点ごとの論理式で検索された文献については、評点が高い論理式によるものほど再現率が低く、適合率が高い様子が示されている。この例では、図3に示した文献リストが利用者に表示された。即ち評点が高いものから順に上位4番目までの論理式が用いられたことを示している。この時点での文献適合率は71%、再現率は38%である。

文献検索効率の推移については、図5、図8のように対話の途中においてキーワードの再現率が小さい時期のほうが、文献再現率が高い場合がある。これは、対話が進んで、用いるキーワードが多くなった分、制約条件が強くなり、文献の再現率が下がったものと考えられる。

これは、本システムでは抽出されたキーワードがすべて同一の重要度であるとして検索を行なっていることが一因である。従って、このような問題点を解決し、対話終了時の検索効率を向上させるためには、利用者が直接言明した言葉や、意味構造において幹の部分にあるキーワードを重要キーワードとし、最初はこのような重要なキーワードだけで論理式を合成し、その後文献再現率を向上させるためにその他のキーワードを用いた論理式を用いるようにすればよい。

また、利用者に表示する文献数条件の最大値を多くすれば再現率の高い文献集合が得られる。

図9は、被験者B、Q₂の意味構造識別子を用いない場合の対話の各時点のキーワード検索効率の推移を示す。

図9において、意味構造識別子を用いない場合には、他の検索要求の意味構造との干渉によって、与えられた検索要求に対する意味構造が破壊され、2回目以降の対話で意味構造の再構築が行なわれていることを示してい

表5: 文献検索効率・キーワード検索効率

被験者	問い合わせ		文献検索効率 (%)		キーワード検索効率 (%)		
			再現率	適合率	再現率	適合率	
							再現率
A	Q1	FB前	44	65	93	92	
		FB後(*)	44	58	93	66	
	Q2	FB前	81	30	79	100	
		FB後(*)	88	27	79	83	
	Q3	FB前	25	60	86	45	
		FB後(*)	25	60	86	45	
	平均	FB前	50	52	86	79	
		FB後(*)	52	48	86	66	
	B	Q1	FB前	92	14	74	100
			FB後(*)	92	14	74	91
Q2		FB前	33	46	84	99	
		FB後(*)	72	48	84	96	
Q3		FB前	40	30	89	90	
		FB後(*)	45	23	89	85	
平均		FB前	55	30	82	96	
		FB後(*)	70	28	82	91	
C		Q1	FB前	85	46	93	100
			FB後(*)	93	43	93	90
	Q2	FB前	97	41	61	67	
		FB後(*)	100	34	61	38	
	Q3	FB前	91	67	71	100	
		FB後(*)	91	67	71	100	
	平均	FB前	91	51	75	89	
		FB後(*)	95	48	75	76	
	D	Q1	FB前	38	59	71	100
			FB後(*)	67	36	74	87
Q2		FB前	60	8	67	93	
		FB後(*)	80	4	67	76	
Q3		FB前	33	38	78	93	
		FB後(*)	33	36	78	83	
平均		FB前	44	35	72	95	
		FB後(*)	60	26	73	82	
平均		FB前	60	42	79	90	
		FB後(*)	69	38	79	78	
制約条件最小論理式			95	18	-	-	

(*) フィードバック前に検索された文献(キーワード)を含んだ値

る。意味構造識別子を用いることによって、複数の検索要求が同時に入力されても各問い合わせに対する意味構造が保持され効率のよい対話が可能になった。

フィードバック処理の結果、フィードバック前に検索できなかった文献が1件以上検索できた割合は64%であった。またフィードバック処理の結果、適合する文献が1件以上検索できた割合は42%であった。

文献検索効率は、制約条件が高い論理式によって検索すると適合率は高くなるが再現率が低くなる。文献の再現率を高めるためには、制約条件の低い論理式を用いれば良い。しかし、この場合には大量の文献が検索されてしまい、再現率は向上するが適合率がかなり小さくなるという問題点があった。制約条件が小さい論理式による文献検索効率の平均を表5に示す。フィードバック処理によって文献適合率はほぼそのまま文献再現率は、平

均15%の改善が見られた。フィードバック処理によれば、制約条件の低い論理式を用いることに比べ文献の適合率を大きく落さずに高い再現率を得ることができる。

このようにフィードバック処理によって、フィードバック前に検索できなかった文献が効率的に検索された。

キーワード検索効率では、フィードバック後の方が適合率が低下する傾向が見られる。しかしながらフィードバック処理によってフィードバック前に得られなかった適合文献が得られていることを考え合わせると、利用者が検索要求に直接関係がないと考えているキーワードであっても文献検索にとって有効なキーワードが存在することが示されたといえる。即ち、適合文献中によく現れやすいが、対話で意識しないキーワード情報が文献検索に有効であることがわかる。

キーワードの適合率、再現率が高くて必ずしも文献の適合率、再現率が高いとも限らない。これは、前述したようにキーワードの重要度の考慮が必要なことと、文献の検索効率が、対象とする文献データベースに依存していることによる。

8 むすび

あいまいな問い合わせ文を理解する文献データベースの知的インタフェースの構成法について述べた。

本システムでは、対話によって、問い合わせ文の意味構造を明確化するとともにこれをユーザモデル内に構築する。そしてユーザモデルにおける意味構造を基に論理式を合成する方法を示した。論理式の制約条件に基づく評点毎に文献データベースをアクセスすることによって、再現率は低い適合率の高い文献集合や、適合率は低い再現率の高い文献集合を生成することができる。

対話において利用者は意識していないが、文献検索にとって必要なキーワードが存在することがある。このため、検索文献の中から利用者が選択した文献に付与されていたキーワード情報を解析してこのようなキーワードを抽出し、これをユーザモデルにフィードバックする手法を提案した。このようなフィードバック処理によって文献適合率はほぼそのまま文献再現率は、平均15%の改善が見られた。これによって、論理式の制約条件を緩和する場合より効率的な検索が可能となる。

人間の使う言葉は、文献データベースの問い合わせに限りず多くの冗長性やあいまい性を含んでいる。自然言語インタフェースの使い易さを向上するためには、このような人間の使う言葉の特性を許すものが望ましい。本手法は、文献データベースのインタフェースの構成法として示したが、自然言語インタフェースの実現において本稿で示したような構成法は広く適用が可能である。

謝辞

本研究の機会を与えて下さったATR通信システム研究所葉原耕平会長、山下紘一社長、日頃より議論いただいた木下茂行主任研究員(現川崎製鉄(株))、および評価

実験に協力して下さった知能処理研究室の皆様にも感謝します。

参考文献

- [1] G. Jakobson, et al. "An Intelligent Database Assistant", IEEE EXPERT, SUMMER, 1986.
- [2] F. N. Tou, et al. "RABBIT: An Intelligent Database Assistant", in Proceedings of AAAI-82, 1982.
- [3] I. Monarch and J. Carbonell, "CoalSORT: A Knowledge Based Interface", IEEE EXPERT, SPRING, 1987.
- [4] R. S. Taylor, "The Process of Asking Questions", American Documentation, Oct., 1962.
- [5] N. J. Belkin, et al. "ASK for Information Retrieval: Part 1. Background and Theory", The Journal of Documentation, Vol. 38, No. 2, June, 1982.
- [6] N. J. Belkin, et al. "ASK for Information Retrieval: Part 2. Result of a Design Study", The Journal of Documentation, Vol. 38, No. 3, Sept., 1982.
- [7] J. Yen, et al. "Specification By Reformulation: A Paradigm for Building Integrated User Support Environment", in Proceedings of AAAI-88, 1988.
- [8] David L. Waltz, "Scientific Datalink's Artificial Intelligence Classification Scheme", THE AI MAGAZINE, Spring, 1985.
- [9] サルトン編著, 神保監訳, "SMART 情報検索システム", (株) 企画センター, 1974.
- [10] 伊藤哲郎著, "情報検索", (株) 昭晃堂, 1986.
- [11] 加納, 木下他, "文献検索システムにおける検索要求意図推論方式の一検討", 情報処理学会第37回全国大会, 1988.
- [12] Y. Kanou, S. Kinoshita, et al., "Query Understanding Using User Modeling in a Document Retrieval System", Proceedings of ICCS Symposium '89, Beijing, pp.105-108, 1989.
- [13] 加納, 木下他, "知的検索のための複数の検索意図を表現するユーザモデル", 人工知能学会研究会資料, SIG-HICG-8902-1, 1989.
- [14] Y. Kanou, S. Kinoshita, et al., "Document Retrieval with a User-Model to Represent Multiple Requests", Proceedings of ICCI '90, Canadian Scholar's Press, pp.269-274, 1990.
- [15] S. Kinoshita, Y. Kanou, et al., "Deep Understanding of Japanese Queries in an Information Retrieval System", Proceedings of the 1988 International Conference on Computer Processing of Chinese and Oriental Languages, pp.358-362, 1988.
- [16] 木下, 加納他, "知的文献検索システムにおける問合せ理解とその評価", 情報処理学会知識工学と人工知能研究会報告, 62-2, 1989.
- [17] 木下, 加納他, "文献データベースの知的インタフェースとその性能評価", 情報処理学会データベース・システム研究会報告, 72-20, 1989.
- [18] S. Kinoshita, Y. Kanou, et al., "Quantitative Evaluation Method for Intelligent Interface to a Document Database, Proceedings of the Working Conference on Data and Knowledge Base Integration, England, 1990.
- [19] 木下, 加納他, "文献データベースの知的インタフェースにおける検索結果からの問合せ理解", 情報処理学会データベース・システム研究会報告, 75-3, 1990.
- [20] 拝原正人, "日本語文献データベースへの知的アクセス", 電子情報通信学会誌, Vol. 72, No. 7, 1989.