ICOT Technical Report: TR-210

TR-210

自然言語処理に基づく
情報検索システムIRIS

杉山健司、秋山幸司、伊吹　潤
川崎正博、内田裕士
（富士通）

October, 1986

ICOT
Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan
(03) 456-3191～5
Telex ICOT J32964

Institute for New Generation Computer Technology

自然言語理解に基づく情報検索システム IRIS

杉山健司，秋山幸司，伊吹　潤，川崎正博，内田裕士
（富士通）

本論文では，自然言語質問文を理解し，その質問文に内容的に答えるようなテキスト群をテキストベースから検索する知的情報検索システム（IRIS）について述べる。本システムの目標は，(1)用語選択労力からのユーザの開放，(2)自然言語質問文に含まれるキーワードとその論理結合以外の情報のテキスト選択への利用，及び(3)テキストのトピック理解に基づくテキストの自動選別である。これらの目標のうち，主に，(1)，(2)を目指したシステムについて議論する。本システムは，対象世界のモデル，質問文／テキスト文解析，キーワード検索式自動生成，内容マッチングから構成される。

IRIS: An Intelligent Information Retrieval System based on Natural Language Understanding

Kenji SUGIYAMA, Kouji AKIYAMA, Jun IBUKI, Masahiro KAWASAKI and Hiroshi UCHIDA
Fujitsu Ltd. and Fujitsu Laboratories Ltd.
（英称注所）　1015, Kamikodanaka, Nakahara-ku, Kawasaki 211, JAPAN

This paper describes the design of an intelligent information retrieval system: IRIS, which retrieves texts through user's natural language query. Three targets of IRIS are dicussed, including (1) to make a user free from the burden of finding suitable keywords and their logical combination, (2) to select the texts by making use of the information other than keywords, naturally contained in a natural language query, and (3) to automatically select specific topic texts. The system solving two former targets are described in detail, which consists of the models of a task domain, the parsers for query and text, the keyword generation expert and the content matcher.

1．はじめに

従来からの情報検索システムは，検索キーをどう統制するかいう観点から，完全統制するものと，完全に自由にするものと，その両者の中間に位置するものに分けることができる。完全統制の場合，何が統制語であるかを知ることがユーザの負担になり，逆に，完全自由の場合，関連する用語を思い付けるかといったユーザの資質によって適合率・再現率が変ってしまうという欠点がある。両者の長所欠点を補うものとして両者の中間のものがあるが，いづれにしろ，ユーザは用語という壁につきあたる。それを緩和する方法として，オンラインシソーラスが開発されてきたが，シソーラスを引く手間と，どう用語を見付けるかはユーザに任されたままである。

一方，人工知能，特に，自然言語理解の分野に目を向けてみると，近年，自然言語インタフェースの研究が盛んに行なわれ，リレーショナルデータベース等の形式化されたデータに対する自然言語アクセスは，可能になりつつある［Ishikawa et al. 86］。我々は，このような背景から，自然言語理解に基づくテキスト情報への知的アクセスによる情報検索システムの高度化を目指して，知的情報検索システム（IRIS: Intelligent Information Retrieval and Information Selection System）を研究開発中である。

IRISでは，従来の情報検索システムの高度化の方向として，まず，次の２つを目指している。

(1) 上に述べたような用語選択の労力からユーザを解放する。

(2) 自然言語質問文によって，従来のキーワードの論理式では表せないような検索条件を自然に表せるので，このような検索条件を有効に使う。例えば，「日立がある会社に技術供与したという内容のテキストは？」という問合せと，「ある会社が日立に技術供与したという内容のテキストは？」という問合せでは，検索条件が異なる。

(1)を実現するには，質問文を解析し，その中からキーワード検索式を自動抽出し，さらにシソーラス情報を使って適切な関連キーワードに展開する必要がある。(2)を実現するためには，質問文だけではなく，テキスト文の方も意味解析し，意味構造同志のマッチングを行う必要がある。

このような自然言語理解に基づく知的情報検索システムの利点は，上記２つの目標達成による直接的利点に加え，間接的にもう１つ，ユーザがそれぞれの情報検索システムに固有な問合せ言語を覚える必要がなくなることである。

本論では，以下，IRISのシステム概要を述べ，上記目標達成に必要となる対象世界のモデル，及び，質問文／テキスト文解析，キーワード検索式自動生成，内容マッチングの方式について議論する。各機能を実現するモジュールは，新世代コンピュータ技術開発機構が開発した逐次型推論マシンPSI のプログラミング言語であるESP ［Chikayama 84］で記述されている。また，IRISの３つ目の目標である自然言語理解によるテキスト情報の選別という話題についても議論する。最後に，情報検索システムの高度化を狙った他の研究との比較も行なう。

2．システム概要

2．1 自然言語理解と対象世界

一般に，自然言語を理解するには，言葉で表現される対象世界をモデル化し，計算機内に表現しておく必要がある。また，現在の技術レベルでは，すべての自然言語文を自由に理解するシステムを作ることは不可能であるので，分野を限定して実用に近いシステム作りを目指す必要がある。勿論，システムのフレームワークそのものまで，この限定された分野のみ効果があるものでは，各分野ごとにシステムを作り直さなければならなくなってしまうので，システムのフレームワークは分野独立に作り，分野依存知識を変更することによって様々な分野に適応できるシステムを目指す。このような意味から，IRISでは，対象世界のモデルを導入し，そこに分野依存知識を格納できるようにしている。

対象世界のモデル化を行うためには，ユーザが持っている対象世界のイメージを知る必要がある。そこで，対象分野として，まず，情報産業界の新聞記事の見出しを取りよせ，アンケート調査により質問文例を約300 文収集した。さらに，日本情報産業新聞，日経産業新聞，通信電点コンピュータから情報産業界に関する記事を約800 件程抽出した。これらの文例を図１に示す。これらの文例を分析することによって後述する対象世界のモデルが得られる。

「韓国ソウルエレクトロン　ＣＡＤ／ＣＡＭで米ＣＶ社と合弁会社」
日本情報産業新聞　1984年11月 5日　2面

「エンジニアリング・パッケージソフト提携が発表　二次元／三次元の解析用　米ＭＤＩ社の製品」
週刊電波コンピュータ　1984年12月17日　8面

(a) 見出し文の例

Q1　日経のトップ記事が見たい。
Q2　最近，ＣＡＤ／ＣＡＭ分野に進出した企業は？
Q3　３日前の記事でパソコンについてのものはあるか？

(b) 質問文の例

図１　収集文の例

2．2 テキストベース

上で述べた対象分野のテキストは，図２のようなマスタファイルとインバーテッドファイルから構成される。マスタレコードの情報は，何時のどの新聞の何面に載ったかを

表す書誌的情報と自然言語テキストである記事見出しから構成される。この記事見出しには，インバーテッドファイルが対応し，あるキーワードを含む記事が即座にわかるようになっている。

現在，IRISでは，インバーテッドファイル中のキーワードは，従来型の情報検索システムが作り出すものを仮定している。すなわち，記事見出しを単語辞書を用いて単語分割し，得られた単語のうち助詞や助動詞を除いたものを全てキーワードとしている。勿論，従来型の情報検索システムで，これとは違った方式でキーワードを決めているものもあるが，このような違いは，先に述べた対象世界のモデル内に各方式に対応した知識を格納することにより，対処できる。

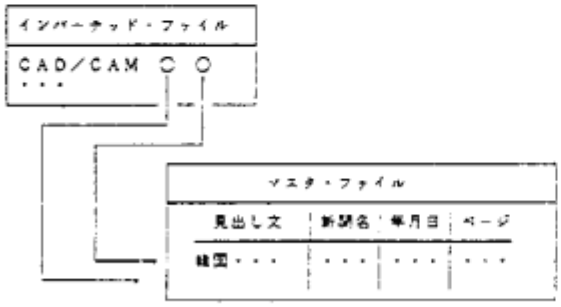

図２　テキストベースの構成

２．３　システム構成

IRISのシステム構成は，図３のようになっている。ユーザの質問文は，まず，質問文解析モジュールによって構文意味解析され，質問文の意味構造が作られる。この意味構造は，図にあるように，キーワード検索式自動生成モジュールと内容マッチャーの両方に送られる。

次に，この２つのモジュールのうち，まず，キーワード検索式生成モジュールが動き出す。このモジュールは，質問文の意味構造から概念的キーワードとその論理結合である検索式を生成する。ここで，概念的キーワードといったのは，必ずしも質問文に陽に現れていない用語でも，人間の検索専門家なら含めるべきだと判断するような用語を含んだものを意味している。例えば，「富士通に関する記事は？」といった質問文の場合，「計算機メーカ」とか「半導体メーカ」といったキーワードをも含むことを意味する。

このようにして得られた概念的キーワードの検索式は，従来から実用化されているキーワードによる情報検索システム（ここでは，それを機械的情報検索サブシステムと呼んで知的情報検索システムと区別している）によって実行され，その結果，テキストベースから対応するテキスト群が得られることになる。

ここで得られたテキスト群は，あくまでキーワードのレベルで得られた結果なので，テキスト文の意味内容にそれ程深くは立ち入っていない。そこで，より深くテキスト内容に立ち入った検索を実現するため，まず，このテキスト群を，さらに構文意味解析（テキスト文解析）し，テキスト文の意味構造を作り出す。さらに，このテキスト文意味構造と先に得られていた質問文意味構造との内容マッチングを行い，最終的に意味内容が一致するテキスト群を検索する。

以上のIRISのコンポーネントのうち，機械的情報検索サブシステムを除いた各モジュールは，対象世界のモデルを使って各々の処理を行う。

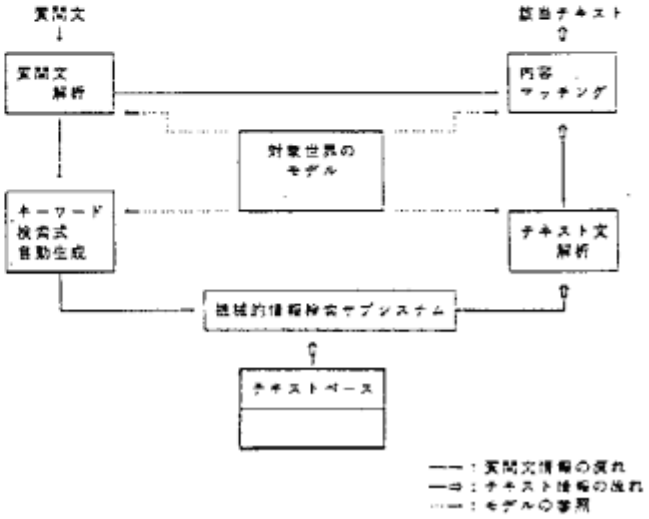

図３　システム構成

３．対象世界のモデル

一般に，多くの情報検索システムの対象であるテキストベースを見てみると，そこに格納されている情報は，大きく，書誌情報と，テキスト情報そのものに分けることができる。そこで，IRISでは，この２つを区別し，それぞれに対応したモデルを導入している。前者を背景モデルと呼び，後者を内容モデルと呼ぶ。背景モデルは，テキストの書誌的情報，即ち，そのテキストがどういう背景で書かれたり，発行されたかを示す。内容モデルは，テキストの意味内容を表す。

２つのモデルのうち，背景モデルは，ちょうど，リレーショナルデータベース内に格納されているレコードの各フィールドの情報をモデル化したものに対応し，リレーショナルデータベースへの自然言語インターフェースであるKID［Ishikawa et al. 86］の分野モデルと同等のものである。内容モデルの方は，KIDの方には対応するものがなく，IRIS特有のものである。

背景・内容モデルとも後述するようにオブジェクト指向パラダイム［Chikayama 84］でいうオブジェクト中心の表現になっている。ある具体的な質問文の意味は，これら２つのモデル中のオブジェクトのインスタンス・ネットワー

クで表現される。テキスト文の方は，２つのモデルの内，内容モデルの方のインスタンス・ネットワークで表現される。具体的な表層文からこれらの意味構造を作り出すため，IRISで取り扱う単語は，すべて，一般的な機能語（助詞，助動詞等）と分野固有の内容語（主に，名詞や動詞）に分類され，内容語は，この２つのモデル中のオブジェクトのどれかに対応付けられている。

内容モデルは，後述する世界知識によって補完される。即ち，内容オブジェクトの意味付けを世界知識で与える。この世界知識は，キーワード検索式自動生成モジュールや内容マッチャーによって使われる。

３．１　内容モデル

対象分野に関する検討から，テキストの意味内容を表現するために必要となるオブジェクトは，述語的なものと名詞的なものに大別される（図４）。述語的オブジェクトは，さらに，意志を持つ実体の物理的・精神的動きをしめす行動オブジェクト，意志を持たない実体の物理的動きを示す動作オブジェクト，実体の意図的・非意図的な変化を表す変化オブジェクト，そして，実体の状態を表す状態オブジェクトの４つにクラス分けされる。名詞的オブジェクトは，組織体や，人，業界を含むある行動を起すような主体や，行動の受け手となるような製品や商品などを含む客体や，行動や変化が起る場所としての舞台などにクラス分けされる。

述語的オブジェクトの４つのクラスの中には，ほぼ同じ概念を表すと考えられる動詞や形容動詞を１つにまとめたオブジェクト（元素オブジェクトと呼ぶ）がある。たとえば，「開発する」，「確立する」，「完成させる」，「実現する」，「実用化する」などの単語は，いづれも，その中核的意味要素として，"作る"という概念を含んでいるので，"作る"という元素オブジェクトを考える。これは，これらの動詞を含む記事，「Ａ社がＢを開発する」，「Ａ社がＢを完成させる」などは，ほぼ同一内容であると判断されるためである。

行動，動作，変化の各述語クラスに属する元素オブジェクトは，各元素に固有の引数パターンを持ち，状態に属する元素オブジェクトは，その元素固有ではなく，その元素に対応する単語の品詞に依存する引数パターンを持つ。引数パターンは，図４に示されているように，述語的オブジェクトと名詞的オブジェクトの意味関係を規定したものであり，格フレームに準じたものである。この意味関係は，先に収集した例文中にある語句の修飾関係をカバーするように決められている。

名詞的オブジェクトの各オブジェクトの間には，図４に例示されているように，所在地や供給者といった依存関係が存在する。上位オブジェクトが持つこれらの関係は，下位オブジェクトにも遺伝される。即ち，図の例では，客体が組織体と供給者という依存関係を持つので，客体の下位オブジェクトである製品や商品等も組織体と同様の依存関係を持つ。これらの依存関係は，名詞的オブジェクト間の２項関係となっているが，前述した述語の引数パターンは，名詞的オブジェクト間の関係という観点からは，多項関係を表現している。

本モデルと非常によく似た文の意味モデルとしては，概念関係モデル［藤澤他 86］があるが，本モデルで表現されているような多項関係は表現されていない。これは，かれらのシステムでは，読者あるいは情報検索者が注目するのは名詞的実体とそれらの間の２項関係であるとして，テキスト文の主題をモデル化しているためである，と考えられる。しかし，より一般的には，ここで扱うような多項関係も主題と考えるべきであると我々は考えている。

３．２　背景モデル

現在対象にしているテキストベースの情報を反映して，背景モデルは図５のような構成になっている。背景モデルのオブジェクトも基本的に内容モデルと同じ枠組みで表現されている。

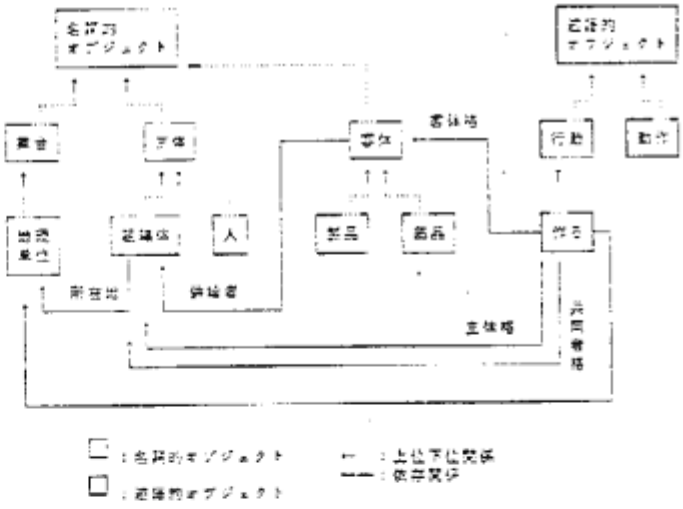

図４　内容モデルの一部

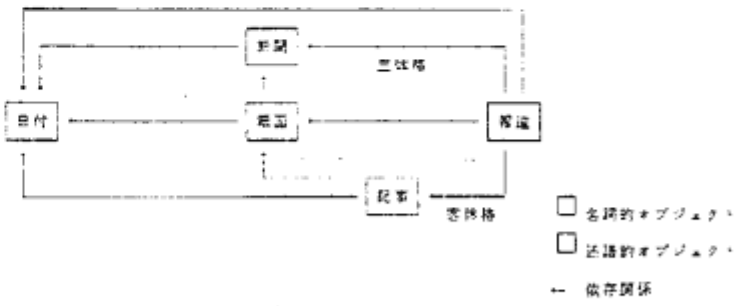

図５　背景モデルの一部

3．3　世界知識

よく議論されるように［Hobbs 84］，総ての知識をシステムに組み込むのは，不可能であるので，組み込むべき知識を選別する必要がある。IRISの現在の応用分野では，例文等の分析から図6のような種類の世界知識があれば良いと考えている。この中には，シソーラス的な情報が含まれている。図に示したもの以外にも必要と考えられる知識，例えば，大企業の子会社に関する知識等もある。しかし，知識を入れることは，非常に労力がいる仕事であるので，その知識の有益度，即ち，検索システム全体の再現率・適合率への寄与という観点から検討して導入することにしている。

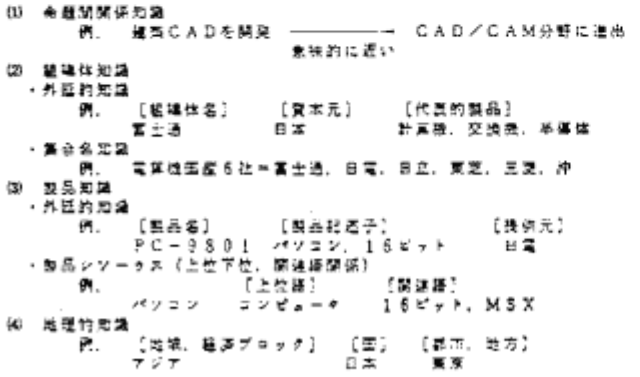

図6　世界知識の種類

4．自然言語文解析

IRISにおける構文意味解析部は，ユーザの質問文と新聞記事見出しという2つの異なる自然言語文を解析し，対象世界のモデルで規定されている意味関係を抽出する役目を担っている。2つの異なった種類の文を扱うため，質問文解析部（Qパーサ）とテキスト文解析部（Tパーサ）がある。

4．1　質問文解析

質問文の中には，書誌情報のみに対する問い合せ（図1(b)のQ1），テキスト内容のみの問い合せ（図1(b)のQ2），書誌情報・テキスト内容両方に対する問い合せ（図1(b)のQ3）の3種類が考えられる。Qパーサでは，これら3種類の質問文から，書誌情報に関する部分と，内容に関する部分を分離抽出する。

Qパーサは，図7に示すように，大きくは，単文統語処理部，背景／内容モデル部，暗黙情報解消部から構成されている。単文の統語処理をする部分は，名詞連続の解析，連体助詞を含む名詞句の解析，述語を中心としたパターンの解析というフェーズ構成を採り，基本的にボトムアップに解析を進める。名詞句，述語解析フェーズでは，背景／内容モデルに対して意味関係に関する問い合せをメッセージの形で行い，文の意味表現である背景／内容オブジェクトのインスタンスのネットワークを作り出していく。この統語処理とモデルとのメッセージ転送に基づく意味表現生成の方式は，KIPS［Sugiyama et al. 84，杉山他 84］のそれにほぼ準じている。

暗黙情報解消部では，述語の暗黙引数の解決を行っている。例えば，「A社が光ディスクの開発を始めたという記事は？」という質問文では，「開発した」のも「始めた」のも「A社」であるが，単文統語処理部・背景／内容モデル部までの処理では，「A社」は，統語上，「始めた」にしか係らず，意味表現上も「始めた」としか意味関係を持たない。そこで，「開発した」の主格が何かといった述語の引数関係を解決するため，暗黙情報解消部がオブジェクトのネットワーク構成を手掛りにして，それが「A社」であることを見つけ出す。

4．2　テキスト文解析

テキスト文の大きな特徴として，助詞が省略され，また，文脈から判る時は，動詞も省略されたりすることである。Tパーサは，このような特徴を持つテキスト文から，テキスト内容を抽出する。

Tパーサは，Qパーサとほぼ同じような構成になっており（図7参照），統語処理とモデルとのメッセージ交換による意味表現の生成方式は同じであるが，対象世界のモデルとしては，内容モデルしか使わない。単文統語処理部の中も，Qパーサの場合と同じフェーズ構成になっているが，助詞の省略等に対処するため，各フェーズの処理内容は，少し違ったものになっている。即ち，名詞連続の解析フェーズでは，内容オブジェクトのクラス分けに基づいた連語範囲の定義情報を，Qパーサの場合より，厳密に適応して，連語の認定を誤らないようにしている。また，連語認定を厳密に行った結果，助詞を伴わない名詞句が多く出来るので，述語の解析フェーズでは，助詞を伴わない名詞句の係り受け関係の解析機能を強化している。

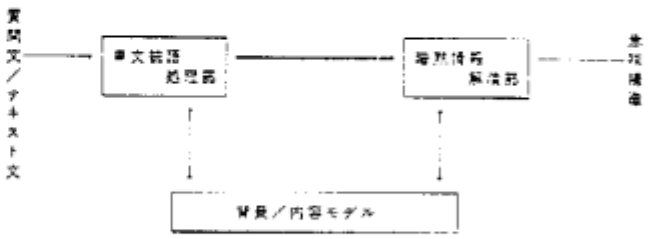

図7　質問文／テキスト文解析モジュール

暗黙情報解消部では，暗黙引数の解決の他に，省略されている動詞の推定処理も行う。暗黙引数の解決は，テキスト文が主見出し文，副見出し文，副々見出し文等，複数の文から構成され，述語とその格要素が離れた文中に存在する可能性が高いため，Qパーサの時と異なるヒューリスティックを使い，1文の範囲を超えて，引数解決を行う。省略動詞の推定は，たとえば，「A氏がB社の社長に」といった見出し文の場合，「就任する」という動詞を補う。この処理のため，内容モデル中の引数パターンを参照する。

5．キーワード検索式自動生成

キーワード検索式自動生成は，Qパーサによって作られた背景及び内容オブジェクトのインスタンス・ネットワークを入力として，機械的情報検索サブシステムで実行可能な検索式を生成する。この検索式は，大きく，背景ネットワークから生成される書誌情報に関する検索条件部と，内容ネットワークから生成されるテキスト内容に関する検索条件部にに分けられる。テキスト内容に関する検索条件は，キーワード論理式で表される。

キーワード検索式自動生成モジュールは，上記2つの検索条件部に対応して，書誌情報条件生成サブモジュールとテキスト内容条件生成サブモジュールの2つに分れている。

5．1　書誌情報検索条件生成

書誌情報条件生成サブモジュールは，背景モデルで表現された書誌情報に関する意味表現からテキストベースに格納されたデータ形式に応じた，即ち，各フィールドに対応した条件式を生成する。この処理は，既存の技術，例えば，KID のようなデータベース・マッピングによる生成方式［Ishikawa et al, 86］や，KIPSのプログラムコード自動生成モジュールのような生成目標のモデルを介する方式［杉山他 84］によって実現できる。RISでは，生成目標のモデルを介する方式，即ち，テキストベースのフィールドをモデル化して，書誌情報に関する条件を生成している。

検索条件を生成する際，暗黙の条件指定に関する情報の補いを行う。例えば，「日経の 6月10日から12月10日までの記事は？」という問い合せの場合，日付に関する知識を使って，この「6月10から12月10日まで」を現在（'86年10月 1日と仮定）に一番近い過去の期間である「1985年 6月10日から1985年12月10日まで」という条件に変換する。

5．2　テキスト内容検索条件生成

テキスト内容条件生成サブモジュールは，内容モデルで表現されている意味情報から，概念的キーワードの論理式を生成する。この処理のためには，シソーラスを引き適切な関連語を探したり，テキストベースに入っているテキスト文の特徴を考慮してキーワードの論理結合を考え出す必要があるが，このような操作は，それ自体が専門技能になるほどに知的な行為であるため，本モジュールは，検索専門家が持っている検索知識［Bates 79］によって駆動される，いわゆる，エキスパートシステムとして実現される。

図8に示すように，大域的作業領域としての黒板は，内容オブジェクトのインスタンスのネットワークである。その初期値は，Qパーサが作り出した内容オブジェクトのインスタンス・ネットワークであり，検索専門家の知識である検索戦術／戦略をモデル化したルールを適用し，このネットワークを変形していくことによって概念的キーワード論理式を生成する。これらのルールの中には，"組織体を示す内容オブジェクトが名称を持てば，その組織体が扱っている代表的製品をもキーワードの候補とし，それらをOR結合する"といったものがある。各組織体の代表的製品が何かを知るため，このルールは，先に述べた世界知識を参照する。図の例では，「富士通」の代表的製品は，「電算機」，「半導体」であるという知識が使われている。

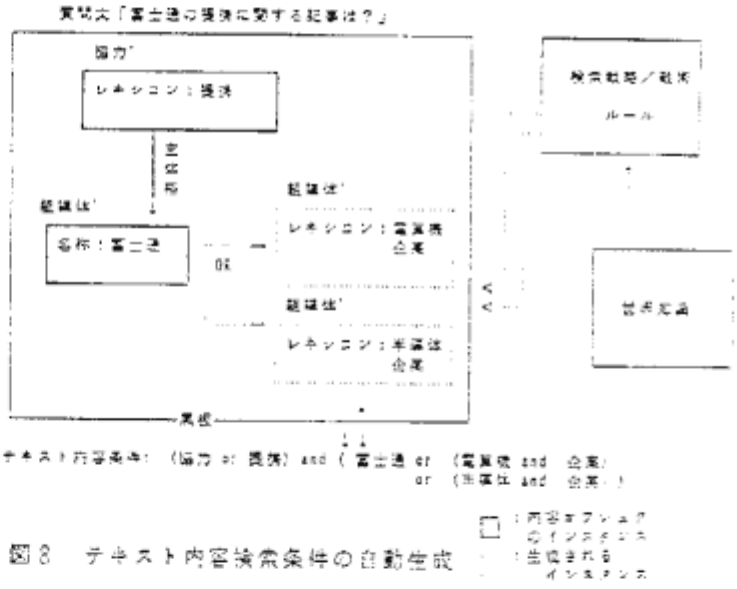

図8　テキスト内容検索条件の自動生成

6．内容マッチング

内容マッチングの働きは，質問文の意味構造と各テキスト文の意味構造を比較し，各テキストが質問とどの程度，意味的に一致するかを判定し，一致するテキストのみを検索結果とすることである。この機能の実現方法を考えるため，人間がどの質問にどのテキストが意味的に一致すると判定するかを調査した。その結果，その判定は，白黒はっきりしたものではなく，何か漠然としたものであり，主に，「何が何をどうしたか」という情報が如何に一致しているかに基づいていた。そこで，内容マッチングでは，意味構造の中でも，内容マッチング上重要と考えられる「何が」，

「何を」，「どうした」といった内容オブジェクトにのみ注目することにし，また，内容一致度は，その漠然さを考慮して数値でモデル化することにした。

内容マッチャーが扱う情報は，個体と命題的関係の２種類に分けられる。個体には，組織体や製品といったものがあり，内容オブジェクトの名詞的オブジェクトにほぼ対応する。命題的関係には，行動や状態などがあり，ほぼ内容オブジェクトの述語的オブジェクトに対応するが，名詞的オブジェクトの中にも，例えば，属性のように内容マッチャーで命題的関係として扱われるものもある。命題的関係に関する推論を実現し易くするため，図９に示すように，命題的関係に関する情報はオブジェクトによる表現ではなく，１階述語論理表現に変換される。そして，マッチャーの推論過程では，例えば，図の質問文述語論理式中の述語記号「協力」とテキスト述語論理式中の述語記号「供与」が意味的に対応し，第１引数である組織体同士が一致しなければならず，第２引数である変数Ｙはテキスト述語論理式中の形態オブジェクトに，第３引数の変数Ｚは変数Ｗに一致しなければならないといったことが，結論付けられる。

命題的関係に関する情報は，世界知識を使った推論により意味の一致度が計算され，個体に関する情報も，個体同士の内容一致度が，世界知識を使って計算される。図９の例では，「“ある組織体Ｘが，あることＹに関して，組織体Ｚに協力する”ということは，“その組織体Ｘが，あることＹを，組織体Ｚに供与する”ことに意味的に近い」ということ，及び「富士通の代表的製品としては，電算機がある」という世界知識が使われる。それぞれの意味の近さは，内容一致度数としてモデル化されているが，これらの一致度の計算方法は，ヒューリスティックに決められている。

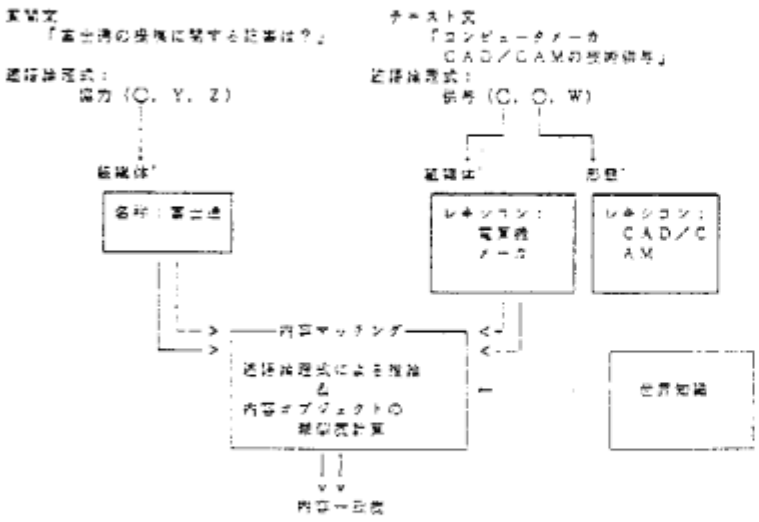

図９　内容マッチング処理

どのような計算方法が最終的に良いかは，大量文例の分析によって決める必要がある。

このように各テキストの内容一致度が計算されるので，ある閾値以上の一致度を持つテキストを内容一致したものと判定している。この閾値の決定も一致度の計算方法とからみ今後の課題である。

７．おわりに

本論文では，従来型の情報検索システムの問題点を整理し，用語選択労力からの開放，自然言語情報の内容検索への利用という２つの目標を掲げて，それぞれの問題を解決する方式，即ち，概念的キーワード検索式自動生成方式，内容マッチング方式を提示した。また，これら２つのモジュールの前提である質問文／テキスト文解析の処理方式も示した。現在，これらのモジュールは，パイロットバージョンが動き出したところであり，これによる評価は，また別に報告したい。

さて，以上のシステム設計では，対象テキストをある分野に限定した。これにより，システムが持つべき知識の範囲を限定し，実用に近いシステム開発を目指した。しかしながら，一般にいろいろなテキストが存在し，ある分野のテキスト情報が種々異なった文書に分散しているのが常である。例えば，現在のIRISの対象テキストである情報産業界の新聞記事も，一般紙の他の記事と混合して出てくる場合がよくある。そこで，IRISの３つ目の研究目標として，一般のいろいろな分野が混在したテキスト情報から，ある分野に限定したテキストだけを自動選択することを挙げている。先に述べた対象世界のモデルは，ある意味では，ユーザの興味の対象をモデル化したものであると捉え直すことができる。従って，IRISの知的検索系と同じ対象世界のモデルを少し異なる観点，即ち，テキスト内容がこのモデルで規定されている範囲に入っているかどうかという観点で使用することによって，この研究目標も達成できるのでないかと考えている。この考え方を出発点として，３つ目の研究目標にも取り組んでいきたい。

最後に，情報検索システムの高度化を目指した他の研究との比較を行う。これらのものとしては，細川らのシステム［細川 79］，佐藤らのシステム［佐藤他 84］，IR-NLI［Guida et al. 83］，知的ファイリングシステム［藤澤 86］，高松らのシステム［高松他 78］，SRI のシステム［Walker et al. 81，杉山 85］などがある。細川らのシステムは，テキストの方を解析し，自立語にロールと呼ばれる一種の格付けを行うものであり，IRISのテキスト解析に対応するが，検索系と融合する所までは，至っていない。佐藤らのシステムとIR-NLIは，質問文解析の結果からキーワードとその論理結合を得，シソーラスを使ったキ

ーワードの自動展開を行うもので，IRISの質問文解析とキーワード検索式自動生成に対応するが，内容マッチングはまだない。知的ファイリングシステムでは，IRISが目標とはしていない画像情報としてのテキストをもその守備範囲にしているが，IRISと共通する研究課題のところを比較してみると，テキストの意味表現を人手で入力するレベルであり，それに基づく検索はテキスト意味構造を利用したメニュー主導型である。これに対し，IRISでは，テキストの意味表現は自動生成しようとしており，検索も自然言語で行おうとしている。高松らのシステムは，IRISの質問文／テキスト文解析，内容マッチングに対応した処理を行っているが，IRISで取り扱っている命題間関係や外延内包関係に関するマッチング処理はない。さらに，高松らのシステムでは，テキスト内容を反映したテキストベースを構築するので，キーワード検索式自動生成は必要なくなる反面，従来型の情報検索システム上に作られたテキストベースを利用できない。SRI のシステムは，テキスト文，質問文の両方を解析し，深い知識に基づく推論により意味的な検索を可能にしようとするものであるが，大規模知識ベース構築の課題が大きく，まだシステムが稼働するところまでは至っていない。IRISでは，この知識を応用分野で閉じるような形で作成することにより，余り大きくなくてもシステムが動くようにしている。また，SRI のシステムには，高松らのシステムと同様，テキスト内容を反映したテキストベースを作るので，IRISのキーワード検索式自動生成に対応するモジュールはない。以上，各システムとの違いを個別的に述べたが，最後に，各システムに共通するIRISとの違いとして，これらのシステムがテキスト内容のみを対象にしており，書誌情報とテキスト情報の両方を受け付けるようにはなっていないことを挙げることができる。

謝辞　本研究は，第５世代コンピュータ・プロジェクトの一環として行われた。本研究に対して御支援頂くICOTの横井俊夫，岩下安男両室長に深く感謝致します。また，日頃から御指導頂く当社の林達也部門長付，上原貴夫部長，内田裕士室長に感謝致します。

参考文献

[Bates 79] Bates, M.J. "Information Search Tactics", Journal of the American Society for Information Science, pp. 205-214, 1979.

[Chikayama 84] Chikayama, T. "ESP Reference Manual", ICOT Technical Report TR-044, 1984.

[Guida 83] Guida, G. and Tasso, C. "IR-NLI: An Expert Natural Language Interface to Online Databases", Proc. of ACL '83, pp. 31-38, 1983.

[Hobbs 84] Hobbs, J.R. "Building a Large Knowledge Base for a Natural Language System", Proc. of COLING-84, pp. 283-286, 1984.

[Ishikawa 86] Ishikawa, H., Izumida, Y., Yoshino, T., Hoshiai, T. and Makinouchi, A. "A Knowledge-based Approach to Design a Portable Natural Language Interface to Database Systems", IEEE Proc. of the International Conf. on Data Engineering, pp. 134-143, 1986.

[Sugiyama et al. 84] Sugiyama, K., Kameda, M., Akiyama, K. and Makinouchi, A. "Understanding of Japanese in an Interactive Programming System", Proc. of COLING-84, pp. 385-388, 1984.

[Walker et al. 81] Walker, D.E. and Hobbs, J.R. "Natural Language Access to Medical Text", IEEE Proc. of the 5th Annual Symposium on Computer Applications in Medical Care, pp. 269-273, 1981.

[絹川 79] 絹川 "情報検索のための日本語解析", 情報処理, Vol.25, No.3, pp.367-371, 1984.

[佐藤他 84] 佐藤，斎藤，菊地 "特許情報検索のための日本語質問文解析", 情報処理学会論文誌, Vol.25, No.3, pp.365-371, 1984.

[杉山他 84] 杉山，秋山，亀田，牧之内 "対話型自然言語プログラミングシステムの試作", 電子通信学会論文誌, Vol. J67-D, No. 3, pp. 297-304, 1984.

[杉山 85] 杉山 "自然言語による内容検索に向けて", 情報処理学会自然言語処理研究会資料47-6, pp. 55-61, 1985.

[高松他 78] 高松，藤田，西田 "係り受け関係に基づく文献の検索", 情報処理, Vol.19, No.12, pp.1150-1157, 1978.

[藤澤 86] 藤澤，畠山，宇野，藤堀，星野 "高度ファイリングの理念と要素技術ー文書理解と知的ファイリングー", 情報処理学会日本語文書処理研究会資料7-4, pp. 1-8, 1986.