HOKKAIDO UNIVERSITY

Title	問題解決機構としての文献情報検索システムの構成
Author(s)	前田, 隆; 桃内, 佳雄; 沢村, 一
Citation	北海道大學工學部研究報告, 84, 101-108
Issue Date	1977-07-11
Doc URL	https://hdl.handle.net/2115/41423
Type	departmental bulletin paper
File Information	84_101-108.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第84号（昭和52年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 84 (1977)

問題解決機構としての文献情報検索システムの構成

前田　隆*　桃内佳雄**　沢村　一**

（昭和51年12月28日受理）

A Construction of Information Retrieval System as a Problem Solving Machine

Takashi MAEDA, Yoshio MOMOUCHI and Hajime SAWAMURA

(Received December 28, 1976)

Abstract

Various information systems, in particular, information retrieval systems or question answering systems are able to consider as a problem solving machine from the point of the user's view.

In this paper, we have presented a general model of such a system, and described a document information retrieval system as a implementation of such a model. To information structure in this system, a method of analysis of hierarchical semantic structure of document information is applied.

Some features of this system are described with examples for interactive retrieval process which is a process to supply the user's request of document informations.

1. まえがき

計算機に一定の情報を蓄積しておき，これらをある特定の仕様に従ってとり出したり，あるいは蓄積情報を利用して他の入力情報を処理し，その結果を出力するようなプログラム（ないしはプログラム言語）はその装置を含めて，一般に情報システムとよばれている。各々のシステムはどのような情報を蓄積しているか，どのようなとり出し方を許すか，あるいはどのような入力情報をどのように処理するかなどその処理対象と処理方式によって，情報検索システム，質問応答システム，言語処理システムあるいは音声や画像処理システムなどとよばれる。

本稿ではこれらの情報システム，とくに情報検索システムおよび質問応答システムはその利用者の立場からは一つの問題解決（補助）システム[12]として見なし得るという観点から，そのようなシステムの一般的モデル[10),13)]について考察する。次にこのモデルの一つの具体化として文献情報検索システムを構成する。このシステムはとくに文献情報の階層的な意味構造分析[8)]における一つのレベルに着目し，この方法を適用した情報構造にもとづくものである。利用者の検索要求をシステムとの会話により，適当な出力を得るまで追求するプロセスが例によって示される。

2. 情報システムのモデル

情報システムの基本的機能は，外部からの何らかの情報を受理し，これに適当な処理を施し，

* 工業数学講座
** 情報システム工学講座

その結果を何らかの形式で出力するということである。外部から情報を入力する主体をこのシステムの利用者とすると，一般に利用者が当面している問題（質問）が入力されることになる。そうすると，この入力に対するシステムからの出力は，その入力に対するシステムとしての解答であると考えることができる。すなわち，このシステムはこうした利用者にとっては一種の問題解決（補助）機構としての意義をもつことになる。システムがどのような解答を生成することができるかは，そのシステムがどのような知識を持っているかに依存する。たとえば，多くの文献検索システムの場合は，利用者が関心を持つ問題に対して，この場合の知識を蓄積してある文献集合であるとすれば，“関連する文献情報”を提供するということになる。このように情報システムはどのような知識をどれだけ蓄積しているか，またそれらの知識を利用してどのような処理が可能であるかによって特徴づけられる。文献情報検索システムは利用者の問題（質問）に対して直接的な解答を生成するのではなく，関連情報の提供にとどまるという意味で，低次の問題解決（補助）機構であると考えることができる。

以上の点およびシステムの応答に対応して利用者の要求を会話的に入力して，問題解決に近づけていくプロセスも含めて，問題解決（補助）機構としての情報システムの一般的モデルは図1のように表現することができるであろう。このようなモデルにおける各サブシステムの基本的動作は次のようなものである。

2.1 入力解析部

利用者は問題（質問）をそのシステムで定められた記述方式にもとづいて表現し，入力するものとする。これらの入力に対して入力解析部はそのシステム固有の「文法・辞書」を参照して入力の構文解析を行なう。この結果，通常の文献検索システムであれば，“索引語”とそれらの間の“論理的関係”が得られ，言語処理システムであれば，“語い”とそれらの間の関係をその文法固有の形式において表現する“構文的構造”が得られるであろう。このモデルでは，入力解析結果の情報を，“語い”とその“論理的関係”とで表わすことにする。

2.2 制御部

ここでは転送されてくる情報を判定して，次の処理の方向を制御する。次の二つの場合がある。

(1) 入力解析部から制御が移された場合は，その“語い”と“論理的関係”から定められる意

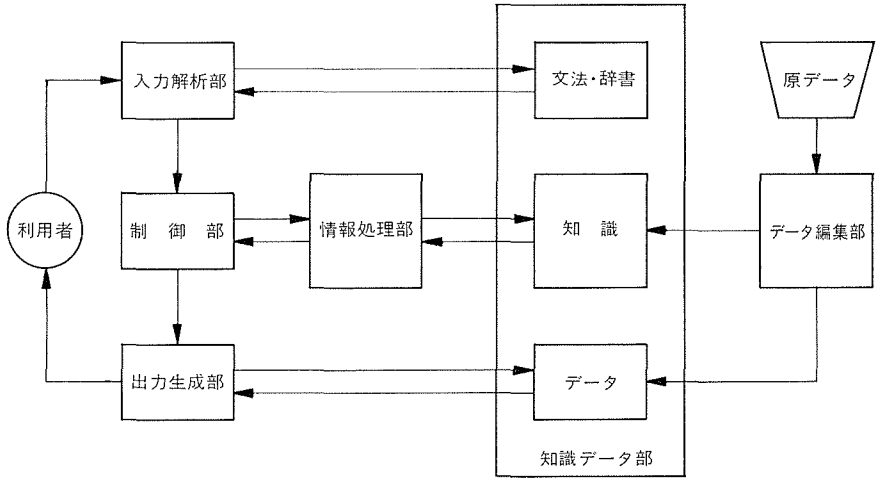

図1 情報システムの一般的モデル

味に従って，情報処理部あるいは出力生成部へと制御を移す。

(2) 情報処理部から制御が移された場合は，その処理結果を含め，出力生成部へと制御を移す。

2.3 情報処理部

情報処理部は一般に計算，探索，推論などのいわゆる“手続き”を含んでおり，必要に応じて「知識」として蓄積されている計算方式，探索手法，推論システム[12]などを参照して，その手続きを実行するサブシステムである。通常の文献検索システムの場合には 2.1 の結果得られた索引語とその論理的関係は，情報処理部においては“検索仕様”とその“データ”となり，ハッシングやクラスタリング技術[9] あるいはインバーテッド・ファイルなどで組織化[14]されている構造化データからその仕様に照合するデータ群を検索することになる。自然語処理システムであれば，入力情報のもつ“構文的あいまいさ”，“省略語”，“代名詞の同定”など自然語理解[15),16]に特有の諸問題を解く部分であって，それらを解くに必要な「手続き」と知識データ部における「知識」とが用意されていなければならない。

2.4 出力生成部

制御部を通じて出力生成部に制御が移される場合は三種類あり，各々に対応した応答をする。

(1) 各サブシステムにおける構文的エラーの存在による場合は，その性質に応じた適切なメッセージを出力する。

(2) 各サブシステムにおける処理の終了状態，内容を何らかの形式で出力することができる。この機能は会話型処理にとっては不可欠のものである。

(3) 最終的出力についても利用者の要求に応じて，たとえばソート・マージや自然語生成などの機能が要求されることがある。生成される出力内容は単なる計算結果から，文献集合，自然語の解析結果，翻訳の結果，あるいはシュミレートされたロボットの行動プラン[11]やプログラム（手続き）などとそれぞれのシステム固有のものとなる。

2.5 データ編集部

原データをコード化して格納する機能と各々のシステムに固有な必要情報の抽出とそれにもとづくデータ群の組織化による“構造化データ”を作成し，「知識」として格納するという二種類の機能をもつ。またデータの追加，更新，削除などデータベース管理などもここを通じて行なうもので，一般にはシステム管理者が実行するものである。

2.6 知識データ部

知識データ部はそれぞれ役割の異なる三つの構成部分からなる。入力情報を構文的にチェックするための「文法・辞書」，各々のシステム固有の計算方式や，探索手法，推論システムなどその対象とする世界に関する「知識」，および処理結果の構成に必要な，これもシステムに固有な数値（表），図形（パターン），文書（文献）などのコード化されている「データ」である。

これらの外に利用者がその目的に応じて一時的に“知識”を随時定義して，これを利用することは可能であるが，これらは知識データ部における「知識」とは根本的に異なる性格のものであり，その利用者の使用が終了した時点で解消されるものとする。

3. 文献情報検索システムの構成

2. で述べた情報システムの一般的モデルを具体的な文献情報検索システムについて検討する。文献情報検索システムとは，“文献の全文，文献代用物（たとえば抄録），またはその名称および所在（すなわち，その完全な書誌事項）を検索できるシステム”[1]である。このように利用者の情

報要求をできるだけ満たすようなシステムを構成しようとするとき，最も重要な問題は，原データである各々の文献の内容をできるだけ正確に抽出し，処理しやすい形式に表現することである[6),7)]。ここでは文献情報の階層的な意味構造分析の一つの手法[3),8)]にもとづいて，文献の情報構造を抽出する。通常，情報構造はその認識個体を事象とよび，事象はいくつかの属性から構成され，それらは各事象固有の属性値をとることにより，具体的に表現されるものとされている[2)]。我々がとる手法においては，各認識個体は文献（論文）であるが，その主要な内容は抄録に表現されているという立場から，抄録を分析の対象（事象）とし，属性に相当するものは，抄録における各文が抄録の構成上果している役割（機能）をとるものとする。そして属性値に対応するものとしては，その文におけるその機能の下でのいくつかの重要語をとるものとする。こうして各文献は書誌事項（標題，著者，雑誌名，巻・号，発行年など）と共に，その主要な内容を表わすものとして，属性としての機能名とその値としての重要語の対によって，形式的に表現されることになる[4),6)]。

この文献情報検索システムは，利用者の問題解決に寄与するため，システムとの会話により，格納されている「知識」や「データ」の一部を利用してできるだけ利用者の情報要求を満たしていく過程の一つの実現と考えることができる。

3.1 システムの概要と特徴

図１のモデルに対応して，このシステムの概要を図２に示す[4),6)]。

図１の一般的モデルと比較するとこの文献情報検索システムにおいては，

(1) 入力解析部で参照する「文法・辞書」は「コマンド系・規約」と具体化されている。

(2) 情報処理部がそこで利用されるべき「知識」の性格に応じて二つの異なる機能に分化されている。一つは「語い」に関する知識をとりだす推論機能，他は蓄積文献に関する知識をとり出す検索機能からなり，利用者はこれらの機能を自由に利用して，システムの知識を参照することができる。

(3) 出力生成部では処理結果の出力形式を選択できるように工夫してある。

(4) データ編集は現在のところ，人間が行なっているが，この自動化についての試みが続けられている[7)]。

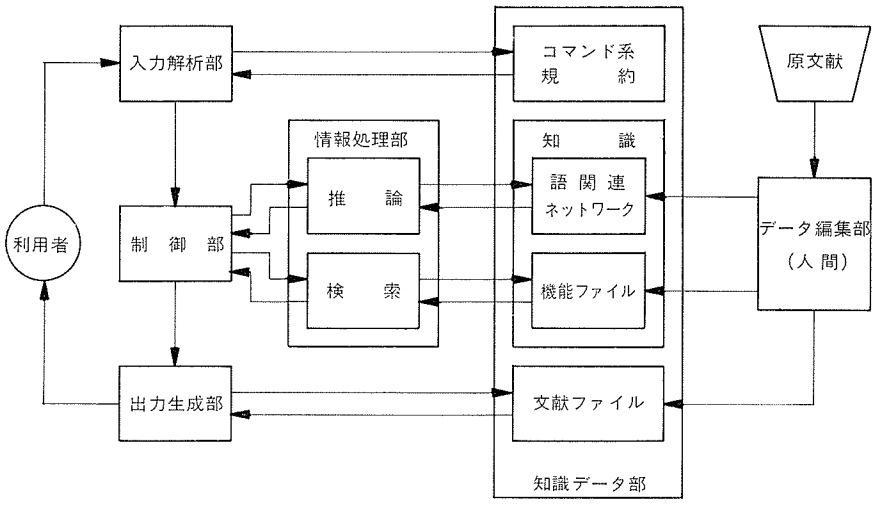

図２ 文献情報検索システムの構成

などを具体的な特徴としてあげることができる。これらの特徴をふまえて，システムとの会話を通じての検索例は 3.3 で述べられる。

3.2　語関連ネットワークと機能ファイル

文献情報検索システムにおける「知識」は，通常システムにおける使用可能な語いに関してそれらの相互の関係を表わしているシソーラス，およびそれらの語いからみての文献の分布を示すインバーテッド・ファイルなどが代表的なものであろう。他に文献集合をそれらの“近似度”に関して何らかの方法でクラスター化しておく場合もある。

本システムにおいては，こうした「知識」として「語い」に関する意味的関連[5),17)]を次の四種類の関係で規定している「語関連ネットワーク」が用意されている。(図 3 参照)

(1) 同義関係，(2) 反意・対語関係，(3) 上位関係，(4) 下位関係。

次に文献に関する「知識」として，各文献の抄録に対する文機能分析[8)]にもとづく，次の 8 種類の機能ファイルを構成している。

①背景 (B)，②主題 (S)，③手法 (M)，④関連 (C)，⑤結果 (R)，⑥適用例 (E)，⑦応用分野 (A)，⑧考察 (D)。

各ファイルの内容は各々の機能をもつ文を含んでいる文献と，その文の意味特徴[5),17)]としての重要語の対から成るものである。(かっこ内の英文字はシステムで用いている記号である。)

3.3　コマンド系と会話による検索

システムの利用は次の 8 種類のコマンドを用いて会話により行なうことができる。端末でのタイプライター，あるいはディスプレー上に応答される出力をみながら適当と思われる時点で，出力形式を指定して，求める文献情報を得ることができる。

(a) Q: =〈質問式〉; ……「語い」の和と積からなる論理式を引数とする「機能」についての論理式。(以下 Q 文と呼ぶ。)

例　$Q: =F_1(W_1 * W_2 + W_3) + F_2(W_4) * F_3(W_5 + W_6)$;

($F_i(i=1, 2, 3)$ は機能，$W_j(j=1, ..., 6)$ は語いを表わす)

(b) QEND; ……1 つの質問の終りをシステムに伝える。システムは次の Q 文を受理できる状態となる。

(c) SEARCH; ……既に与えてある Q 文による検索命令。応答として，該当文献数を“HCNT=n”の形式で出する。

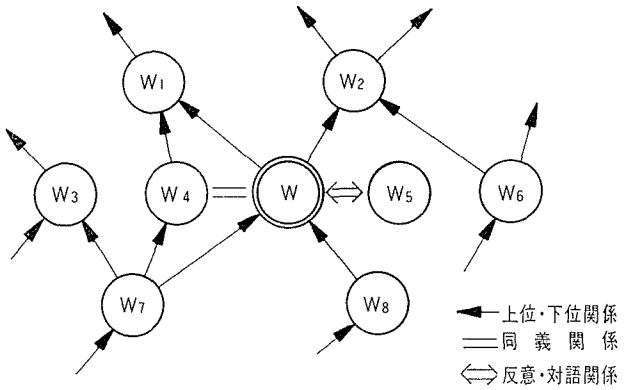

図 3　語関連ネットワーク

(d) INFER (W, I); ……「語い」W についてタイプ I (=0, 1, 2, 3) の推論を行なう。図3に対応して次のような語いが参照される。

I=0: 同義語 (W_4), 反意・対語 (W_5)

I=1: 類義語 ⓐ 上位について (W_6), ⓑ 下位について (W_3)

I=2: 上位語 (W_1, W_2), 下位語 (W_7, W_8)

I=3: 近接語……その語（語幹）を部分ストリングとして含む語い

(e) SUBST (F, W_1, W_2); ……既に与えてある Q 文において機能 F の語い W_1 を W_2 でおきかえる。

(f) LIST; ……Q 文の内容を各機能ごとに出力する。(Q 文の確認に用いる)

(g) OUTPUT (I): ……検索結果の出力命令で，三つの選択ができるようになっている。

I=0: 書誌事項と抄録全文

I=1: 書誌事項と抄録の文のうち，Q 文で指定した機能に対応する文集合

I=2: 書誌事項のみ

(h) END; ……システム利用の終了を伝える。総質問個数などを出力して終る。

END 文が使われるまでその他のコマンドを何回でも使うことができる。

図4は上記コマンド系を用いて，会話により適当な応答を求めていくプロセスの一部を示している。(1), (2), (3) の三つの Q 文を中心とした会話例で，# 印のついた文は利用者が入力したコマンドを示しており，# 印はシステム側からの入力要求である。# 印のつかない文は（最初の4行を除いて）それぞれすぐ上のコマンドに対するシステムの応答を示している。

まず (1) の場合，Q 文の意味は「機能」が B，すなわちその文献の主題の「背景」について述べている文において「語い」として “FORMAL LOGIC” あるいは “INFERENCE” を含むもの，ということである。この Q 文が受理されたという意味で，# 印が出力されてくるので，受理された Q 文の内容を確認するため，次に「LIST;」と入力する。次の行の「B FORMAL LOGIC+INFERENCE」がシステムの応答である。これにより Q 文が正しく受理されていることが確認されたので，# 印の次に検索命令「SEARCH;」を入力する。これに対する応答は「HCNT=0」，すなわちこの Q 文に該当する文献は登録されていない，ということである。そこで「語い」を変えてみることにし，“INFERENCE” の代りに “REASONING” とする代入命令「SUBST (…);」を用いる。次の行の「B: INFERENCE/REASONING」はこれに対する確認応答であり，その次の「LIST;」命令およびその応答「B FORMAL …」で正しく代入されていることがわかる。次にこの語いを変えた Q 文に対して「SEARCH;」を入力すると，今度は「HCNT=1」すなわち該当文献が1件であるという応答である。そこで「OUTPUT (0);」，すなわち「抄録」全文を含めた出力命令を入力しているが，これに対する出力結果は省略してある。次の「QEND;」でシステムは次の Q 文を受理できる状態に戻る。(2) の Q 文は S，つまり「主題」について述べている文で，かっこ内の5個の語いのどれかを含むもの，という意味であり，これに対する「SEARCH;」の結果，該当文献は9件となっている。現在は Q 文の各機能に対して語いは最大5個までしか許していない。

(3) は Q 文での機能が S (主題)，A (応用分野)，D (考察) と複数個からなる論理式の場合である。このプロセスで語いに関する推論，「INFER (SEARCH, 2);」を用いてシステムの保持している知識を参照している。推論のタイプは2であり，従って “SEARCH” の上位語，下位語がそれぞれ「SUPERORDINATE」，「SUBORDINATE」と分けられて出力されている。尚，図の右下に「INFER (SCENE, 3);」に対する出力が示されている。「QEND;」につづいて，(4)

```
**************************
BUNKEN KENSAKU SHORI KAISHI
**************************
*** @-BUN NYURYOKU ***

(1) #@:=B(FORMAL LOGIC+INFERENCE);
    #LIST;

    B  FORMAL LOGIC+INFERENCE;

    #SEARCH;

         HCNT=     0
         **********
    #SUBST(B,INFERENCE,REASONING);

    B:INFERENCE   /REASONING

    #LIST;

    B  FORMAL LOGIC+REASONING;

    #SEARCH;

         HCNT=     1
         **********
    #OUTPUT(0);

      *** SHUTURYOKU KAISHI ***
           ... (略)
      *** SHUTURYOKU OWARI ***

    #@END;

    *** @-BUN NYURYOKU ***

(2) #@:=S(THEOREM PROVER+THEOREM PROVING+PROBLEM SOLVING+PROBLEM SOLVER+ROBOT);
    #SEARCH;

         HCNT=     9
         **********

    #@END;

    *** @-BUN NYURYOKU ***

(3) #@:=S(SEARCH)*A(GAME TREE)+D(CHESS PROGRAM);
    #SEARCH;

         HCNT=     1
         **********
    #INFER(SEARCH②);


    WT='SEARCH' ;
    IT=     2 ;

        * SUPERORDINATE ...
      ( GAITO SURU SUP_SUB WORD WA ARIMASEN )

        * SUBORDINATE ...
          BRUTE FORCE SEARCH
          ORDERED SEARCH
          HEURISTIC SEARCH
    #SUBST(S,SEARCH,HEURISTIC SEARCH);


    S:SEARCH                          /HEURISTIC SEARCH

    #LIST;

    S  HEURISTIC SEARCH;

    A  GAME TREE;

    D  CHESS PROGRAM;

    #SEARCH;

         HCNT=     1
         **********
    #@END;

(4) #END;

    **************************
    BUNKEN KENSAKU SHORI OWARI
    @ NO KOSU=   7
    **************************
```

```
#INFER(SCENE(3));


WT='SCENE' ;
IT=     3 ;

     * CONTIGUITY (FOR A STEM) ...
       HEURISTIC SCENE DESCRIPTION
       POLYHEDRAL SCENE
       ROAD SCENE
       SCENE ANALYSIS
       SCENE DESCRIPTION
       THREE-D SCENE
```

図4 システムとの応答例

で「END;」と入力し，システム利用の終了を伝えている。最終メッセージで「Q の個数=7」となっているが，上記三つ以外は省略している。

以上の例で本システムの大まかなふるまい，会話型検索のし方，特徴などを理解することができるであろう。尚，現在はシステムにおける登録文献が少ないので検索の質や効率などについての検討は更に充実した段階で考察する予定である。

4. あ と が き

情報システムを利用者にとっての問題解決（補助）機構とみなす立場から，その一般的モデルを構成し，その基本的機能について検討した。個々の情報システムはとくに「情報処理部」の能力および対応する「知識」の充実さによって種々に特徴づけられることが明らかにされた。さらにこうした立場から，文献情報の階層的な意味構造分析の手法を適用して，比較的低次の問題解決（補助）機構である文献情報検索システムを構成した。文献内容を反映している抄録の文機能分析による情報化表現は利用者の情報要求の表現にも多様性をもたらすことになっている。また，システムとの会話を通じて語いに関する「知識」などを参照しつつ，情報要求を満していくというプロセスの一部が例を通して示された。

今後，登録文献の充実およびその下での検索効率・質などに関する検討などが残されている。更にデータ編集の自動化の実現も大きな課題である。

謝 辞

日頃種々の点でご助言をいただいている本学部情報数理工学第一講座河口至商教授に深く感謝致します。

尚，本システムは北大大型計算機センター FACOM 230-75 を使用し，全プログラムは PL/I 言語で約1100ステップからなっている。プログラム作成やファイル利用などに関して，長田博泰氏，相良劭氏をはじめセンター関係者に大変お世話をいただいた。ここに記して感謝の意を表します。

参 考 文 献

1) F. W. ランカスター著，中村幸雄監訳：情報検索システム（1972年），紀伊国屋書店.
2) 中原啓一著，電子通信学会編，情報検索（昭和49年），電子通信学会.
3) 桃内佳雄，前田 隆，沢村 一：電気四学会北海道支部連合大会講演論文集（1975年）.
4) 前田 隆，桃内佳雄，沢村 一：電気四学会北海道支部連合大会講演論文集（1975年）.
5) 前田 隆，桃内佳雄，沢村 一：情報処理学会第16回全国大会講演論文集（1975年）.
6) 前田 隆，桃内佳雄，沢村 一：第8回計測・制御に関する北海道研究集会論文集（1975年）.
7) 前田 隆：情報処理学会第17回全国大会講演論文集（1976年）.
8) 桃内佳雄，前田 隆，沢村 一：北大工学部研究報告（本号），(1977年).
9) D. B. Crouch: A Clustering Algorithm for Large and Dynamic Document Collections, Southern Methodist University, Ph. D., 1972.
10) C. J. Crouch and D. B. Crouch: Information Systems COINS IV, Plenum Press (1974), 219-237.
11) R. E. Fikes and N. J. Nilsson: Artificial Intelligence **2** (1971), 189-203.
12) N. J. Nilsson: Problem-Solving Method in Artificial Intelligence, McGraw-Hill, 1971.
13) B. A. Shapiro: A Survey of Problm Solving Languages and Systems, Maryland University, TR-235, 1973.
14) G. Salton: Automatic Information Organization and Retrieval, McGraw-Hill, 1968.
15) R. C. Schank: Conceptual Information Processing, North-Holland, 1975.
16) T. Winograd: Understanding Natural Language, Academic Press, 1972.
17) 池上嘉彦：意味論（1975年）大修館書店