情報システム　34－3
（１９９１．５．２１）

人工知能技術を応用した
データベース利用技術に関する調査

鍋田　茂子　　川崎　浄正
㈱ＣＳＫ

経済統計データベースの検索には、経済統計に関する専門知識と、データベースシステム自体に関する知識が必要であり、さらに過去の利用例もしばしば参照される。本調査研究では、特に貿易統計に範囲を限定して、一般的なユーザの検索要求を調査し、適切なデータ入手を導くために必要な知識を分析し、エキスパートシステム的なアプローチで試作システムを開発した。

対象とした経済統計分野は、多くの情報源からのデータを取り扱う為、適切なデータが存在しない場合に、次善のデータを探索する機能やユーザの検索経験に応じて知識を更新する機能が求められる点が特徴となっている。試作システムを評価し、実用的な検索支援システム構築のための課題をまとめた。

Using Artificial Intelligence to Retrieve Financial Statistical Databases

Shigeko Nabeta　　Jyoushou Kawasaki
CSK Corp.

We are surrounded by a large amount of information. But it is not so simple to get suitable financial statistics.　How can we make it easy to get information we need? To solve the problem, recent AI techniques are considered useful. In this paper, we applied AI techniqes to the domain of information retrieval for financial statistics, in which the expertise on financial statistics, databases, and IR commands is required, and the users need near optimal solutions even when the databeses do not have exact solutions.

1．はじめに

近年における情報化の進展とともに、データベースの構築が急速に進められている。

対象となる分野は拡大し、利用される局面も多様化しているが、現実に利用者が、データベースにアクセスし、的確にデータを利用するためには、各データベースの所在、構造、用法等のデータベースそれ自体に関する知識と、データそのもの及び利用上の制約事項、ノウハウ等のデータ利用に関する知識が不可欠である。しかも、これら利用者に求められる知識は、益々複雑、多様化しており、データベースの利用普及の隘路の一つとなりつつあるのが現状である。

このため、本調査研究では、データベース利用の効率化を目的として、昭和６３年度より、３ヵ年の計画で、データベースと利用者の間のインタフェース向上の試みを行った。

[JIPDEC 1989] [JIPDEC 1990] [JIPDEC 1991]

2．経済統計データベース検索の現状

検索の現場を見ると、官庁統計や、業界統計、また、国連やＩＭＦ等の国際機関の作成した統計などが、電子的に検索可能になっている。

データベース利用の現状を整理すると以下のようになり、データベース利用の様相は、一般化と高度化の２つの方向に向かっている。

①データベース及び経済統計に関して、経済統計の専門家以外の人が利用するケースが増えている。（特に貿易統計データベースについては、国際化の進展により従来関心を持たなかった層の利用が増えた）

②従来の利用者（経済統計の専門家）のデータ利用ニーズが多様化している。（データのグラフ化、外部情報と内部情報の統合加工等の要求が増えた）

このような状況下で、ユーザが感じる問題点を調査すると、特に経済統計データベース検索の特徴として、次のような問題点が抽出された。

①データベース及び経済統計に関する専門知識のガイダンスがないと、検索が難しく理解しにくい。

②既存のガイダンス機能が使いにくい（メニュー構成上の問題等による）

③品目分類について、データベースの分類と利用者の必要とする分類にギャップがあり、検索がしにくい。

④データの性質（例えば、調査対象範囲の定義により、同じ名前のデータでも性質が異なる）に関する情報が、利用者に判りにくい。

⑤欠測データに関する情報が判りにくい。

⑥異なるデータベースからの情報の統合等高度な加工がしにくい。

3．データベース検索モデル

２で述べた問題点に対して、実際はどのような問題解決が行われているかを明らかにするために、専門家でない人の発する検索要求の姿を調査し、これに対して、専門家である助言者がどのようにデータ入手を導くかを調査し、データベース検索のプロセスをモデル化した。

3．1　ユーザの検索要求

経済統計の範囲を特に貿易統計に絞って実験を行うこととし、「通産ジャーナル」の記事より貿易統

計に関する１００数件の検索要求文を作成した。これらは、必ずしもデータベースの検索と直接結びついている訳ではないが、データベースを思い浮かべる前に、ユーザが欲しいと思う情報の原型と考えられる。[JIPDEC 1989]

統計やデータベースに関し深い知識のないユーザの、自然な表現（検索コマンド的な表現ではなく、概念的な表現や抽象的な表現も許した表現）を考えたいので、ユーザが要求する「検索」は、この１００数件によって代表されるものと想定し、これをシステムによる支援の対象とした。

以下に、このユーザが要求する「検索」の例を示す。

ユーザの要求サンプル

- 最近の日本、韓国、台湾等のアジア勢の対欧輸出構成について知りたい。
- 輸出関連の製造業の状況を見たい。
- アジアＮＩＥＳからの輸入状況を見たい。
- 我が国の国民総生産に対する製品輸入の割合を知りたい。
- 我が国の開発途上国からの製品輸入の割合を知りたい。
- 米国から日本へのハイファイのステレオスピーカーの輸出について知りたい。
- 輸入報告額と前年同月比が知りたい。
- 米国の貿易赤字と日本の貿易黒字に占める為替レートの割合を知りたい。
- 我が国の米国スーパーコンピュータ関連部品の輸出量を見たい。

上記調査を経てユーザの検索要求を、①報告国、②相手国、③品目、④事項という４つの属性によって表現した。例えば報告国（日本）、相手国（アメリカ）、品目（コンピュータ）、事項（輸出額）等である。

これ以外に、加工、期間、期種、等の属性があるが、これらは、データベース選択に関しては補助的な属性なので、データベースを選択するための属性として上記①から④を使用した。

また、検索要求によっては、全ての属性に意味のある値が入るとは限らないことが、この問題の特徴になっている。

例えば、「日本からアメリカへの輸出額」の場合、品目に相当するものは、’全品目’　となり、全域を示す用語になる。

また「日本の海外資産」の場合、相手国、品目は、それ自体意味を持たなくなる。

こららの属性は貿易統計に関する検索要求を表現するためのものであり、他の経済統計に関しては、別のフレームが存在するが、事項として取りうる値を拡張し、他の属性に関して、無意味であるという値の取り方を許すと、国単位の統計に関しては、比較的広範囲な検索要求を吸収できる。

表1　貿易統計に関する検索要求

属性名	意味	取りうる属性値
報告国	貿易統計の主体となる国名（どの国からみた統計かを示す）	日本、アメリカ合衆国等の国名及び、ASEAN 諸国等の国群と、これら俗称のうちよく使われるもの（米国東独、 NIES 等）
相手国	報告国に対して、貿易の相手となる国名	報告国と同じ
事項	知りたい情報の種類	輸出額、ＧＮＰ、為替レート等
品目	貿易でやりとりされる物の名称	石油、コンピュータ、電気製品等基本的に概況分類に従って分類された品目とこれらの俗称

3．2　検索要求から必要なデータを探索するための過程

3．1　の検索要求に対して実際に、専門家がどのようなアドバイスを行い、検索の支援を行っているかを、ヒアリング及び、通商白書の検索事例から調査し、検索支援に必要な知識を下図のように整理した。

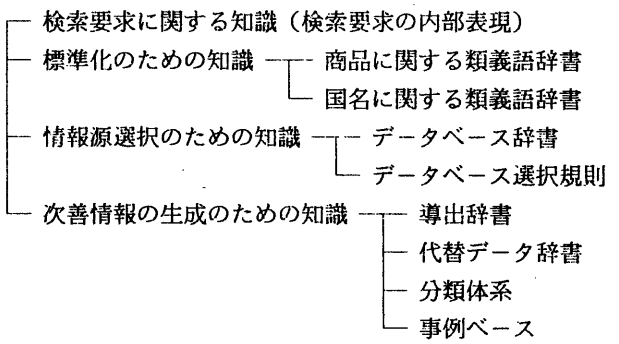
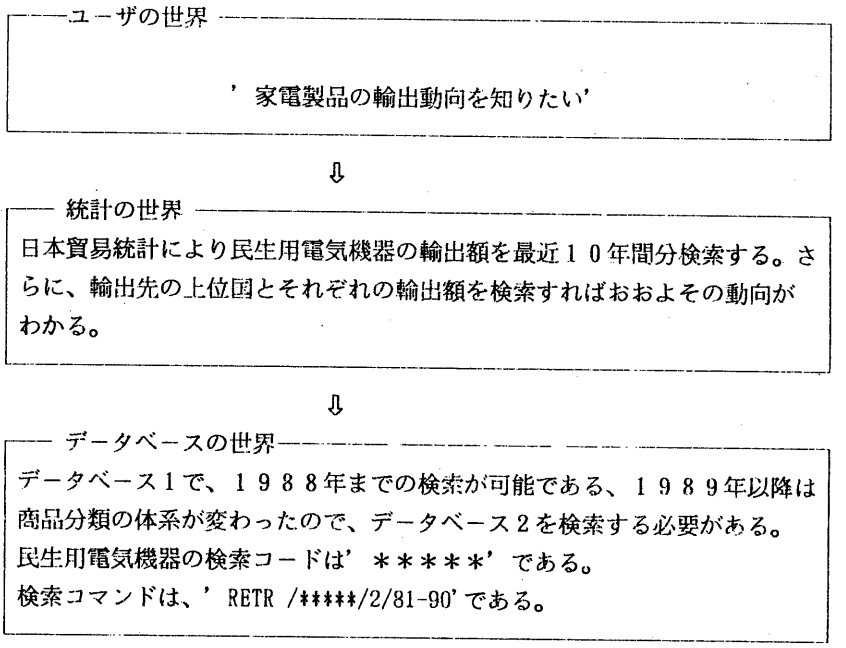


図3　検索に必要な知識と対応する辞書

これらの知識を使ったデータベース検索の問題は次に示す3つの世界のうえで考えると判りやすく整理できる。[JIPDEC 1989]



上図の「ユーザの世界」は、ある情報を必要としているユーザが（データベースや、統計に関する専門知識なしに）表現した、欲しい情報の姿である。日常的な言葉で表現されておりデータベースや、経済統計の専門家から見ると曖昧な表現であったり、複数の解釈が可能であったりするものも含まれる。「統計の世界」は、統計の観点から情報を分類した世界で、統計の専門家が表現した情報の姿である。～という統計の～という分類に計上されたデータを～の期間に関して合計したもの・・というように、どの統計にどういう形で収録されている情報であるかによって情報を表現する。

「データベースの世界」は、データベースの観点から情報を分類した世界で、データベースの専門家が表現した情報の姿である。どのデータベースに収録されており、どのようなコマンドによって検索ができるという形で情報を表現する。

専門家でない人が情報を検索することは、「ユーザの世界」から「統計の世界」を通して「データベースの世界」へ、検索要求を変換することに他ならない。

「ユーザの世界」から「統計の世界」へ情報を変換するためには、世の中に存在する統計の性質や特徴に関する知識を持ち、一般常識の範囲で表現された要求を解釈して統計に当てはめることが必要である。また’統計の使われ方’に関して、この要求に対しては、この統計が利用されることが多い・・・といった慣例的な知識も必要である。

さらに、類似情報に関する知識を使って、検索に失敗した場合は別の情報源を利用するなどの検索戦略が存在すると思われる。

「統計の世界」から「データベースの世界」へ情報を変換するためには、データベースのコマンド体系に関する知識を持ち、統計がデータベースにどのような条件で収録されているかを知る必要がある。また統計上は存在する筈だが欠測である・・・というようなデータを検索してみて始めて判るような知識も存在する。

このように、情報検索の様相を整理して考えると、解決すべき問題とは、「ユーザの世界」から「データベースの世界」への検索要求の変換に必要な知識を分類し、システム的に変換を行う仕組みを作る

事であろうと思われる。

またこの知識としては、辞書的な明文化されたものの他に、ノウハウとして、因果関係的に表現できるものや、因果関係ではなく過去の検索事例そのもののような形をしたものが存在することが伺える。

4．プロトタイプシステム

4．1　システム構成

システムは、パソコン上に、Prologを使って開発した。さらに、ホストコンピュータ上のデータベースをアクセスするためのフロントエンドプロセッサとして、パソコン上で稼働する通信ソフトを使っている。（現在アクセス可能なデータベースは3種類である）

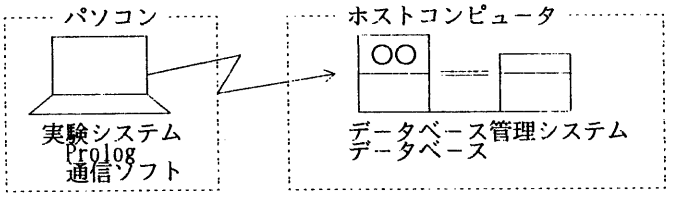


図2　システムの稼働環境

4．2　システムのソフトウェア構成

システムは、①制御機構、②知識ベース群、③事例ベースの3つの部分に分かれる　制御機構は、知識や事例を利用してユーザが入力した検索要求から検索コマンドを生成し、検索を行い必要に応じて、知識ベースの更新を行う。制御機構は、さらに、①検索要求入力部、②事例ベース推論部、③データベース選択部、④コマンド生成部、⑤検索部、⑥知識登録部に分かれる。

知識ベース群は、検索に必要な知識を、何種類かの統一した形式に整理したもので、①類義語辞書、②導出辞書、③代替データ辞書、④データベース選択規則、⑤データベース辞書の5種類からなっている。そのなかの一部は、実験システムの学習機能により、処理終了後に更新が行われる。

事例ベースは、検索支援の履歴を統一した形式で、蓄積するもので、検索要求の種類を区別するためにインデクスを付与されている点が特徴である。

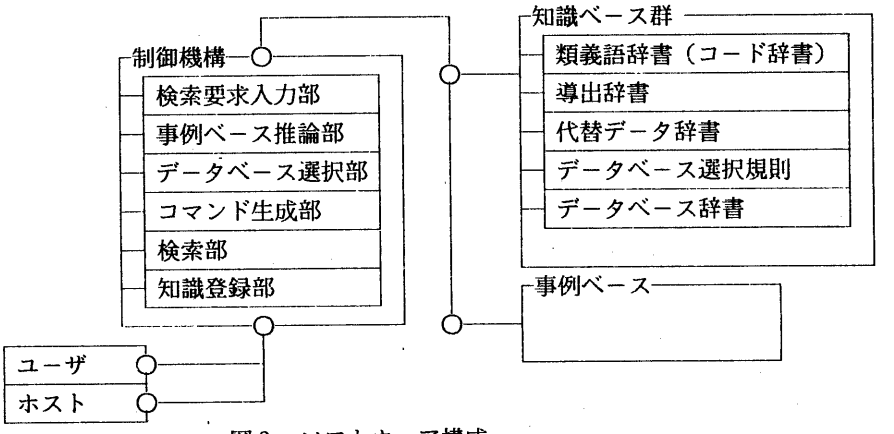


図3　ソフトウェア構成

4．3　画面構成と基本的な操作方法

画面の構成と各部分の呼称を下図に示す。

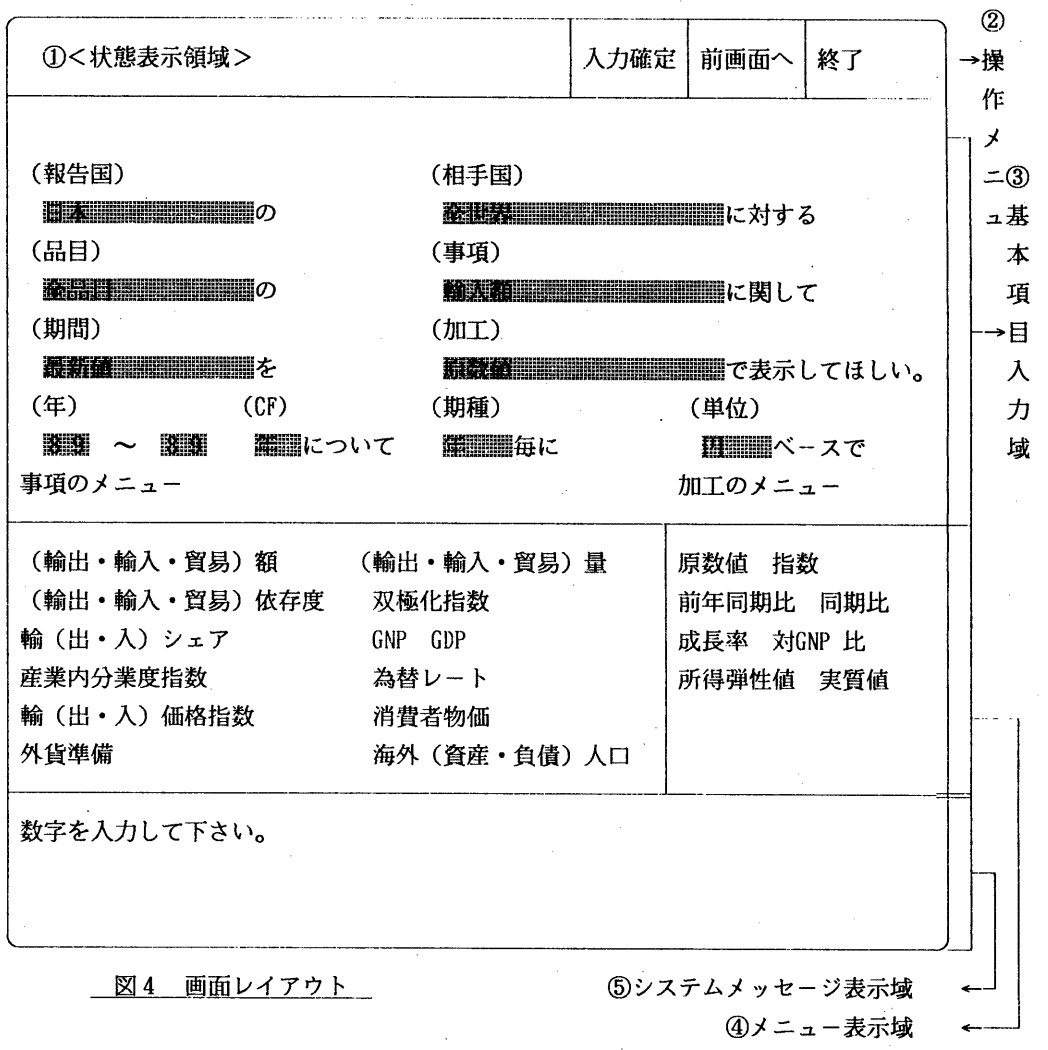


図4 画面レイアウト

入力画面は、図1に示すように、①状態表示領域、②操作メニュー、③基本項目入力域、④メニュー表示域、⑤システムメッセージ表示域の5つの部分から構成される。

状態表示領域はシステムの処理状況をモニタリングするための領域である。操作メニューは各画面に共通する操作指示のためのメニューで、ファンクションキーと対応している。基本項目入力域は、検索要求の各属性を入力するテンプレート画面である。メニュー表示域は事項および加工に関する選択可能なメニューを表示する領域である。システムメッセージ表示域は、システムからユーザへの操作指示や、次善情報に関する説明を表示する領域である。

ユーザは、システムからの指示によって入力したい枠を指示し、その枠に入力を行う。
入力したい枠にカーソルを移動して、リターンキーを入力することにより、指定された枠が入力可能な状態になる。

4．4 知識ベース構成

システムが学習機能を持つため、終了時に新たに次の辞書が作成される。

①国名の類義語辞書：国名の新規登録を行った場合に作成される
②品目名の類義語辞書：品目の新規登録を行った場合に作成される
③導出辞書：類義語辞書の新規登録をおこなった場合に作成される
④事例ベース：１件以上の処理を終了した場合に作成される
各辞書の記述様式を以下に示す。（辞書は全てprologの辞書になっている）

① 類義語辞書（国名）
国名の類義　語辞書は、国名と関する類義語及び検索コードを収録するもので、ユーザによって新たな類義語が登録された場合に、更新が行われ知識獲得の機能を持つ。
形式
dic(種別、基本名詞、〔類義語のリスト〕、レベル、〔検索コード〕）.
・種別は、国名、品目名の区別を表し、国名の場合は、r-c,品目名の場合はgoods と記述される。
・レベルは、知識獲得の際にユーザに品目分類の階層を示すために使われる。
・検索コードは３種類のデータベースに対応して
国名の場合は、〔ＤＢ１のコード、ＤＢ２のコード、ＤＢ３のコード〕が記述される。（品目レベルの属性を持つデータベースは１種類しかない）

② 類義語辞書（品目）
品目の類義語辞書は、品目と関する類義語及び検索コードを収録するもので、ユーザによって新たな類義語が登録された場合に、更新が行われ知識獲得の機能を持つ。
形式
dic(種別、基本名詞、〔類義語のリスト〕、レベル、〔検索コード〕）.
・種別は、国名、品目名の区別を表し、国名の場合は、r-c,品目名の場合はgoods と記述される。
・レベルは、知識獲得の際にユーザに品目分類の階層を示すために使われる。
・検索コードは３種類のデータベースに対応して
品目名の場合は、〔輸出コード、輸入コード〕が記述される。（品目レベルの属性を持つデータベースは１種類しかない）

③ 導出辞書
導出辞書は、品目及び事項に関する上位と下位の関係を収録する。
形式
red(品目名または事項、〔下位項目のリスト〕、＄メッセージ＄）.
・下位項目のリストは、品目または事項に記述された項目の下位項目である。（即ちこれらを足し合わせると品目名または事項と同等になる）
・メッッセージはこの知識が利用される時にユーザに表示される。

④ 代替データ辞書
代替データ辞書は、事項に関して概念的に近いデータの関係を収録する。
形式
simil(データ名１、データ名２）.
例
simil(ＧＮＰ，ＧＤＰ）.

⑤5　データベース選択辞書

データベース選択辞書は与えられた検索要求からデータベースを選択するための規則を収録する。

形式

db-sel(〔検索要求の属性構成〕、修正の種別、ＤＢ名、’メッセージ’)．

・検索要求の属性構成は、検索要求の　報告国、相手国、品目、事項、期間、加工の範囲を示す。

・修正の種別は、検索要求に対して、当該ＤＢを適用する場合に必要な修正の種類を示す。

ｓａｍｅ：修正不要

ｃｈｌ２（項目）：検索要求の報告国、相手国を入替え、事項を項目で置き換える。

⑥　データベース辞書

データベース辞書は、データベースの収録状況を収録する。

形式

dbd(ＤＢ名、〔属性構成〕、〔事項のリスト〕、〔収録期間〕、〔年・年度区分〕、〔収録機種〕、〔収録単位〕)．

・属性構成は、当該データベースの属性の組。

・事項のリストは、収録されている事項のリスト。

⑦　事例ベース

事例ベースは過去の検索事例を蓄える。〔NABETA 1991〕

形式

case([日本，米国，車，輸出額]，[日本，アメリカ，乗用車，輸出額]，db1，
　　1　　2　　3

[1,1,1,輸出額]，[$RETR 317010503/304/80FY-82FY$,$DISP V3$]，
　　4　　5

[$RETR 3/ 1/ 5$,$DISP V3$])．
　　6

下線部の意味は次のとおりである。

①ユーザの入力した検索要求。（属性毎に表現されたもの）

②検索要求を類義語辞書を使って基本名詞に変換したもの。

基本名詞は、国名、品目等の一般的な呼称で、データベース検索コマンド生成に必要な検索コードが基本名詞との対応表によって与えられている。但し、属性’事項’に関しては、メニュー選択によって入力が行われるので、入力された属性と基本名詞は同じものである。

③データベース名

④属性指標

検索要求を示す属性を次の規則で、１から３までの数字に指標化したもの。

ｉ属性の値が全世界、全品目等の全域を示す用語になっている場合：２

ii日本のＧＮＰ等の検索要求の場合の相手国や品目のように、属性自体に意味の無い場合：３

iii ｉ ii以外の用語が入力されている場合：１

属性’事項’に関しては、データ構造の性質上、必ず指標は１になるので、指標の代わりに、入力された値そのものを指標として使う。

⑤検索コマンド

①の検索要求に対応する検索コマンド。（成功したもの）

⑥一般化検索コマンド

⑤の検索コマンドのうち、報告国コードを示すものを¯1 に、相手国コードを示すものを¯2 に、品目コードを示すものを¯3 に、事項を示すものを¯4 に、期間を示すものを¯5 におきかえたもの。

5．実用化に向けた今後の課題

検索支援システムに対する要件は5種類程度に分類される利用目的に応じて異なる。今回は、調査業務を中心に、試作システムの開発をおこなったが、実際は研究業務、企画業務等の利用目的に応じてインターフェイスの異なる、共通の知識ベースを持ったシステム構成が考えられる。〔ＪＩＰＤＥＣ　１９９１〕

知識構造と知識獲得の面からは、さらに、データの説明書に当たる書誌的な情報の取扱について検討をすすめる必要がある。過去の検索経験を利用する考え方は膨大な知識を効率よく蓄えるために有効であり、運用形式と併せて検討を進めていく必要がある。

検索対象分野の拡張に伴い、複数の系列の比較に関する問題が発生することが予想されるが、この問題に関しては、統計調査の時点からの統計分類の標準化が望まれる。

6．おわりに

3年間の調査研究を通して、データベースの利用の様相を調査し人工知能技術を使って行うことのできる検索支援の形態を実験によって確かめ、実用的な検索支援システム構築のための課題をまとめた。

試作システムは、規模の小さいものであるが、ユーザの立場からの検索支援という点に、本研究の意味があると思われる。テンプレート型のインターフェィス、事例ベース推論による過去の経験の利用等の特徴を生かし、さらに一般的な情報検索モデルを検討していく予定である。

参考文献

[JIPDEC 1989] （財）日本情報処理開発協会情報処理技術の応用に関する調査研修報告書－人工知能技術を応用したデータベース利用技術に関する調査. -63-R003,1989.

[JIPDEC 1990] （財）日本情報処理開発協会情報処理技術の応用に関する調査研修報告書－人工知能技術を応用したデータベース利用技術に関する調査. -01 R007,1990.

[JIPDEC 1991] （財）日本情報処理開発協会情報処理技術の応用に関する調査研修報告書－人工知能技術を応用したデータベース利用技術に関する調査. -02 R002,1991.

[Nabeta 1991] 鍋田茂子、寺野隆雄：事例ベース推論を利用した情報検索ノウハウの蓄積と利用. 情報処理学会研究報告、91-AI-75, pp69-78,1991.