

文献情報データベースを情報源とする知識ベースシステム

宇陀則彦 石塚英弘 山本毅雄
図書館情報大学

問題解決を行う知識ベースシステムを構築するうえで課題となるのは、問題領域の知識獲得と解決に適した知識の構成及び表現である。本論文は、専門的な知識を多量に蓄積したデータベースを情報源とすることで問題領域の知識を多量にかつ効率的に獲得できることを示す。また、問題解決のための知識を個別的知識と一般的知識という2種類の知識から構成し、これを利用することにより問題解決ができることを示す。個別的知識は、データベースの内容を確実な事実としてとらえたものとし、一般的知識は、個別的知識に関連する概念の階層関係を表現したものとする。構築した知識ベースシステムは、フレームで表現された個別的知識と一般的知識を適宜利用して問題解決を行う。なお、情報源のデータベースには、詳細な情報を含んでいることで有名な CAsSearch を選んだ。

Knowledge Base System Acquiring Knowledge from Bibliographic Database
Norihiko Uda Hidehiro Ishizuka Takeo Yamamoto
University of Library and Information Science
1-2 Kasuga, Tsukuba, Ibaraki 305, Japan

Knowledge acquisition of specialized domain and knowledge representation suitable for problem have difficulty in building knowledge base system. In this paper, we describe effective knowledge acquisition from database which has a large quantity of special knowledge. And we describe problem solving using knowledge which consists of specific knowledge and general knowledge. Specific knowledge is contents of database as fact. General knowledge is hierarchical structure in relation to specific knowledge. Our knowledge base system performs problem solving using frame represented specific and general knowledge. We use CAsSearch which has detailed information as database of information sources.

1 はじめに

知識ベースシステムは、知識を利用して問題解決を行うシステムである。

知識ベースシステムによって問題解決を行うためには、対象となる問題領域の知識を獲得し、問題解決に適切な構成にして計算機上に表現して利用しなければならない。

その際、多量で効率の良い知識の獲得と、問題解決のための知識の構成と表現は大きな課題である。

これまでの知識ベースシステムでは、専門家や専門書を情報源として知識を獲得している。とくに、エキスパートシステムの場合、専門家から知識を獲得する必要があるが、それはインタビューによって行われている [1][2][3]。しかし、インタビューによる獲得は、知識が多量の場合、専門家が自分の知識を明確に表現できない場合困難である。また、専門書からの自動的な知識の獲得は、自然言語理解の技術、とくに文脈解析が現時点では不十分なため困難である。

これに対して、専門的な知識を多量に蓄積し表現しているものにデータベースがある。データベースのデータは、個々の事実が明確に記述してあるため、専門家や専門書から知識を獲得する場合の問題は生じない。そのため、データベースのデータを転用し、知識ベースシステムに利用する試みがいくつかなされている。例えば、リレーショナルデータベースのテーブルを命題・述語論理型知識の形式に変換して知識ベースを構築したもの [4]、テーブルをフレームに変換して知識ベースを構築したものなどがある [5]。そのほか、あらかじめ用意した知識抽出用ルールを用いて、データベースから診断用の知識を抽出する試みもある [6]。

また、データは必要な情報が陽に表現されているものであるのに対し、知識は陰に表現されるものを含んでいる。知識を利用する場合には、陰に表現されたものを陽に表現することが必要である [7]。

一方、問題解決のための知識は、セマンティックスが明示的でなければならない。ところが、データベースのデータは、表現は明示的であるがセマンティックスは暗示的である。それゆえ、データベースのデータを問題解決のための知識として利用するためには、セマンティックスを明示的にしなければならない。

本研究では、文献情報データベースを情報源にすることによって、セマンティックスが暗示的である文献情報データベースのデータを知識ベース化し、明示的な知識として利用できることを示す。また、この方法によって多量の知識を効率よく自動的に獲

得できることを示す。

本研究では、情報源の文献情報データベースとして、詳細な情報を含んでいることで有名な CASearch を選んだ。

2 CASearch の構成

CASearch には、書誌事項、キーワード、一般事項索引、化学物質索引が収録されている。一般事項索引、化学物質索引は、フレーズ形式で表現されており、このフレーズを名詞句として復元することでキーワードより詳しい論文の内容に関する情報が得られる。また、文脈解析は必要ないので、セマンティックスを明示的にして知識ベースを構築するのは抄録などから知識ベースを構築するより容易である。

Fig.1. に TOOL-IR を用いて得たある論文の CASearch 出力例を示す。

```
( 70 )
CA: 111/15/178023F SC: CA105003 DT: J
TI: Chemical weed control in spring turnip rape with and without stubble
cultivations
AU: Jenneus, Birger
LW: Swed. Farmers' Selling Purch. Assoc. CI: Linköping PC: S-582 58
NA: Swed.
JN: Swed. Crop Prot. Conf. VO: 30th IS: 2 PP: 112-14 PY: 89
CO: SCPRE4 LA: Eng
KW: herbicide spring turnip rape Elymus; weed control spring turnip rape
-----
GCH: Barley GCM: spring
GTM: herbicides for control of Elymus repens in spring turnip rape residual effects in
-----
GCH: Brassica napus oleifera GCM: spring
GTM: Elymus repens control in, herbicides for
-----
GCH: Elymus repens
GTM: control of, in spring turnip rape, herbicides for
-----
GCH: Herbicides GCM: postemergence
GTM: Elymus repens control by, in spring turnip rape, residual effect in
spring barley in relation to
-----
GCH: Weed control
GTM: of Elymus repens, in spring turnip rape, herbicides for
-----
CRN: 74051-80-2 CMF: C17H20NO3S
CHP: 2-Cyclohexen-1-one
CSB: 2-[1-(ethoxyimino)butyl]-5-[2-(ethylthio)propyl]-3-hydroxy-
CNC: NT1
CTM: Elymus repens control by, in spring turnip rape, residual effect in
spring barley in relation to
-----
CRN: 76578-14-8 CMF: C16H17ClN3O4
CHP: Propanoic acid CSB: 2-[4-[(6-chloro-2-quinoxalinyloxy)phenoxy]-
CMD: ethyl ester CNC: NT1
CTM: Elymus repens control by, in spring turnip rape, residual effect in
spring barley in relation to
-----
CRN: 87237-48-7 CMF: C19H19ClF3NO5
CHP: Propanoic acid
CSB: 2-[4-[[8-chloro-5-(trifluoromethyl)-2-pyridinyloxy]phenoxy]-
CMD: 2-ethoxyethyl ester CNC: NT1
CTM: Elymus repens control by, in spring turnip rape, residual effect in
6076 spring barley in relation to
```

Fig.1. CASearch の出力例

CASearch の書誌事項は、別々の項目として収録されており、TOOL-IR では、項目を CA:, TI: 等のラベルを用いて表している。

一般事項索引 (General subject index) は、概念見出し語 (conceptual heading)、概念補足説明 (concept modification)、説明文 (text modification) の3種類からなり、ラベルはそれぞれ "GCH:", "GCM:", "GTM:", である。

GCH は論文の重要な概念を表している。GCM は GCH の概念を補足し、GCH をより理解しやすいものにしていく。GTM はカンマで区切られたいくつものフレーズからなり、フレーズは、論文の内容に関して重要な順に並んでいる。

化学物質索引 (Chemical substance index) は、CAS レジストリ番号 (registry number)、化学物質索引見出し語 (heading parent)、置換基 (substituent)、説明文 (text modification) の4種類からなり、ラベルはそれぞれ "CRN:", "CHP:", "CSB:", "CTM:" である。CHP と CSB は化学物質の名前を表す。CTM は GTM と同様に論文の内容を表す。

3 知識ベースシステムによる問題解決の構想

問題解決を行う知識ベースシステムを構築する場合には、一般の原理を見つけたして適用するより、問題領域固有の個別の知識を重要視したほうが効果的である^[8]。

本研究では、知識を個別の知識と一般的知識に分け、その2種類の知識を適宜操作することによって問題解決を行う。

ここでは、文献情報データベースに収録されている論文の内容を確実な事実として捉えたものを個別の知識とし、個別の知識間の関係や関連した事柄を記述したものを一般的知識とする。

個別の知識は、自然言語処理の格文法の手法を用いて、文献情報データベースから知識ベースのフレームに自動的に変換されるが^[9]、それは知識ベースマネジメントシステムによって行われる。格文法の格はフレームのスロットに相当する。一般的知識は、知識ベースマネジメントシステムによって動的に抽出することもできるし、人手によってあらかじめフレームで表現しておくこともできる。

本研究の問題解決の方法は、まず、対象の事実が記述してある個別の知識に答を求める。答が個別の知識にあれば、問題解決は終了する。もし、答が個別の知識になければ、関係や階層構造を記述した一般的知識を利用して求める事実に関連する概念を調べ、その概念によって再び個別の知識に答を求める。

その事実が個別の知識にあれば、それを答えて問題解決は終了する。

例えば、ある事実が個別の知識にない場合、その事実の同位概念となっている事実を一般的知識で調べ、その事実を個別の知識に求める。その事実が個別の知識にあれば、それを答える。その事実も個別の知識になければ、さらに、別の同位概念、あるいは上位概念となっている事実を一般的知識で調べ、その事実で個別の知識を探索する。

以上のように、個別の知識と一般的知識を利用して問題解決を行う。求める答は常に事実としての個別の知識に求め、一般的知識は、問題解決を補助する知識として利用される。

問題解決の過程で個別の知識の関係を記述した新しい知識を抽出することもできる。この知識は、一般的知識に書き込まれ、問題解決のための新しい知識として利用される。

例えば、問題解決の過程である化学物質がある現象の原因であるという事実が個別の知識にあることが分かっていたら、その化学物質と現象を原因・結果の関係として一般的知識に書き込み、新しい知識として利用できるようにする。

4 システムの構成と機能

4.1 全体の構成

本研究のシステムは、大きく知識ベースと知識ベースマネジメントシステムから構成される。

知識ベースは、さらに個別の知識を表現する知識ベースと一般的知識を表現する知識ベースから構成される。本論文では、それぞれを個別の知識ベース、一般的知識ベースという。

知識ベースマネジメントシステムの構成は、さまざまな捉え方があるが^{[10][11]}、本研究では、知識獲得機構、推論機構、説明機構から構成されるものとする。

知識獲得機構は、文献情報データベースから知識を獲得し、個別の知識の知識ベースに変換する部分であり、変換のためのルールと辞書を持っている。推論機構は、個別の知識と一般的知識を利用して問題解決を行う部分である。また、説明機構は推論の過程を示したり、一般的知識の構造を示す部分である。

Fig.2. にシステムの構成を示す。

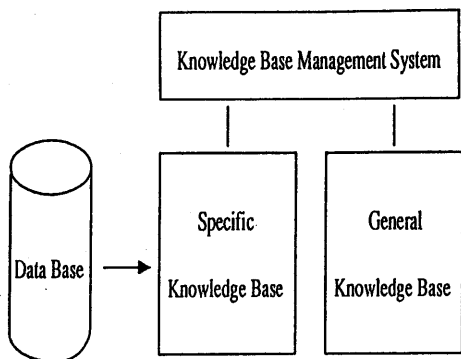


Fig.2. システムの構成

4.2 知識ベース

4.2.1 知識ベースの構成

個別的知識ベースは、書誌事項を表現している *bib_frame*、一般事項索引を表現している *gsi_frame*、化学物質索引を表現している *csi_frame* という3つのフレームから構成される。一般的知識ベースは、*con_frame* というフレームから構成される。

bib_frame は、CASearchの書誌事項のラベルの名称をそのままスロットとした。

gsi_frame は、一般事項索引のセマンティックスを表す名称をスロットとした。名称と意味は次のとおりである。

cano - CAS レジストリ番号を表す。

subject - 主題を表す。

instrument - 道具、手段を表す。

time - 時間を表す。

purpose - 目的を表す。

place - 場所を表す。

source - 材料、原料を表す。

micro_range - 狭い概念を表す。

macro_range - 広い概念を表す。

relation - 関連を表す。

csi_frame は、化学物質索引のラベルの名称をそのままスロットとした。

con_frame は、1つの概念が1つのフレームを形成しており、スロットによってフレーム間の関係を示している。知識ベース管理システムは、概念間の関係を示したフレームをたどることによって推論を行う。

con_frame のスロットの名称と関係は次の2種類である。

isa - 階層関係を表す。

cause - 原因と結果の関係を表す。

A *isa* B の関係では、A がフレームの名称で B がスロット *isa* のフィルターとなる。A *cause* B の関係では、B がフレームの名称で A がスロット *cause* のフィルターとなる。

これら4つのフレームは互いに独立しており、知識ベース管理システムが、推論の過程において適宜利用する。

4.2.2 個別的知識ベース

個別的知識ベースは、*bib_frame*、*gsi_frame*、*csi_frame* から構成されており、これらは CASearch のデータから知識ベース管理システムによって自動的に作成される。*bib_frame* と *csi_frame* は、ラベルの名称をそのままスロットにしたのでフレームへの変換は容易であった。*gsi_frame* は、自然言語処理の格フレームに類似した手法を用いた変換処理を行って名詞句を復元し、セマンティックスを明示的に表現した。

以下に、*gsi_frame* の変換処理について述べる。

1) 句の復元処理^[9]

CASearch の一般事項索引では、GCH (General concept heading) を索引見出し語とし、GTM (General text modification) などの項目と合わせて論文の内容を表現している。GTM は索引見出し語の説明であるため、カンマで区切られたフレーズから構成された独特の形式をとっている。これらフレーズの順序を入れ換えて、GCH を適切な位置に挿入すれば元の名詞句を復元できる。

GCH は、GTM の最初のカンマで区切られたフレーズと結合する。結合の手順は、前置詞が右にあれば GCH を前置詞の右に、前置詞が左にあれば GCH を前置詞の左に挿入すれば良い。2番目以降のフレーズとの結合の手順も同様に前置詞の位置で決まる。

2) スロットへの割り当て処理

復元した名詞句を格文法に類似した手法を用いて解析し、*gsi_frame* のスロットに割り当てる^[9]。

自然言語処理では格の判定法として、

a) 格特有の前置詞

b) 意味マーカの利用

c) 動詞によって接続する格が異なる事の利用があるが、本研究では、a) と b) の方法を用いる。動詞のない句を扱うので c) は適用できない。

Table 1 にスロットとその目印の前置詞、意味マーカを示した。意味マーカは、単語と一対一に対応して辞書に記述されている。

スロットへの割り当て方は、以下のように行う。

まず、前置詞から次の前置詞までを一つの句として区切る。by, with, for, after, before, at の句は、前置詞だけで判断できるので、そのまま句をスロットに割り当てる。in, on, from の句は、前置詞だけではセマンティックスがあいまいなので、意味マーカによって判断する。句の中に複数の単語がある場合は、もっとも優先順位の高い意味マーカによって句をスロットに割り当てる。

Table 1. 前置詞、意味マーカ、格の対応

slot	preposition	marker
subject	-	-
instrument	by, with	-
purpose	for	-
time	after, before	-
place	at, in, on	point
source	from, in, on	ma
relation	in, on	rel
micro_range	in, on	job
macro_range	in, on	con

4.2.3 一般的知識ベース

一般的知識ベースは、関係の名称をフレームのスロットとして設定し、概念間の関係や関連知識を記述している。

一般的知識は、あらかじめ人間が部分的に抽出し、その後、知識ベースマネージメントシステムが問題解決を通して動的に抽出する。

4.3 知識ベースマネージメントシステム

4.3.1 概要

知識ベースマネージメントシステムは、ユーザに次の3つの機能を提供する。

- 1) 書誌事項の検索
- 2) セマンティックスを含んだ検索
- 3) 問題解決のための質問

知識ベースマネージメントシステムは、個別的知識知識ベースと一般的知識ベースを利用し、推論機構、知識抽出機構、説明機構によって、1) ~ 3) の機能を実現する。

知識ベースマネージメントシステムは、ユーザの質問を解析し、知識を個別的知識ベースから探索する。求める知識が個別的知識にあればよいが、ないときは一般的知識を利用することによって推論を行い、問題解決を行う。

また、推論の過程において知識ベースマネージメントシステムが知識を動的に抽出したなら、知識を新たな一般的知識として知識ベースに書き込み、次の推論に役立てる。その際、個別的知識ベースを探索することによって、抽出された知識がどのくらい信頼性のある知識であるかが検証できる。

さらに、知識ベースマネージメントシステムは、ユーザの理解を深めるため、説明機構により推論の根拠を説明することができる。

4.3.2 書誌事項の検索

bib_frame は書誌事項をフレーム形式で表現しており、知識ベースマネージメントシステムは、bib_frame から書誌事項を検索できる。これは、従来の検索と変わりはない。

4.3.3 セマンティックスを含んだ検索

gsi_frame は、一般事項索引をフレーム化することにより、CASearch では表現されていなかったセマンティックスを表現している。そのため、セマンティックスを含んだ検索を gsi_frame から行うことができる。すなわち、データベースの検索では、キーワードを入力し、AND, OR 等の論理演算を行い、人間がセマンティックスを考慮して適切な情報かどうかを判断していたが、知識ベースの検索では、格とキーワードの両方を同時に入力することによって、システムがセマンティックスを考慮して適切な情報を判断する。結果として、知識ベースの検索では、データベースの検索に比べてノイズを減少させることができる。

例えば、除草剤 "Roundup" は、イネ科 (Graminate) の植物である "Elymus repens" に対して有効であるかという質問を行う場合、"Elymus repens" を subject として、"Roundup" を instrument として入力すればよい。

4.3.4 問題解決のための質問

本システムでは、ユーザの質問に対して個別的知識を含んでいる gsi_frame を探索し、求める情報がなければ、一般的知識を含んでいる con_frame を利用して推論を行う。

推論の過程で動的に抽出された知識は con_frame に書き込まれ、推論に役立てることができる。

以下に、例として3つの質問を設定し、知識ベースマネージメントシステムの動作を説明する。

[質問1: 除草剤 "Roundup" は、イネ科の植物である "Elymus sibiricus" に有効か]

1) 質問文の入力

質問は、英文に近い形で行う。質問文は、
"what is [Elymus sibiricus] by [Roundup] ?"
とする。

2) 質問文の解析

質問文は、"what" と "by" は意味マーカ、"Elymus sibiricus" と "Roundup" はキーワードであると解析され、意味マーカによってキーワードをどの格で探索するかを決定する。

3) 個別的知識の探索

"Elymus sibiricus" を subject として、"Roundup" を instrument として、また、格を変えて、"Elymus sibiricus" を source として、"Roundup" を subject として gsi_frame を探索する。そして、この事実があればその内容を出力する。

4) 一般的知識の利用

gsi_frame に、該当する事実がなかったら、con_frame に補助となる関連知識がないかを調べる。con_frame には、"Elymus sibiricus" の上位概念として "Graminate" があり、その下位概念として "Elymus sibiricus" の他、"Elymus repens", "Agropyron repens" という階層関係がある (Fig.3. 参照)。

5) 個別的知識の再探索

"Elymus repens", "Agropyron repens" を subject として再び gsi_frame を探索する。この事実が gsi_frame にあればその内容を出力する。

"Elymus sibiricus", "Elymus repens", "Agropyron repens" がなければ、より広い概念である "Graminate" を subject として gsi_frame を探索する。

なお、実際にシステムにこの質問を行うと、"Roundup" は "Elymus repens" と "Agropyron repens" に有効であると答える。

[質問2: 温室効果を引き起こす原因は何か]

1) 質問文の入力

質問文は、
"what causes [greenhouse effects] ?"
とする。

2) 質問文の解析

質問文は、"what" と "causes" は意味マーカ、"greenhouse effects" はキーワードであると解析される。

3) 個別的知識の探索

gsi_frame を探索し、"fluorochlorohydrocarbons", "chlorofluoro hydrocarbons", "hydrocarbons", "carbon dioxide" の4つの化学物質が温室効果の原因であると答える。

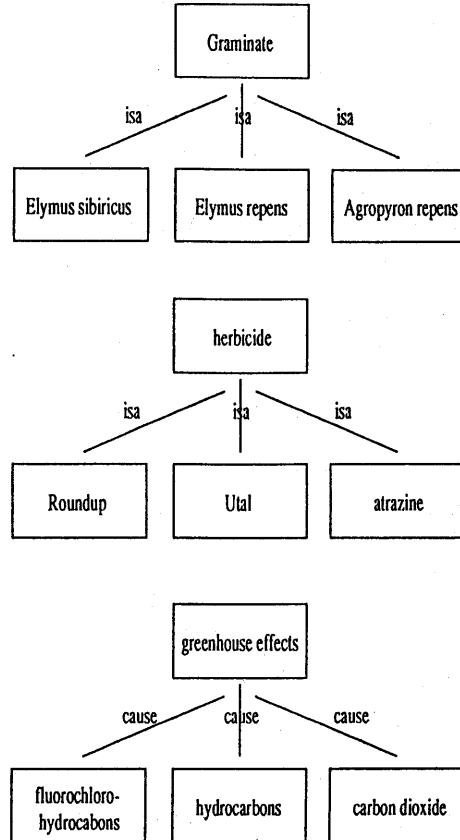


Fig.3. 一般的知識の階層関係

4) 知識の動的抽出

この質問によって、上の4つの化学物質が温室効果の原因であることが判明する。そこで、知識ベースマネジメントシステムは、この新たな知識を con_frame に書き込む。

ただし、新たに抽出された知識すべてを自動的に書き込むわけではない。一般的知識を推論の過程において有効に活用するためには、妥当な知識だけを書き込む必要がある。そのため、動的に抽出された知識を書き込む場合は、どの知識を書き込むかをユーザに問いあわせることによって決定する。それと同時に、一般的知識の重複検査を自動的に行う。

[質問3: 化学物質 carbon dioxide によって何が起るか]

1) 質問文の入力

質問文は、
"what happens by [carbon, dioxide] ?"
とする。

2) 質問文の解析

質問文は、"what" と "happens" と "by" は意味マーカ、"carbon dioxide" はキーワードである

と解析される。

3) 個別的知識の参照

gsi_frame を探索すると、"greenhouse effects", "climatic effects of air pollution", "environmental and societal impacts of greenhouse effect", "global climate changes and temp. increase"などが "carbon dioxide" によって起こると答える。

4) 知識の動的抽出

質問2と同様に、ユーザに問いあわせながら con_frame に知識を書き込む。

なお、システムの動作例は、Fig.4. に示した。

4.4 説明機構

知識ベースマネージメントシステムは、個別的知識と一般的知識を利用して問題解決を行う。しかし、ユーザは、問題解決の答を示されただけでは、答に満足しなかったり、答を理解できないことがある。その場合に知識ベースマネージメントシステムは、どのように問題解決を行ったかを示すことができる。そのことによって、ユーザは、答を理解しやすくなる。

説明機構は、質問の前に talk モードにすることによって起動し、shutup モードにするまで有効である。

talk モードにして質問1を行うと、システムは、Elymus sibiricus と Roundup に関連した一般的知識を簡単な図で表示した後、問題解決の答を表示する。このように、一般的知識を簡単な図で表示することによって、どのように一般的知識が利用されたかをユーザに示す。

talk モードにして質問2と質問3を行うと、問題解決の答を表示した後、動的に抽出した知識を con_frame に書き込み、書き込んだ一般的知識を簡単な図で表示する。このように、書き込んだ動的知識を簡単な図で表示することによって、書き込んだ知識が一般的知識として妥当かどうかをユーザは確認できる。

一般的知識の図の表示を、問題解決の過程においてだけではなく、独立に行うこともできる。

例えば、温室効果についての一般的知識を表示させたいときは、

"show me knowledge about [greenhouse effects]."

と入力すれば、温室効果についての一般的知識が簡単な図で表示される。

Input the query by list as follows.

what is [KEYWORDS LIST] by [KEYWORDS LIST] ?

-> what happens by [carbon dioxide] ?

water net basin supply and hydroelec. generation potential happen by carbon dioxide.

water resources happen by carbon dioxide.

modeling of water cycle and balance effect of concn. doubling happen by carbon dioxide.

(中略)

greenhouse effect happen by carbon dioxide.

greenhouse effect happen by carbon dioxide.

seasonal cycle annual increase happen by carbon dioxide.

seasonal cycle annual increase happen by carbon dioxide. Input

the query by list as follows.

what is [KEYWORDS LIST] by [KEYWORDS LIST] ?

-> show me knowledge about [greenhouse effects]

[greenhouse, effects]

| cause

—— [fluorochlorohydrocarbons, as, refrigerating, agents]

| cause

—— [hydrocarbons, as, refrigerating, agents]

| cause

—— [fluorochlorohydrocarbons]

| cause

—— [chlorofluoro, hydrocarbons]

| cause

—— [carbon, dioxide]

[carbon, dioxide]

| cause

——> [greenhouse, effects]

| cause

——> [environmental, and, societal, effect]

| cause

——> [climatic, effects, of, air, pollution]

Input the query by list as follows.

what is [KEYWORDS LIST] by [KEYWORDS LIST] ?

-> halt.

Input the kind of contents -> halt.

Input the command of top level -> halt.

See you again.

Fig.4. システムの動作例

5 考察

本研究では、文献情報データベースから知識ベースに変換し、個別的知識と一般的知識という2種類の知識を利用して問題解決を行うシステムを構築した。

ここでは、知識ベースシステムによる問題解決からみた知識について考察する。

1) 知識の獲得

問題解決を行うには、まず、対象とする領域の知識を獲得しなければならない。そのためには、対象とする領域の専門家へのインタビューや専門書からの抽出という方法がある。しかし、その方法では、知識を多量に獲得することができず、知識を構成するのに非常に労力を必要とするので、知識ベースシステムによる問題解決を行う知識としては、不十分である。

そこで、本研究では、文献情報データベースを情報源にすることにより、問題解決のための知識を多量にかつ効率的に得ることが可能になった。このことより、知識獲得の問題に対して一つの解決法を示すことができた。

2) 知識の構成

知識ベースシステムによる問題解決にあたっては、問題領域を正確に把握し、問題解決に役立つ知識を抽出し、個々のシステムそれぞれに適切な知識の構成を行わなければならない。

これまで知識の構成に関しては、問題解決のために開発された手法である ISM 法を理論とアルゴリズムの両面から発展させた知識構造モデリング法^[12]、概念階層上で表現されている知識を一階述語論理に変換し、サーカムスクリプションを実行したもの^[13]、構造化ネットワーク Rete を利用し、それを部分的に修正したもの^[14] などがある。

本論文は、文献情報データベースから知識ベースを構築している。データベースから抽出した個別的知識は、個々の事実に関するものである。しかし、個々の事実だけでは、問題解決のための知識としては不十分である。問題解決の答が直接個別的知識にないときは、それ以外の一般的知識が必要である。

本研究の知識の構成は、セマティックスを明確にした事実としての個別的知識、個別的知識に関連する概念の関係を記述した一般的知識からなるものとした。この構成は、文献情報データベースから知識ベースを構築し、問題解決を行うということからみて自然で妥当な知識の構成だと考えられる。

3) 知識の利用

問題解決を行う知識ベースシステムを構築する

場合、問題の領域に適切な知識を構成し利用しなければならない。

本研究では、個別的知識と一般的知識を利用して問題解決を行う。個別的知識には、個々の事実が記述してあり、問題解決手順としては、最初に個別的知識を探索し、個別的知識に答がなければ一般的知識を利用する。この方法により、問題解決を有効に行うことができた。

さらに、問題解決を通して抽出された知識は、一般的知識に書き込まれ、問題解決のための新たな知識として利用される。知識の抽出の方法として、(i) 人間のアドバイスによる抽出、(ii) 例による抽出、(iii) 問題解決による抽出の3つの方法があるが^[15]、本研究では、(i) と (iii) の方法を併用している。これによって、問題解決を繰り返すことにより、一般的知識が増加し、問題解決がより効果的に行われるようになる。また、人間が一般的知識をチェックすることにより、一般的知識の矛盾をある程度解消することができる。

6 今後の課題

個別的知識と一般的知識を利用して問題解決を行うことにより、知識の獲得と知識の利用について新しい手法を示した。しかし、知識の管理、すなわち、一般的知識が多量になった場合の知識の矛盾の解消についての問題が残っている。知識の管理については、ATMS^[16]などのシステムがあり、それに関する研究も数多くある。今後、知識の管理についても考えていきたい。

そのほか、入力文の改善、説明機構の充実など知識ベースシステムとユーザとのインタフェースも改善したい。

7 まとめ

今回、本論文は、文献情報データベースを情報源にすることによって、知識を自動的に獲得できることを示した。また、セマンティックスが暗示的であるデータベースのデータを明示的な知識として知識ベースシステムを構築することにより、問題解決のための知識として利用できることを示した。

8 参考文献

- [1] Boose, J.H.: Personal Construct Theory and the Transfer of Human Experience, Proc. of AAAI'84, pp. 27-33 (1984).
- [2] Kahn, G., Nowlan, S. and McDermott, J.: Strategies for Knowledge Acquisition, IEEE Trans.

PA and MI, Vol. PAMI-7, No. 5, pp. 511-522 (1985).

[3] 滝 寛和, 椿 和弘: 知識獲得のための知識表現「専門家モデル」, 人工知能学会誌, Vol.5, No.2, pp. 59-68 (1990).

[4] 小口 琢夫, 近藤 秀文: 知識ベース管理システム 知識ベースとデータベースの融合方式, 情報処理学会第31回全国大会, pp. 1003-1004 (1985).

[5] 森原 一郎, 牛島浩一, 小野寺 尚文: KBMS における DB/KB 変換方式, 情報処理学会第38回全国大会, pp. 571-572 (1989).

[6] 柳 吉洙, 志村正道: 故障診断用エキスパートシステムにおける 知識獲得, 人工知能学会誌, Vol.1, No.1, pp.93-100 (1986).

[7] 志村正道: 知識工学概説, 情報処理, Vol.26, No.12, pp.1457-1462 (1985).

[8] 諏訪 基: 知識工学の動向, 情報処理, Vol.26, No.12, pp.1463-1467 (1985).

[9] 石塚 英弘, 王 忠清, 山本 毅雄: 化学文献情報データベースの知識ベース化 - 設計と試作, 1989 年情報学シンポジウム講演論文集, pp.53-61 (1989).

[10] 近藤 秀文, 小口 琢夫: 知識ベース管理システムの基本構想, 情報処理学会第31回全国大会, pp. 1383-1384 (1985).

[11] 服部 文夫, 清水 信昭 ほか: 知識ベース管理システム, 情報処理学会 知識工学と人工知能研究会, 41-6 pp.41-48 (1985).

[12] 大内 東, 栗原 正仁, 加地 郁夫: 知識構造モデリング法の構成と 具象化ツール, 人工知能学会誌, Vol.3, No.5, pp.599-606 (1988).

[13] 中川裕志, 萱島 信, 森 辰則: 自然な構造の階層的知識, 人工知能学会誌, Vol.3, No.3, pp.329-336 (1988).

[14] 荒屋 真二: 知識の構造化に関する考察, 情報処理学会 知識工学と人工知能研究会, 50-7, pp.1-8 (1987).

[15] 渡辺正信: エキスパートシステムにおける知識獲得, 情報処理, Vol.28, No.2, pp.167-176 (1987).

[16] de Kleer, J.: An Assumption-Based Truth Maintenance System, Artificial Intelligence, Vol.28, pp.127-162 (1986).

9 参考文献

[1] 辻井潤一: 知識の表現と利用. 昭晃堂,(1987).

[2] 上野晴樹, 石塚 満: 知識の表現と利用, オーム社, (1987).

[3] 中島秀之: 知識表現とProlog/KR. 産業図書, (1985).