

02—R 002

情報処理技術の応用に関する 調査研究報告書

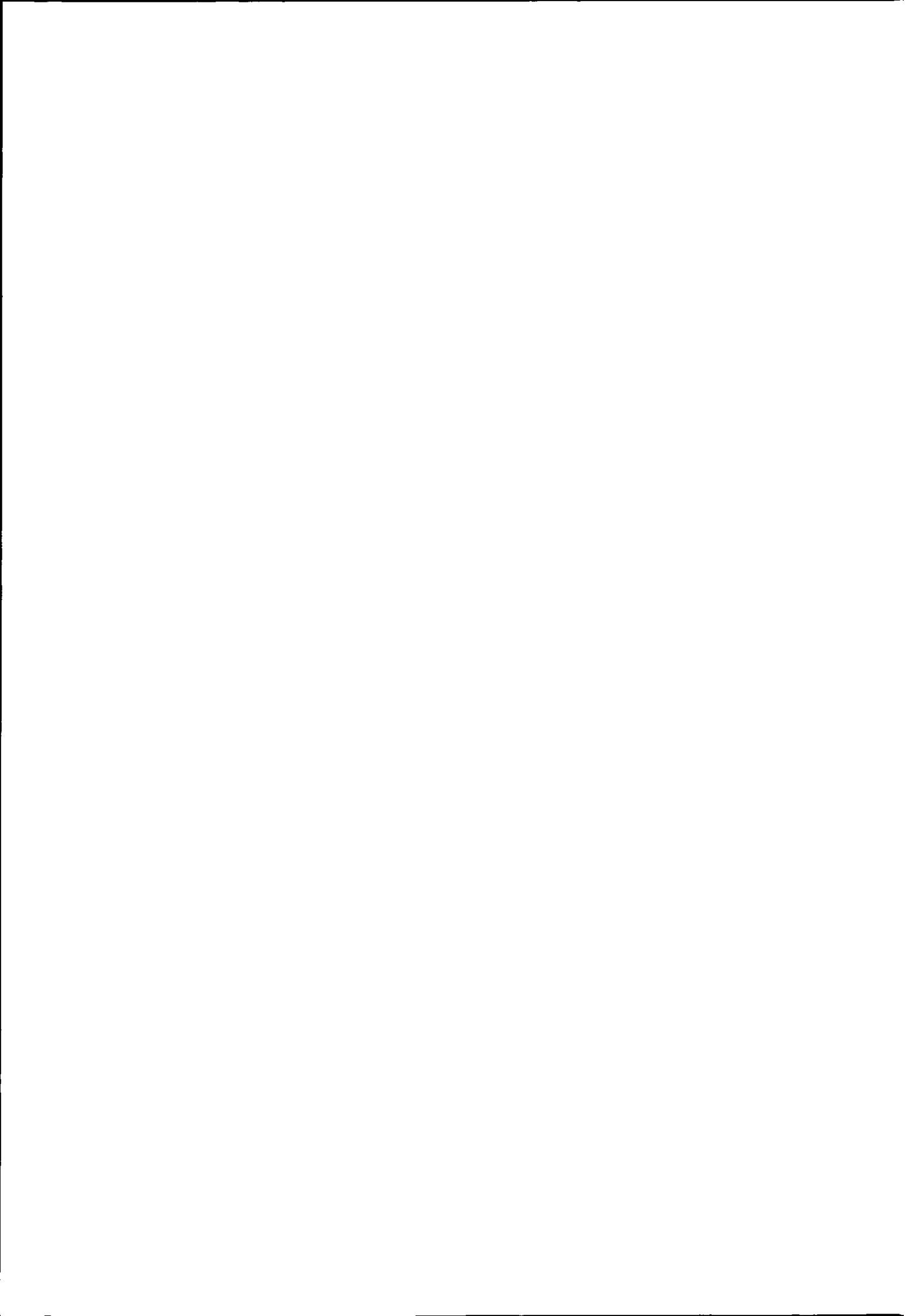
—人工知能技術を応用したデータベース利用技術に関する調査—

平成 3 年 3 月



財団法人 日本情報処理開発協会

この報告書は、日本自転車振興会から競輪
収益の一部である機械工業振興資金の補助を
受けて平成2年度に実施した「情報処理技術
の応用に関する調査研究」の成果をとりまと
めたものであります。





は し 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、従来の個別専門分野から、日常生活レベルに至るまでより多面化してきており、しかも必要に応じて遠隔地からでも瞬時に入手できるようになっている。

しかしながら、現実にご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、収録コード体系、キーワード、検索コマンドの種類、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、分析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とのインタフェース向上の試みをおこなった。

昭和63年度は、実際の経済統計データベース利用の現状調査、人工知能技術の現状調査、海外におけるデータベース検索の技術動向調査を中心に、利用支援のための知識、利用支援システムを実現するための技術に関する検討を行い、実用システムのための目標設定を行うことを活動内容とした。

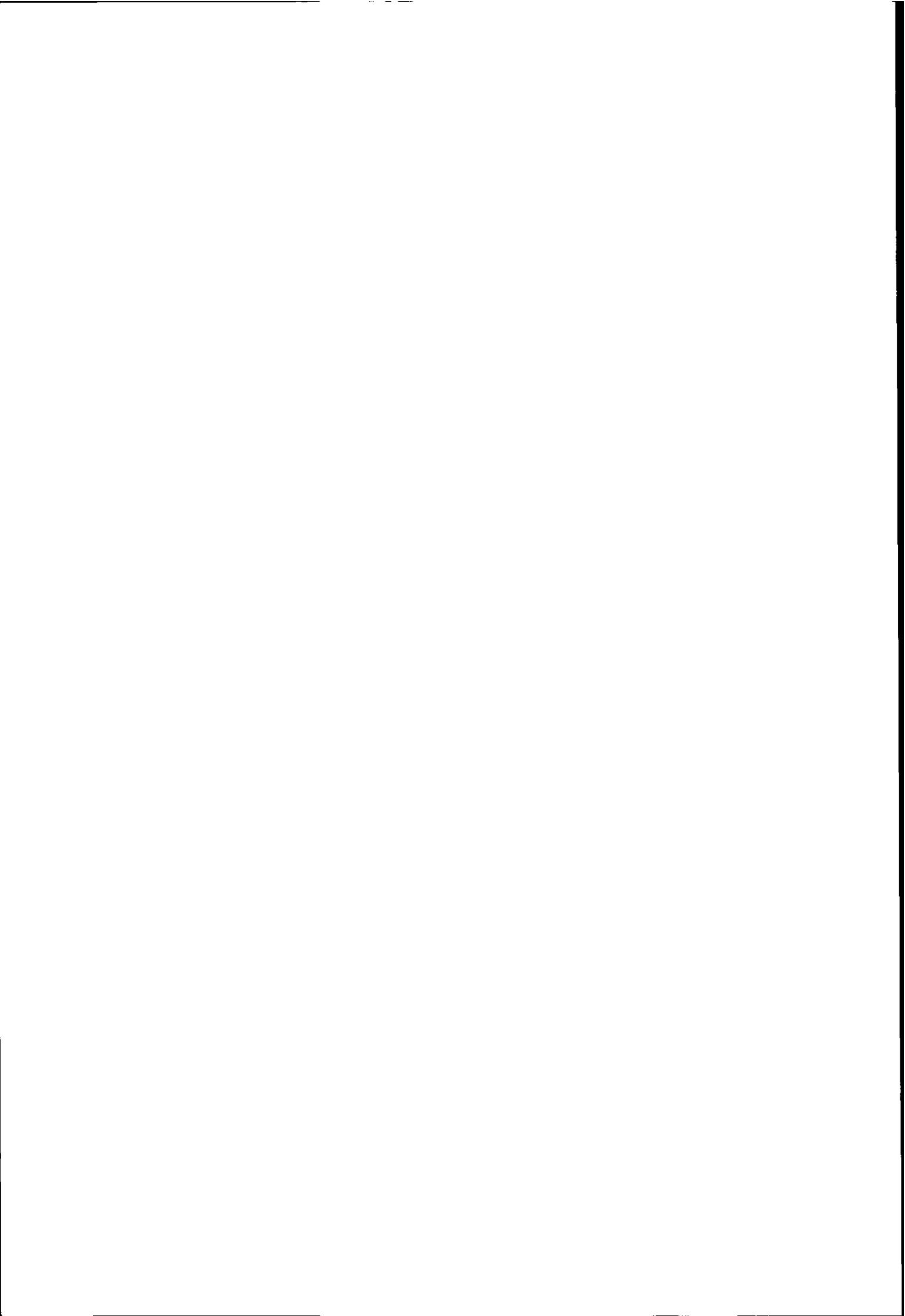
平成元年度は、前年度の調査結果及び知識ベースの設計方針に基づき、貿易統計を例題として、必要となる知識を整理し、知識ベースを試作し、それら知識を利用する試作システムを開発し、知識表現の適否、知識管理方法の妥当性、動作の確認、データベース利用上の有効性等について評価、検討を行い、実用システムへの課題等を整理した。

平成2年度は、前年度までの結果に基づき、知識の追加等を行い、実際のデータベースを利用し、実用システムとしての運用可能性について評価、検討した。またこれらのシステムを実用化するうえでの課題についての検討をおこなった。

本報告書は、この3年間の成果をとりまとめたものである。今後の調査研究のための基礎資料として、利用活用されることはもちろんであるが、関心のある方々に読まれ、データベースを利用した情報活用の一助になれば幸いである。

尚、本調査研究は、株式会社CSK総合研究所に委託して実施した。

平成3年3月



「人工知能技術を応用したデータベース利用技術調査」委員名簿

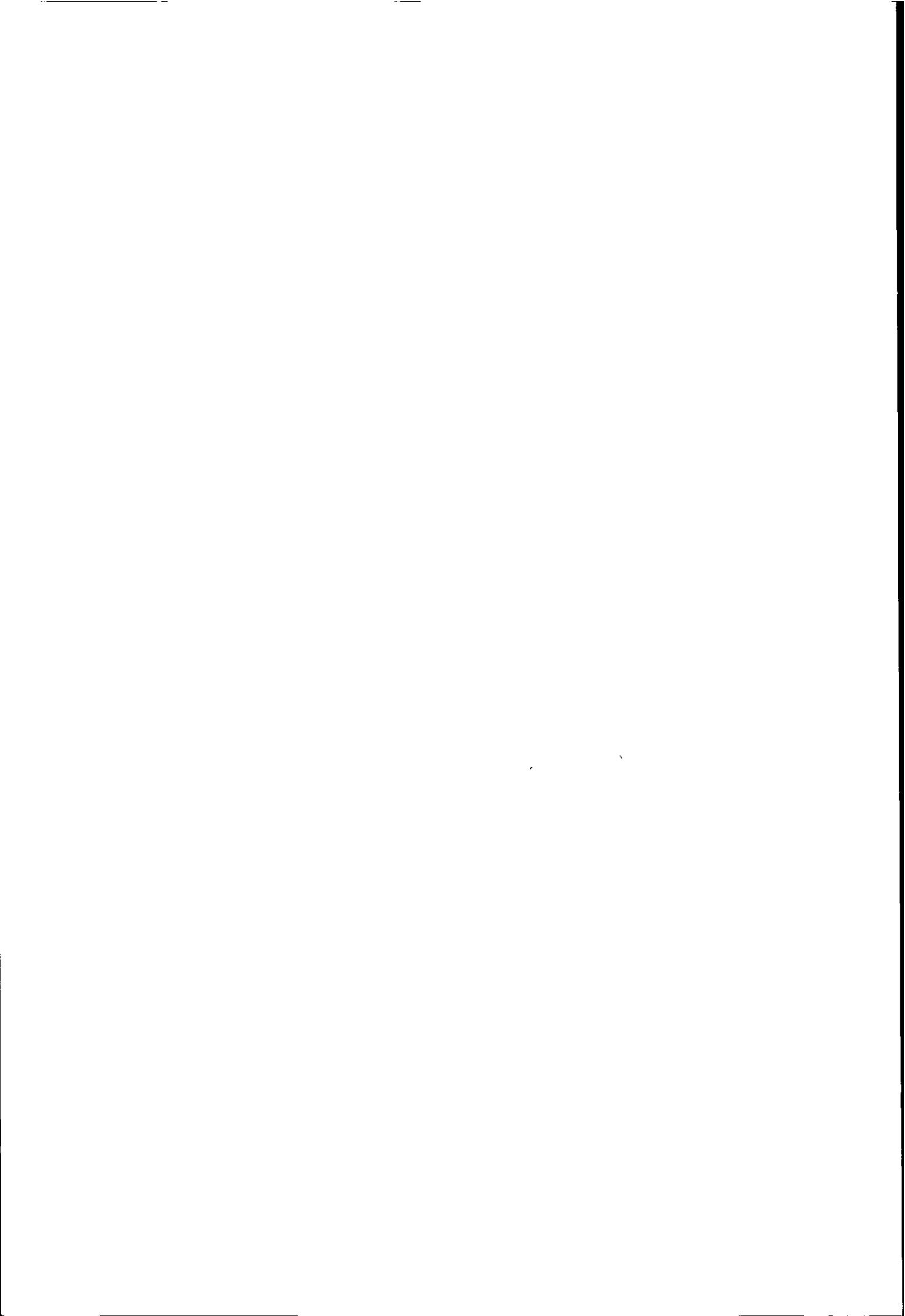
(順不同、敬称略)

(本委員会)

委員長	刈 一博	㈱新世代コンピュータ技術開発機構
委員	鶴野 公郎	慶応義塾大学
委員	石川 徹也	図書館情報大学
委員	黒川 恒雄	工学院大学
委員	佐藤 英人	東京国際大学
委員	藤森 聿子	技術研究組合国際ファジィ工学研究所
委員	斉藤 茂樹	通商産業省 大臣官房 情報管理課
委員	小紫 正樹	通商産業省 大臣官房 情報管理課
委員	仲田 雄作	通商産業省 大臣官房 情報管理課
委員	上原 明	通商産業省 通商産業研究所 政策情報システム部
委員	矢田 光治	㈱CSK総合研究所

(専門委員会)

委員長	鶴野 公郎	慶応義塾大学	(兼本委員会)
委員	石川 徹也	図書館情報大学	(兼本委員会)
委員	長田 博	アジア経済研究所	
委員	宮澤 利成	日本貿易振興会	
委員	野村 玲子	日本貿易振興会	
委員	篠原 昇	㈱アイ・エヌ情報センター	
委員	水島 豊	㈱日本エネルギー経済研究所 エネルギー計量分析センター	
委員	中島 淳元	㈱QUICK	
委員	佐藤 安夫	ログビスタ株式会社	
委員	仲田 雄作	通商産業省 大臣官房 情報管理課	(兼本委員会)
委員	太田 博親	通商産業省 産業政策局 調査課	
委員	杉原井 康男	通商産業省 貿易局 貿易保険課	
委員	上原 明	通商産業省 通商産業研究所 政策情報システム部	(兼本委員会)
委員	栗川 正仁	通商産業省 通商産業研究所 政策情報システム部	
委員	石橋 延弘	通商産業省 通商産業研究所 政策情報システム部	



目 次

第1章 本調査研究の経緯と成果の概要

1 1	調査研究の目的	3
1 1 1	調査研究全体の目的	3
1 1 2	本年度の目的	4
1 2	調査研究の背景	4
1 2 1	全体的な背景	4
1 2 2	本年度の調査研究の背景	6
1 3	調査研究の経緯と成果の概要	7
1 3 1	昭和63年度の調査と成果	7
1 3 2	平成元年度の調査と成果	8
1 3 3	平成2年度の調査と成果	9
1 4	調査研究の結論	9

第2章 第2期試作システムの概要

2 1	拡充要求仕様書作成の経緯と変更部分	13
2 1 1	第1期試作システムに関する評価	13
2 1 2	ユーザの利用目的に関する考察	17
2 1 3	第1期試作システムに関する拡張の方向性と技術的ポイント	17
2 1 4	拡充要求仕様書の作成と内容変更の経緯	21
2 2	第1期システムとの相違点(改善項目)	24
2 2 1	可能な検索支援の範囲	24
2 2 2	第2期試作システムのモジュール構成	24
2 2 3	知識獲得の仕組み	26
2 2 4	次善情報の種類と支援の仕組み	27
2 2 5	その他の改善項目	29
2 3	知識獲得及び次善情報提供のためのインタフェース	31
2 3 1	知識獲得のためのインタフェース	31
2 3 2	次善情報提供のためのインタフェース	36

2 4	知識ベースの構造と働き	42
2 4 1	IRS汎用分類	42
2 4 2	類義語辞書	43
2 4. 3	導出辞書	44
2. 4. 4	代替テータ辞書	45
2. 4. 5	テータベース選択辞書	46
2 4 6	テータベース辞書	47
2 4 7	事例ベース概要	48

第3章 第2期試作システムの評価

3 1	システムに対する要求とその背景	53
3. 1 1	一般的なテータベース利用の現状と課題	53
3. 1. 2	経済統計テータベースの特徴と人工知能技術への期待	56
3 1. 3	経済統計テータベース検索支援システムへの要求	57
3 2	第2期試作システムの限界とその要因	61
3 2. 1	要件の整理	61
3. 2 2	経済統計に関するガイダンス機能及びデータに関する 利用上の注意に関する案内について	62
3 2. 3	複数系列の総合的な利用について	64
3. 2 4	次善情報の提供について	66
3. 3	要求仕様に対する実現方法の妥当性及び実現された機能の有効性	68
3 3 1	要求仕様に対する実現方法と評価	68
3 3. 2	各機能の実現形態と実現方法に関する再検討	70
3. 3 3	システム各機能の連携状況の妥当性	74
3 4	知識ベース構成の妥当性と有効性	75
3 4. 1	知識ベースの役割	75
3 4. 2	本調査研究における知識ベースへの期待	75
3. 4 3	第2期試作システムの知識ベース構成とその妥当性	76
3. 5	実用化の可能性に関する評価	79

第4章 実用化に向けた今後の課題

4. 1	対象ユーザの利用形態と必要な機能	85
4. 1. 1	利用目的の種類	85
4. 1. 2	システムへの要件と利用目的の関係	87
4. 1. 3	多目的利用への対応方法	89
4. 2	知識ベースの構造と知識獲得の方法	90
4. 2. 1	必要な知識の構造	90
4. 2. 2	書誌的情報の取扱	91
4. 2. 3	過去の検索経験に関する知識	93
4. 2. 4	知識獲得の方策	93
4. 2. 5	知識の品質保証及び知識のメンテナンス	94
4. 3	分野の拡張指針	96
4. 3. 1	検索要求の受入れ機能の拡張	96
4. 3. 2	複数系列の比較に関する処理	96
4. 4	統計分類に関する提言	101
4. 4. 1	現行の統計分類に関する問題点	101
4. 4. 2	統計分類上の問題に関する取組	102
4. 5	稼働環境	104
4. 5. 1	必要な稼働環境	104
4. 5. 2	パーソナルコンピュータの技術動向	104
4. 5. 3	実用的なシステムのための稼働環境の選択	105

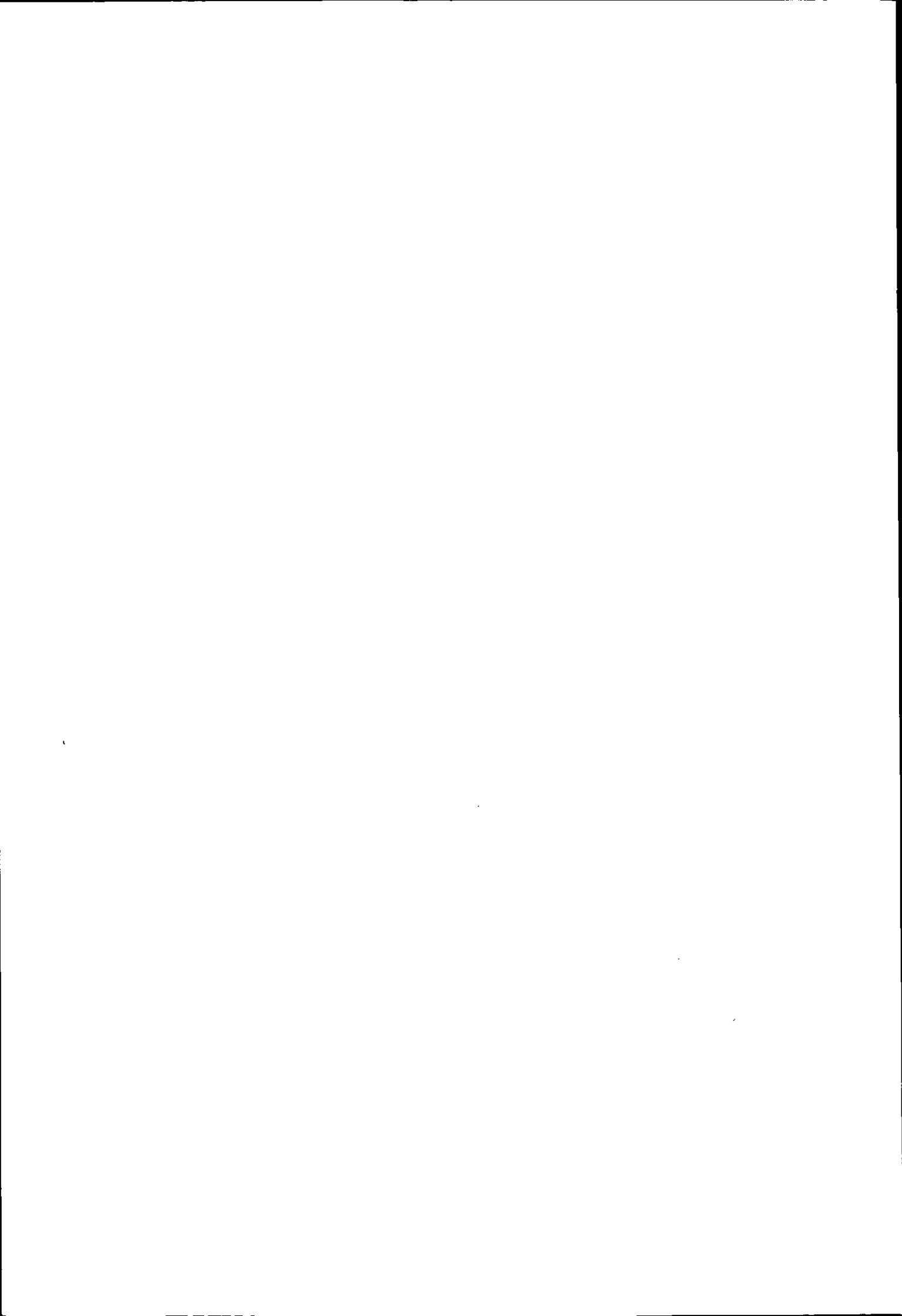
付録

1. システム拡充要求仕様書
2. ネットロにおける貿易統計データベースの工夫
3. 分類体系の問題
4. Frame Work of COMPASS
5. システムテスト結果概要

参考文献



第 1 章 本調査研究の経緯と成果の概要



第1章 本調査研究の経緯と成果の概要

本調査研究は、人工知能技術を応用して利用者とデータベースの間のインタフェースの向上を図ることを目的に、昭和63年度より3ヵ年計画で始められたもので、本年度は調査研究の最終年度に当たる。

本章では、調査研究の目的、背景、経緯、及び3年間の成果について概説する。

1. 1 調査研究の目的

1. 1. 1 調査研究全体の目的

本調査研究では、経済統計に関するデータベースの利用に焦点を絞り、データベースに関する知識、データベース利用に関する知識等を調査し、人工知能技術を応用して知識ベース化の検討を行い、利用者とデータベースの効率的なインタフェースの向上を図ることを目的として調査研究を行う。

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おり、対象となる分野も拡大しつつある。また、利用の局面も、従来の専門的な利用から日常生活の場面での利用へと拡大し、多面化してきている。

利用の多面化に伴い、データベースによる情報提供の手段も多様になり、遠隔地からの利用や一般家庭からの利用が可能になり、また、CD-ROMによるオフラインでの利用も行われている。

このような状況のなかで利用者がデータベースを真に活用するためには、データベースの所在、その構造、収録レコード・キーワード体系、検索コマンド体系等のデータベースそのものに関する知識と共に、収録されているデータの意味・内容や利用上の制約事項、分析の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、利用者に求められるこれらの知識は、量的にも多く、益々複雑、多様化しており、データベースの利用普及にあたっての隘路の一つとなりつつあるのが現状である。

この解決には、分野毎に必要な知識を整理し、データベースの利用者が、収録されているデータと共に整理された必要な知識を活用できる環境を整備する必要がある。

1. 1. 2 本年度の目的

昭和63年度は、経済統計データベース利用の現状調査、人工知能技術の現状調査、海外におけるデータベース検索の技術動向調査を基に、利用支援のための知識、利用支援システムを実現するための技術に関する検討を行い、実用システムのための目標設定を行った。

平成元年度は、初年度の成果を受けて、知識の整備及び、実用システムに必要なマンマシン・インタフェースに関する検討を重点的に行い、試作システムIRS (Intelligent Retrieval System)を構築した。

本年度は、2年目に開発した試作システムを基に、更に実用的なシステムへの課題を設定し、試作システムの拡張を行い、検証を行うと共に、異なる統計分野への拡張も含めた今後解決すべき問題点を検討することを目的とする。

1. 2 調査研究の背景

調査研究の背景については既に、初年度の報告書1章のなかで、データベース利用の現状と人工知能技術に関して詳しく説明し、さらに2年目の報告書1章でもこれを要約しているが、最終年度の報告に当たり改めて問題意識と解決方法を提示する。

1. 2. 1 全体的な背景

①経済統計データベース利用の現状

経済統計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。

経済統計データベース利用の現状

・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々が利用するケースが増えている。(例、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)

・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。

(例 データのグラフ化、外部情報と内部情報との統合加工等の要求が増えた)

次に挙げたデータベース利用の問題点にも、これらの傾向が反映されている。

データベース利用上の問題点

- ・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。
- ・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）
- ・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。
- ・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。
- ・欠測データに関する情報が判りにくい。
- ・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。

②人工知能技術の動向とデータベース検索支援への適用について

人工知能技術のなかに、種々のものが含まれるが、本調査研究のための技術的手段として期待されるものは、エキスパートシステムである。

エキスパートシステムは、専門家の知識を知識ベースと呼ばれる一種のデータベースに蓄え、これを使って、従来のシステムより、さらに柔軟に、問題解決を行うシステムである。

①で述べたデータベース利用上の問題点を、エキスパートシステムの技術によって解決する場合の適用局面を下記にまとめる。

エキスパートシステム技術の適用可能局面

- ・ユーザの要求に対する適切なアドバイス機能
- ・経済統計に関する専門的知識の知識ベース化による検索システムの高度化
- ・統計的な概念の表現方法として、人工知能技術を応用することによる、柔軟なユーザ・インタフェースの構築

即ち、①専門家でないユーザの要求を解釈する。②専門家の検索技術をシステム化する。③検索したい内容（検索要求）を柔軟に表現する。の各要求について、人工知能技術の適用による効果が期待される。

1. 2. 2 本年度の調査研究の背景

本年度は、試作システムの拡張を通じて、実用化のための課題をまとめることが目的であるが、課題抽出にあたって関わりの深い技術的背景としては次のものが挙げられる。

(1) ハードウェアの性能向上

試作システムIRSは、パソコン上で構築されたが、本調査研究の実施中にも、ハードウェア技術は、日々向上している。

パソコンの従来型のOS（オペレーティングシステム）である、MS-DOSの他に、新たなOSとしてOS/2が普及しはしめ、メモリの制約や複数処理の同時実行等の面で環境が改善されつつある。

また、ワークステーションの小型化、廉価化も進み、パソコンとワークステーションの境界が不鮮明になりつつある。

さらに、情報を蓄えるメディアとしてCD-ROMが普及し、データベース自体が大型コンピュータ上にある必然性が減少しつつある。

(2) 人工知能技術の高度化

従来のルールベース推論技術の他に、高次推論と呼ばれる新しい推論方法が研究され、一部実用的なシステムへの利用が始まった。従来のルールベース型推論では解決しにくい問題への対応可能性が増してきたと言える。

また、機械学習に関する研究も進み、学習機能を持つエキスパートシステムや、人間の神経回路網の仕組みを真似たニューロコンピューティング技術が紹介され、利用されるようになってきた。

本年度の調査研究では、試作システムの拡張の一環として高次推論の一種である事例ベース推論を取上げ、また一部の辞書に関して機械学習の機能の実現を試みた。

1. 3 調査研究の経緯と成果の概要

本調査研究は、1. 1の調査研究の目的に従って、3ヵ年計画で実施された。
以下に、各年度での、経緯と成果の概要を解説する。

1. 3. 1 昭和63年度の調査と成果

昭和63年度は、データベースの構造、収録コート・キーワード体系、検索コマンド等のデータベースに関する知識・ノウハウと、統計に関する利用上の制約、分析等、データ利用に関する知識・ノウハウ等について、調査分析を行い、これらの知識ベース化の方法について設計を行った。

具体的には下記の調査方法によって実施した。

- ①実際のデータヘース利用現場に対して、データベースシステムの現状と問題点についての報告を依頼する。
- ②データベース検索への人工知能技術の利用例について調査を行う。
- ③第2回「AIと統計に関する国際ワークショップ」に調査員を派遣する。
- ④「通産ジャーナル」より、貿易情報に関する記事を抽出し、検索要求文を作成する。
- ⑤ ④で作成した検索要求文について、専門家による意味の解釈をおこない、不足情報や参照すべき経済統計、使用すべきデータベースを付加したワークシートを作成する。
- ⑥3種類程度のデータベースを想定して、検索支援の要件を整理する。

昭和63年度の成果としては、下記が挙げられる。

- ①5種類のデータヘースシステムに関する事例から、実際のデータベースシステムで工夫されている点や問題点が明らかになった。
- ②人工知能技術を応用したデータヘース検索システムの事例から、検索支援に必要なユーザ・インタフェースや、仕組みが明らかになった。
- ③海外での事例調査のなかから、経済統計データベースの利用に対する人工知能技術の応用例に関する資料を収集し、海外での調査研究の状況を把握することができた。
- ④ユーザの検索要求に関するワークシートの作成を通して、検索要求の姿が明らかになった。

⑤理想的な経済統計データベース検索システムに関する目標設定を行い、外部仕様を作成した。

1. 3. 2 平成元年度の調査と成果

平成元年度は、前年度の調査結果に基づき知識情報の整備を行うと共に、インタフェース機能の概念設計及びシステム設計を行い、実験システム（IRS）の構築を行った。また、ヨーロッパでの統計データベースに関する支援システムの動向を調査した。

調査方法は、下記によった。

- ①前年度に作成した、検索要求のワークシートを基に、ユーザの検索要求に対する専門家の支援の方法を分析する。
- ②検索要求と、専門家の支援との関係から、データベース検索に関する概念的なモデルを作成する。
- ③概念的なモデルをもとに、システムモデルを作成する。
- ④パソコン上で、PROLOG言語とフロントエンドプロセッサを利用してホストコンピュータ上にあるデータベースをアクセスする機能を持った検索支援システムの設計を行い、試作システムを開発する。
- ⑤開発したシステムを評価し、今後の課題を整理する。
- ⑥スウェーデン、フィンランド、イタリアに調査員を派遣し、統計データベースシステムへの人工知能技術の応用例を調査する。

平成元年度は下記の成果があった。

- ①検索システムのための概念モデル、システムモデルを作成し、検索支援の様相と、システムとしての仕組みを整理した。
- ②試作システム（IRS）を作成し、検索支援のための機能に関して試作システムに基づく評価を行った。
- ③ヨーロッパでの調査結果に基づき、試作システムとの比較を行った。

1. 3. 3 平成2年度の調査と成果

平成2年度は、試作システムに対しさらにインタフェース機能の充実を図り、実用システムとしての適用可能性について評価検討を行った。

平成2年度の調査方法は下記の通りである。

- ①前年度の試作システムに関する評価をもとに、再度拡充要求仕様を作成し評価検討する。
- ②経済統計データベース以外の分野に関する検索支援の事例に関してヒアリングを行う。
- ③統計分類に関する問題点について他システムでの対策事例に関して、ヒアリングを行う。
- ④①の仕様にに基づき、前年度開発した試作システムを拡張する。
- ⑤拡張した試作システムを評価し、実用化のための課題を抽出する。

調査研究の実施により、下記の成果が得られた。

- ①試作システムに関する拡充仕様書の作成により、システムに対する要求を明確にした。
- ②文献データベースに関する事例、及び国土情報データベースに関する事例のヒアリングにより、試作システムの位置づけや、意義を確認した。
- ③統計分類の問題について、独自の統計分類を定義した事例、統計分類の定義に関しての変更や検索に対する柔軟性を持たせたデータベース設計の事例、統計モデルを組み込んだデータベースシステムの事例に関するヒアリング等により、統計分類に関するユーザ・インタフェースの有り方や、内部でのデータの構造についての留意点が明らかになった。
- ④前年度の試作システムを拡張した。特に、知識獲得の機能と、必要な情報が存在しない場合の次善情報生成の機能を新たに付加し、これらの機能の評価を行った。
- ⑤試作システムを基に評価を行い、今後の課題を抽出した。

1. 4 調査研究の結論

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。これによって導かれた結論は、次の通りである。

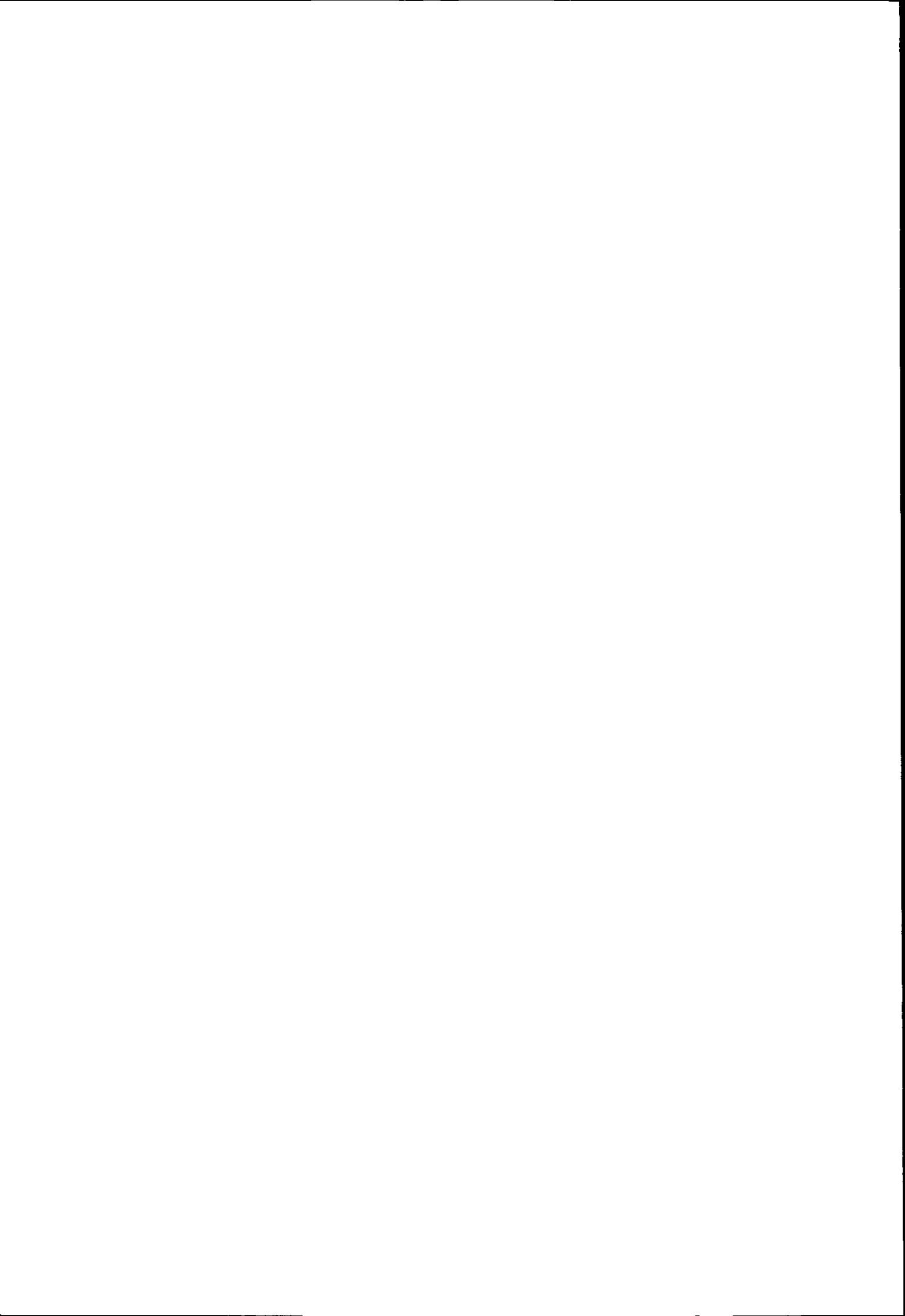
検索支援システムに対する要件は5種類程度に分類される利用目的に応じて異なる。従って、検索支援システムが機能の異なる複数の部分システムから構成されるべきである、というのが第1の結論である。

第2の結論は、検索支援に必要な知識についてである。専門家でないユーザを対象とした検索支援には、①検索要求に関する知識、②ユーザの用語を統計概念に置き換える知識、③情報源選択のための知識、④次善情報生成のための知識、⑤書誌的情報に関する知識、⑥過去の検索経験に関する知識が必要である。⑤以外は実験により知識ベース化の可能性と有効性を確認した。

さらに上記⑤の知識に含まれる複数系列の比較の問題について知識ベース化の方向性を提案すると共に、統計調査そのものを企画の段階で標準化する事を提言した。

最後に、データベースシステムの稼働環境の動向を論じ、将来の稼働環境においても検索に必要な知識の表現技術が本質的であることを確認した。

第2章 第2期試作システムの概要



第2章 第2期試作システムの概要

本章では、第2期試作システムに関して、第1期試作システムとの相違点を中心に解説する。(前年度に開発した試作システムと、本年度に開発した試作システムを区別するため、前年度のシステムを第1期試作システム、本年度のシステムを第2期試作システムと呼ぶ)。

2. 1 拡充要求仕様書作成の経緯と変更部分

付録1に第1期試作システムに対する拡充要求仕様書を掲載する。本節ではこの拡充要求仕様書の作成と検討の経緯を説明し、第2期試作システムで実際に実現した機能との相違点について説明する。

2. 1. 1 第1期試作システムに関する評価

(1)第1期試作システムの概要

第1期試作システムでは、初年度に作成した外部仕様に基づき、検索支援に必要なシステムの枠組みを作成した。

① 環境

第1期試作システムは下記の環境にインプリメントした。

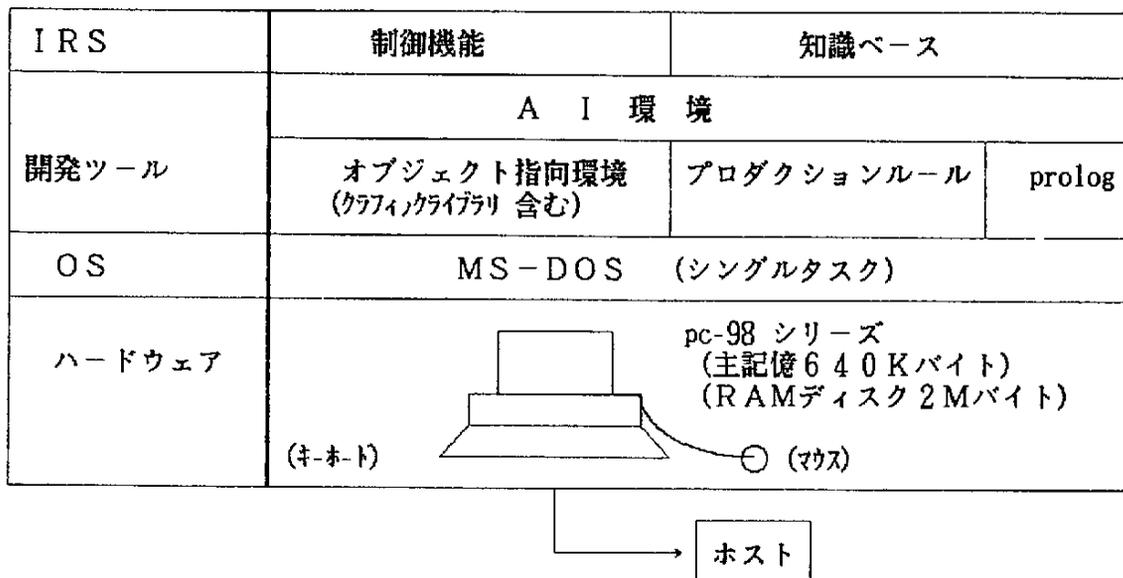
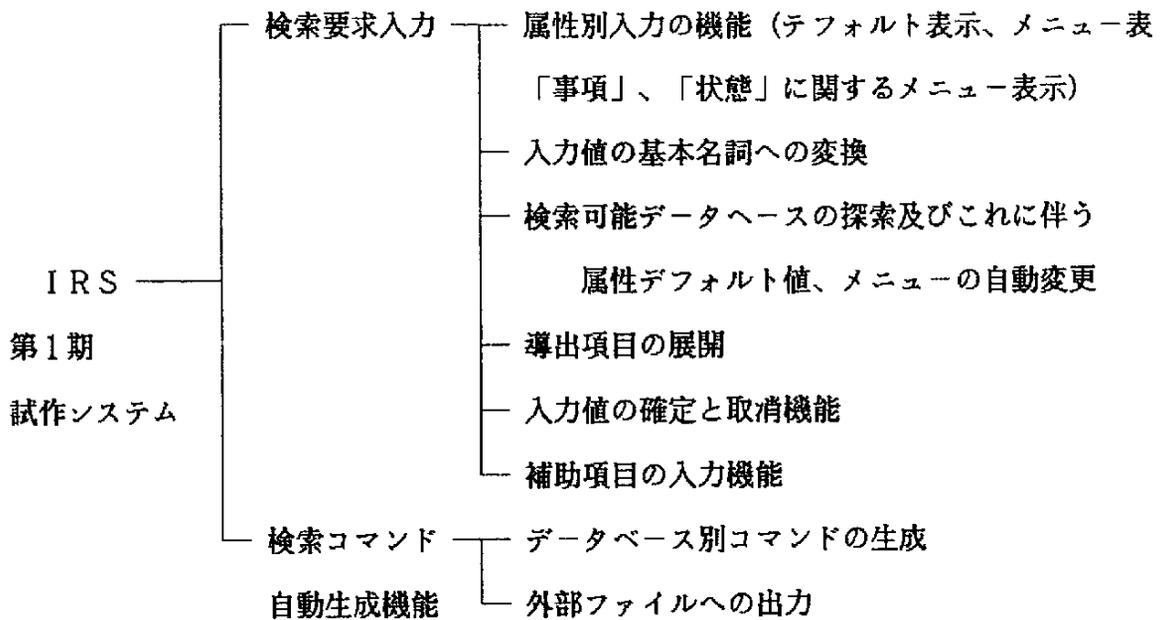


図2-1 第1期試作システムの稼働環境

② 機能概要

第1期試作システムで実現した機能は次のようなものである。



③ 主要知識ベースのメモリ専有量 (概算)

第1期試作システムのコンピュータ上の記憶領域は、表2-1のようになった。

表 2-1 主要知識ベースのメモリ専有量

知識ベース名	使用メモリ (KB)
DB辞書	—
類義語辞書	—
国名	8 (180か国)
品名	28 (900語)
導出辞書	—
導出式	—
状態辞書	—
検索コード辞書	—
国名	8
品名	28
系列コード	2

④ 試作システムの処理と制約

第1期試作システムに関して、処理の流れと、知識との関連、及び現在のシステムの制

約をまとめたものが、図2-2である。

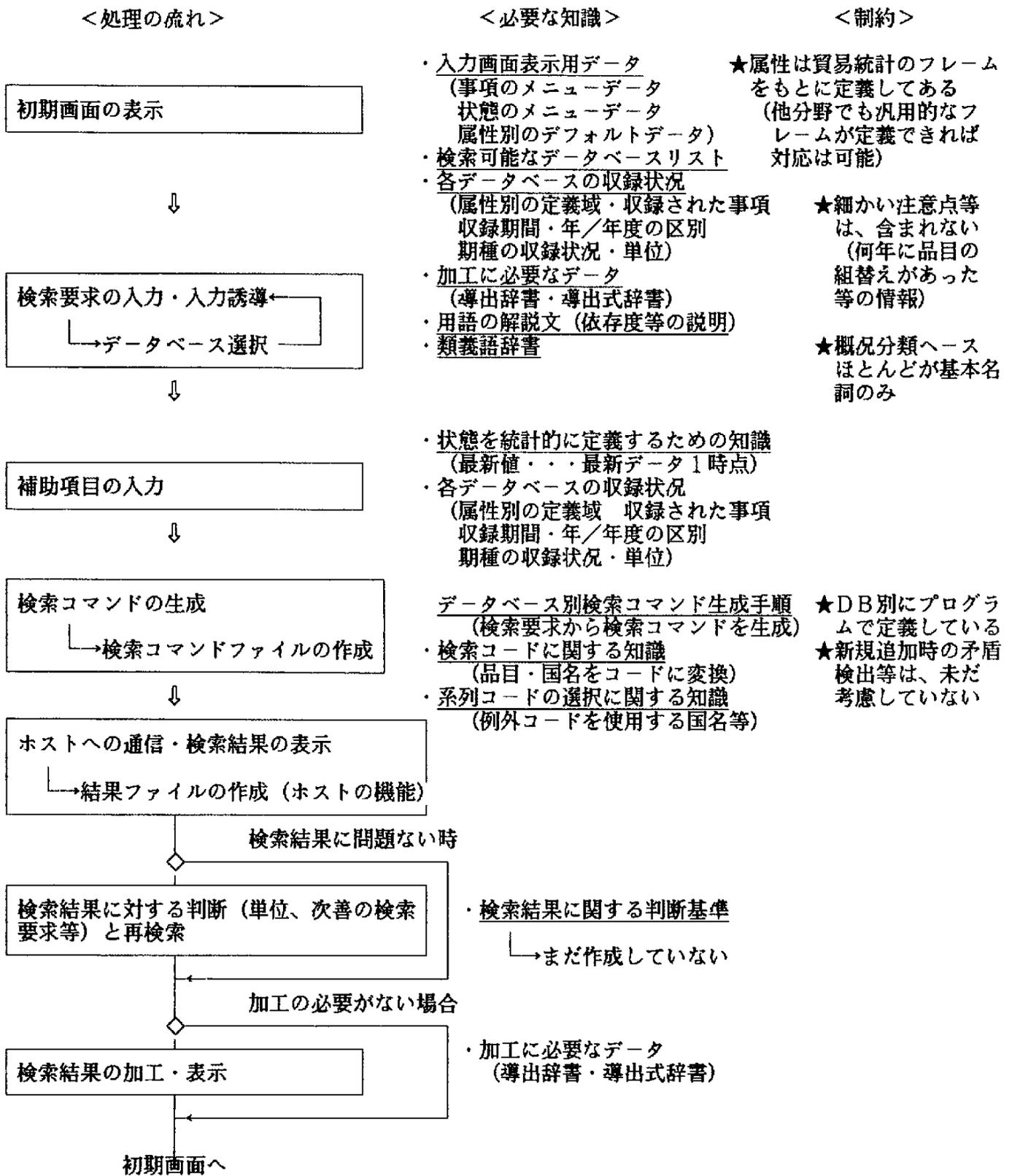


図2-2 第1期試作システムの処理の流れと知識の関係

(2)以降で第1期試作システムに関する評価を以下に要約する。(詳しくは、平成2年度報告書6章の解説を参照されたい。)

(2)問題解決能力に関する評価

第1期試作システムでは、3種類のデータベースを対象として、日常的な用語で表現された検索要求を受け付ける。また、「最新値」や「最近10年の推移」等、ユーザの利用目的に近い表現方法での選択が一部可能である。さらに導出辞書の機能を持ち、直接データベースに収録されていない情報を知識ベースを使って合成することができる。

但し、複数系列の比較や、定期的なデータの監視等については、対策ができていない。

(3)操作性に関する評価

第1期試作システムでは、データの取消や、成功した手続きの保存機能がない。また、テンプレート型の入力画面に対して、マウスとキーボードを併用して、ユーザが検索要求を入力する仕組みになっているが、操作に慣れないユーザにとって、マウスとキーボードの併用は負担になると思われる。

さらに、検索要求入力時に、入力内容に応じて自動的に検索可能なデータベースが絞り込まれ、メニューや入力枠の内容が変化する機能(入力誘導)については、実際に欲しい情報とは異なる情報に誘導される危険がある点が指摘されている。

即ち第1期試作システムは、情報が存在しないという事態を避けるために自動的に入力を誘導したが、必ずしもユーザにとって満足のゆく結果にはならなかった。この問題は、むしろ必要な情報が存在しない場合の次善情報をどのようにユーザに示すか、という問題として捉えられるべきであることが指摘された。

(4)知識の拡充及び保守に関する評価

第1期試作システムでは、知識とプログラムを切離して、知識ベースに知識を蓄えることには成功したが、知識ベースの拡充や保守の面での機能はまだ充分ではない。

特に品目の類義語辞書は、類義語の種類が多く、個人的な用語の使い方も考えられるので、さらに柔軟に新規登録や、変更を受け付けるようにしておく必要がある。

また、実用的な知識ベースを作成するために、運用時に知識を獲得する機能の重要性が指摘された。

以上の第1期試作システムに対する評価のなかで、第2期試作システムのための拡充要求仕様に反映させたものは、特に、次善情報に関する機能と、知識獲得の機能である。

2.1.2 ユーザの利用目的に関する考察

前年度の調査研究で、①研究業務、②企画業務、③定例的調査業務、④臨時調査業務、の4種類の利用パターンを、データベースのユーザに対するアンケートにより抽出した。

データベースの利用者は多様であり、検索支援に必要な要件は、利用目的によって異なると考えられる。

第2期試作システムでは、先ず定例的調査業務を想定して拡充要求仕様を検討した。

2.1.3 第1期試作システムに関する拡張の方向性と技術的ポイント

第1期試作システムの評価を終了した時点で、以下の各点か、拡張項目として挙げられた。

(1) 知識ベースの充実

システムの問題解決能力は、知識ベースの充実に依存する。

第1期試作システムの知識ベースは、枠組みができた程度の少ない知識であり、解決できる問題の範囲も限られていた。

知識ベース充実の方策として、知識の獲得機能を付加するという方向で、拡張指針がまとめられた。

また、知識ベースに関する管理責任に関しては、類義語辞書はユーザ、導出辞書は統計専門家、検索コマンドに関する管理はデータベース専門家、というように、担当範囲を設定したが、この点に関しては、なお検討の余地があり、本報告書第4章で改めて考察する。

(2) 次善情報の提供機能の追加

ユーザの要求に近い情報を提供するためには一旦入力を全て受け付けた後にデータの有無を判断し、必要に応じて次善情報の生成を行う必要がある。

また次善情報は1種類ではなく、複数のものが存在すると思われるので、調査分析を行う必要がある。

さらに、検索を行う前にデータがないことが判る場合と、検索後に判る場合の2種類のケースについて対策が必要である。

(3) 分野の拡張

第1期試作システムの問題解決の範囲は、貿易統計に関する情報のなかの一部である。

他分野の統計等を利用して横断的な情報検索を行う場合に、このアプローチで処理可能か否か検討が必要である。

(4) 要求指定方法の拡張

第1期試作システムでは、検索要求は報告国、相手国、品目等の属性単位で入力していたが、実際の検索には、さらに、目的に近い表現、例えば'対米輸出の動向'等の要求指定機能により、過去の実績(白書での利用等)のある情報を得たい場合がある。

また、同時に複数の系列に関する検索要求を指定したい場合も多いので、対策が必要である。

(5) 検索可能なデータベースの追加

第1期試作システムでは3種類のデータベースを検索対象としたが、さらに多くのデータベースを検索する場合は、一つの検索要求に対して検索可能なデータベースが複数存在するケースが多くなる。従って、検索可能なデータベースを追加する場合は、的確なものを選択するための機能をさらに充実させる必要がある。

また、現在数値データのみを対象としているが、文献データも検索範囲に含めた場合の対策も必要である。

(6) データの加工機能の追加

データ検索の結果を報告書にまとめたい場合は、グラフ作成や表作成の機能およびデータの統計的処理による分析の機能等が必要になる。

(7) 検索履歴の保持

データ検索者は、定期的に同じ検索を繰り返すことが多いことから、一度検索に成功した場合、検索に必要な手続きを保存し、2回目からは簡単に情報検索を行える機能が実用的なシステムには必要である。

以上の7項目のなかで、本年度の拡張範囲としては、特に、知識ベースの充実と、分野の拡張性の検討が重要とされたが、分野の拡張に関しては検討にとどめ、知識ベースの充実、次善情報の提供機能、検索履歴の保持か、第2期試作システムの拡充要求仕様として採用された。

知識ベースの充実に関しては、知識の更新、追加の機能が、拡充されることになった。

・次善情報の提供機能に関しては、次善情報の分類を行い、次善情報の提供に必要な辞書の仕組みを新たに実現した。

・検索履歴の保持に関しては、アラカルトメニューの機能を追加することにより、よく利用される検索要求をメニュー化すると共に、事例ベース推論の機能を使って、過去の検索経験を利用した処理の高速化を試みた。

次ページに、前年度の報告書より、拡張の方向性に関して、項目と技術的ポイントをまとめた表を再録する。

表 2-2 拡張の方向性

項目	技術的ポイント	備考
<p>知識ベースの充実 類義語辞書の充実</p> <p>状態辞書の充実</p> <p>導出辞書の充実</p> <p>説明文の充実 (データの属性等)</p> <p>知識の追加更新機能の充実</p> <p>知識の整合性チェック</p>	<p>☆類義語の収集</p> <p>☆基本名詞への変換速度の維持</p> <p>☆品目分類の体系化 (既存の分類体系ではなくユーザオリエントな分類体系を考慮する必要がある)</p> <p>☆経済白書等を使って、状態として登録すべきテクニカルタームを抽出する</p> <p>☆テクニカルタームを検索条件に展開して、知識ベースを作成する</p> <p>☆経済白書等を使って、導出辞書に登録すべき (因果関係によって導かれる) テクニカルタームを抽出する</p> <p>☆データの属性として、ユーザに提示すべき情報の収集、整理</p> <p>☆知識ベース追加更新用インターフェイスの仕様検討</p> <p>☆知識の矛盾検出のしくみ (知識間の関連性を考慮する必要がある)</p>	<p>◇類義語辞書は、大規模になればホストに移行も検討</p> <p>◇状態辞書と導出辞書の切り分け方は、再検討の余地がある</p>
<p>次善情報の提供機能の追加</p>	<p>☆品目分類、情報の因果関係等から次善情報を選択するための情報尺度を定義する</p> <p>☆上記尺度を使って再検索を行う仕組みを作る</p>	
<p>分野の拡張可能性 (P P I S 関連分野)</p> <p>国際経済 (財政・金融)</p> <p>国内経済 (マクロ経済・産業・企業活動・物価・労働・産業連関表等)</p> <p>資源情報、地域別経済情報、公害情報</p> <p>企業財務、商品市況、法令情報等</p>	<p>☆分野別フレームの検討</p> <p>☆分野間のメタ知識の抽出と知識ベース化 (品目分類と産業分類の関連に関する知識等)</p> <p>☆複合分野検索のためのインターフェイス検討</p> <p>☆知識ベースの高速逆引き機能 (例えば貿易→自動車→自動車産業→産業)</p>	
<p>要求指定方法の拡張</p> <p>過去の事例により指定</p> <p>アラートメニューによる指定</p> <p>複数の検索要求入力</p>	<p>☆事例の収集と知識ベース化</p> <p>☆検索手順の知識ベース化</p> <p>☆入力インターフェイスの検討</p>	
<p>検索可能なデータベースの追加 (貿易統計の)</p> <p>数値情報データベースの追加</p> <p>文献情報データベースの追加</p>	<p>☆検索コマンドの追加</p> <p>☆キーワード検索・キーワード合成等</p> <p>☆数値情報と文書情報の併用システムとしての要件抽出 (入力フレームの共有等)</p>	<p>◇文献検索は、本質的には別システム</p>
<p>データの加工</p> <p>統計パッケージとのリンク</p> <p>統計分析手法のアドバイス</p>	<p>☆データのワークシートと統計ソフト用ファイルへの変換</p> <p>☆統計分析のノウハウ収集と知識ベース化</p>	
<p>検索履歴の保持</p> <p>前回入力値を次回デフォルト値として利用</p> <p>メニュー・辞書の並べ換え (前回使用したものを先頭に置く)</p> <p>前回の検索要求をメニューに追加</p>	<p>☆履歴ファイルの作成 (ユーザ別カルテ)</p> <p>☆知識ベースのメンテナンス機構</p> <p>同上</p>	

2. 1. 4 拡充要求仕様書の作成と内容変更の経緯

2. 1. 1 から 2. 1. 3 で解説した第1期試作システムに関する評価を基に、本年度の実現可能性を考慮して、第2期試作システムのための拡充要求仕様書を作成した。

(付録1参照)

拡充要求仕様書に対して、実現した部分、仕様変更した部分、実現しなかった部分について、その経緯を述べる。

また、検討の過程で参考にした他分野事例に関するヒアリング及び、統計分類の考え方に関するヒアリングで、使用した資料が、付録2、付録3、付録4である。

参考文献(3)は、他分野での事例紹介として、テキストデータヘース検索におけるポストインデキシング化の試みとして、キーワードの自動抽出システムを紹介するものである。この例での問題意識と比較すると、試作システムに関しては、類義語辞書の利用により、インデキシングレスの検索システムの可能性が考えられる点に意義があると言える。

付録2は、統計分類への対応事例として、紹介された事例で、複数統計分類の対応関係をシステム化し、さらにユーザにとって判り易いものとするための、独自分類に関する資料である。拡充要求仕様書では、ユーザが新規に類義語を登録する場合に、直接特定分類のコードを入力させようとしていたが、付録2の例は、ユーザにとって判りやすい分類体系で指定させることが、データヘースシステムのメンテナンス上必要であることを示している。(付録1、付-12 ページ参照)

付録3は、分類体系の捕らえ方に関するシステム設計の立場からの考え方の紹介である。

この資料では、ユーザに用語の意味を指定させる場合のユーザ・インタフェースのアイディアとして、分類体系の木構造を表示する方法を提案している。この提案を参考にして拡充要求仕様書の類義語辞書新規獲得用のインタフェースを再検討した。

付録4は、複数の分野に跨がる横断的な検索の例として、経済モデルを組み込んだデータベースシステムを紹介している。ここでは、データ間の因果関係をモデルとして捕らえ、検索可能な情報を使って、必要な情報を得るための工夫が紹介されている。

さらに、他分野の事例としては、国土情報に関する検索システムの例が紹介された。このシステムは、データベースの論理構造と、概念構造が異なる点で、試作システムと共通しており、特にユーザ・インタフェースに関して注意すべき点が指摘された。この事例からは、試作システムで評価すべき点は、入力テンプレート方式になっている点であり、さらに分野を拡張した場合は分野に応じた入力テンプレートが必要である、また、データの記述は他分野からの参照を考慮すると概念スキーマとして記述すべきである等の点が指摘された。

上記検討を経て、拡充仕様書の類義語辞書の新規獲得機能に関して、分類体系を表示しユーザに対応する品目を指定させる方式に変更を行った。

また、ユーザに示す分類体系としては、このシステム専用のIRS汎用分類を定義することとした。(IRS汎用分類に関しては、2.4節で改めて解説する)

次に、拡充要求仕様書で説明している機能構成のなかで、実現したものと実現しなかったものを解説する。(図2-3参照)

次ページの機能構成図において、'新規に作成する部分'のなかで実現したものは次善情報の生成に関する機能、類義語新規登録機能である。アラカルトメニュー入力に関する機能は、属性確認・知識更新の機能と入力値の訂正の機能を除き実現した。

また、アラカルトメニューの内容は、数件に留まり、必要な式の枠組みを用意する事のみを検証した結果になった。これは、実際に用意されたアラカルトメニューに対する要求(通商白書の図表タイトルから抽出したもの)のなかで、与えられた3種類のデータベースで検索可能なものが少なかった点によるものである。

また、本年度改造を加える部分に関しては、検索コマンドの編集・知識更新については、インタフェースを用意した。コマンド保存も実現した。検索要求の入力のなかで、上位10か国といった上位指定については実現しなかった。これは、ホストとパソコンの通信方式が原始的なものであったため、上位指定のためのコマンドをホストに送るための処理が追加しにくい事が原因である。

知識ベースのメンテナンスのための機能に関しては、市販のエディタを利用することとした。これは、知識ベースメンテナンスのための要件が十分に抽出できなかったためである。入力値の訂正に関しては、実現した。

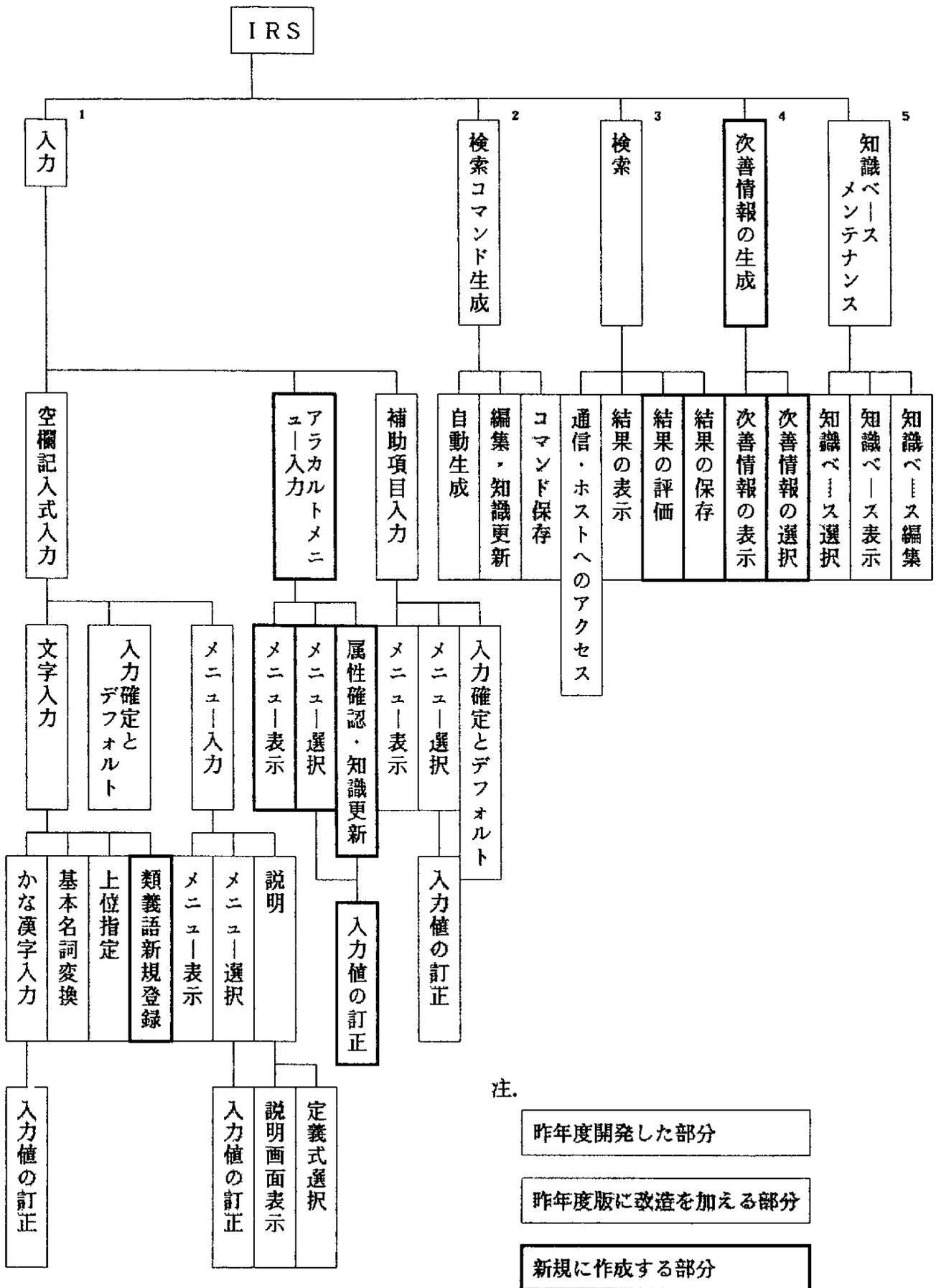


図 2 - 3 第 2 期試作システムの機能構成

2. 2 第1期試作システムとの相違点（改善項目）

第1期試作システムでは、検索要求入力のためのインタフェースと、コマンド生成機能を中心に開発を行い、検索要求入力と同時にデータベースの選択を行う点がシステムの特徴となっていた。

第2期試作システムでは、検索失敗の少ない検索支援を行うために、次善情報生成の機能（指定品目に関するコードが存在しない場合や、欠測の場合に対応するため）と、類義語辞書の知識獲得機能を中心に機能の強化を行った。

第1期試作システムの特徴であったデータベース選択機能付のインタフェースは、一旦ユーザの検索要求を全て受け入れてから、最も適切なデータベースを選ぶ方式に置き換えた。

本節では、第1期試作システムと第2期試作システムの相違点を中心に、第2期試作システムでの機能の実現方法について説明する。

2. 2. 1 可能な検索支援の範囲

検索可能な検索支援の範囲は、下記に示す範囲である。

- ① 検索可能なデータベース
日本貿易統計、各国間貿易統計、世界経済統計の3種。
- ② 検索可能な国名
全世界。
- ③ 検索可能な品目
民生用家電製品。
- ④ 検索可能な事項
輸出、輸入、双曲化指数、輸出／輸入シェア、GNP等15種類。

検索可能な範囲に関しては、第1期試作システムと比べて、若干の拡張程度である。

2. 2. 2 第2期試作システムのモジュール構成

実験システムは、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の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。そのなかの類義語辞書は、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。

(ここで、画面の表示等、入力制御のための知識は、知識ベース群から除き、専門家の知識に相当するものだけを知識ベースとして扱う)

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ックスを付与されている点が特徴である。

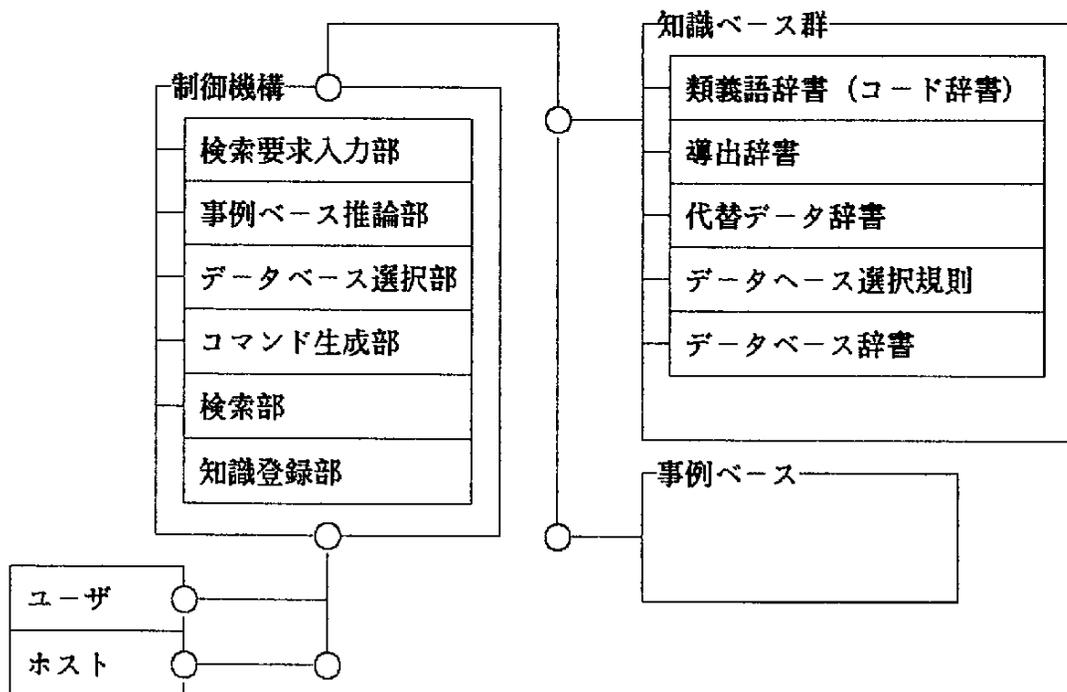


図2-4 モジュール構成

上記モジュール構成において、第1期試作システムとの相違点を挙げると、事例ベースに関する処理が追加された点と、代替データ辞書が追加された点である。また、ホストコンピュータとの通信機能は、第1期試作システムと比べて強化された。

事例ベースに関する処理は、既に経験した検索事例を蓄積し、新たな問題に適用することによって処理効率をあげるための仕組みである。

2 2. 3 知識獲得の仕組み

知識獲得の機能は、類義語辞書に関してのみ実現した。

類義語辞書に関する知識獲得は次のような方法で行われる。

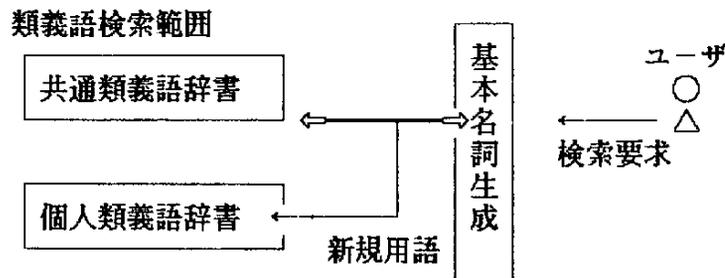


図 2 - 5 知識獲得の仕組み

- ①ユーザが入力した検索要求の内、報告国、相手国、品目に関して類義語辞書とのマッチングを行う。
- ②類義語辞書に未登録の用語が存在した場合は、類義語辞書新規獲得用の画面を表示し、ユーザが入力した用語が基本名詞のどれに相当するかの指定をさせる。
(現在のところ、該当する品目がない場合の処理は行わない)
- ③ユーザが指定した基本名詞と、入力された用語を使って、類義語辞書用の登録データを生成する。
- ④検索処理が終了した段階で類義語辞書の更新を行い、新たにユーザが指定した新規類義語に関する登録データを従来の類義語辞書の先頭に付け加える。

ここで、品目に関する類義語辞書では基本名詞に相当する用語を、IRS汎用分類として生成した。IRS汎用分類は、ユーザにとってわかり易い用語になっているが、各統計分類とは必ずしも1対1に対応していない。このためユーザが指定した用語が実際は検索不可能な場合もあり、改善情報の処理が必要となる。

基本的に辞書を含む検索支援システムは、個人別のフロッピーに保存されて運用されるために、個人別の類義語辞書が生成されることになる。

2. 2 4 次善情報の種類と支援の仕組み

次善情報のなかには、次の4種類が含まれる。

- ① 指定コードが存在しない場合に、下位コードを積み上げて合成する。
分類体系上の下位コードばかりでなく、他のデータに演算を施して得られる数値が代替情報として利用可能な場合も、このグループに含める。
- ② 指定コードが存在しない場合で、下位コードの積み上げ方も辞書にない場合、上位コードで代替する。
- ③ 事項に相当する用語が存在しない場合、GNPの代わりにGDPを使う例のように、概念的に近い用語で代替する。
- ④ 入力された属性を組み換えることによって、近いデータを提供する。

さらに、類義語辞書によって、入力の用語を基本名詞に変換する処理も一種の次善情報と考えると、5種類の次善情報が定義されたことになる。

第2期試作システムの処理の流れを次ページのフローで示す。

次善情報生成は、次ページのフロー中の☆印を付した部分で行っている。

(START)

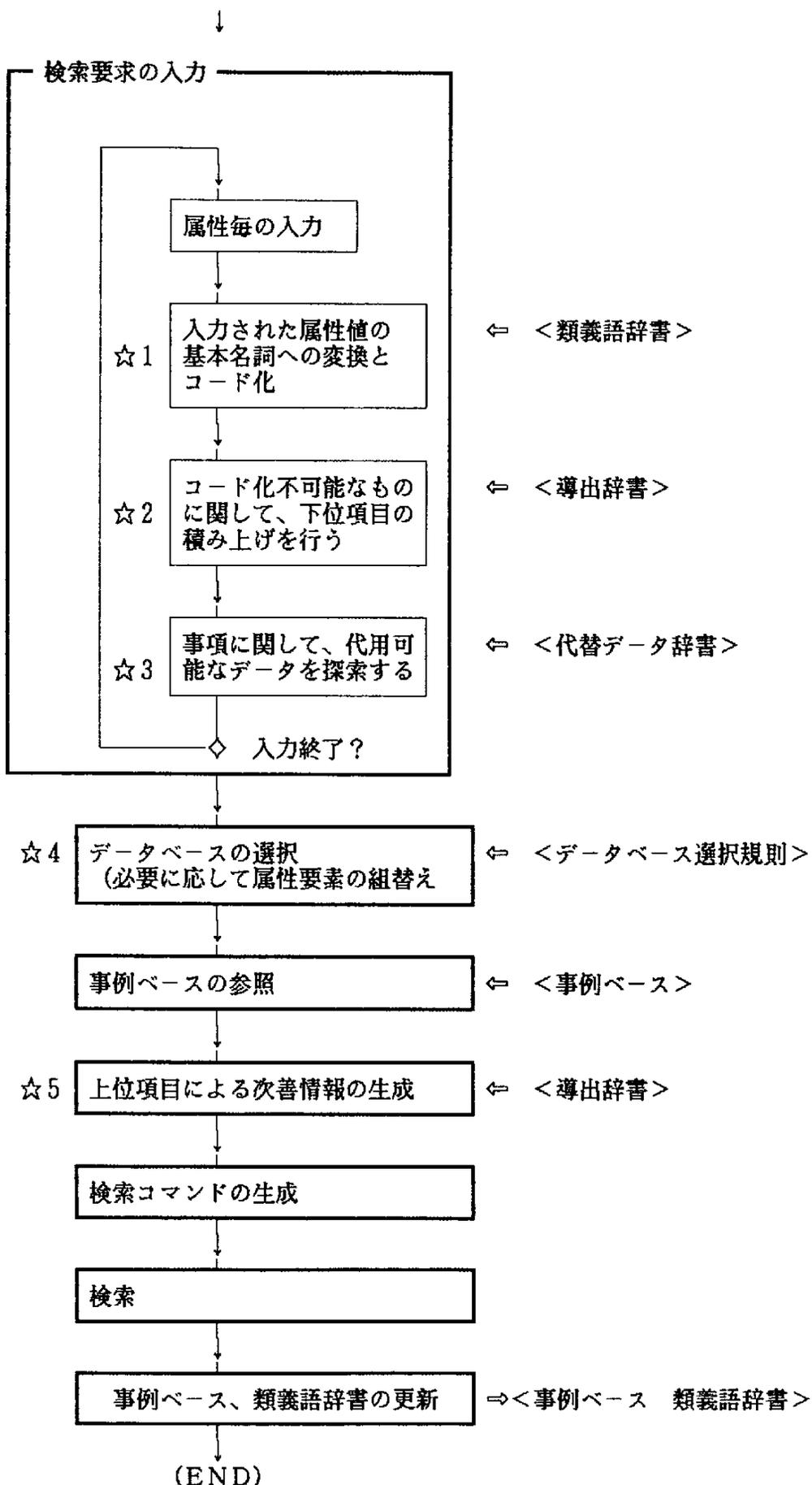


図2-6 試作システム(IRS)の処理の流れ

☆1から☆5の改善情報の処理について辞書との関係や、処理の内容を記述する。

☆1 入力された属性値の基本名詞への変換は類義語辞書を使って行われ、入力された属性値は、基本名詞に置き換えられる。

品目に関しては、ここで、コードの生成が行われ、検索可能性を確かめる。

☆2 下位項目の積み上げは導出辞書を参照して行う。

発見された下位項目のリストと入力された属性を置き換える。

☆3 事項に関しては、代替データ辞書により代替可能なデータ名を探索し記憶する。

(事項に関しては、置き換えずに、入力値と代替情報の両方を検索する)

☆4 データベースの選択は、データヘース選択規則を参照し、入力属性の属性値の組とマッチするデータヘースを探索する。

必要な場合属性を入れ換えたり、事項名を変更する処理も行う。(今のところ、相手国が日本で、商品名がある時、報告国と相手国を入れ換えて、輸出と輸入を入れ換える処理だけが可能)

☆5 上位項目による改善情報の生成は導出辞書を使って行う。

2. 2. 5 その他の改善項目

改善情報の提供と知識獲得の仕組みが第2期試作システムの大きな特徴であるが、その他にも第2期試作システムの特徴といえる項目があるので次に紹介する。

(1) データベースの選択方式

データヘースの選択を、検索要求の入力か全て終了してから行う方式に変えた。(第1期試作システムでは、1件属性を入力する毎にデータベースの選定を行っていた。)

検索要求が全て終了した時点で、検索要求にマッチする、データベースをデータベース

選択規則から探索するが、適切なデータベースが存在しない場合は、検索要求を一部変更してマッチするものを探す。

一部変更に関する知識もデータベース選択知識に収録されているが、現在はそれほど充実した知識内容にはなっていないので。

(2) 事例ベース推論の導入

事例ベース推論は、過去の経験を新たな問題解決の道具として使うための推論方式で、ルールベースになりにくい知識の蓄積に有効である。

第2期試作システムでは、事例ベースは、コマンド生成機能の一部として効率向上の目的で使われているため、表向きの機能にはなっていないが、本来の性質から考えると辞書の枠組みに収まらない、発見的な検索戦略や、属人的な知識の収集に、今後期待が持てる。

また、事例ベースは、一件の処理が終了するたびに結果に応じて事例ベースを蓄積してゆくので、知識獲得の機能も持っていると言ってよい。

(3) アラカルトメニューによる検索要求の入力を可能にした。

属性毎に検索要求を入力する方式の他に、さらに、検索目的に近い表現での検索要求の入力が可能になるよう、アラカルトメニュー方式の検索要求の導入を行った。

アラカルトメニューの各項目は、' 通商白書図表タイトルに関する重要度アンケート' (平成2年度報告書、付録2) より抽出した。

現在のシステムでは数件しかデータがないが、この知識が充実するとさらに少ない手間でデータベースの検索ができるようになる。

2. 3 知識獲得及び次善情報提供の為のインタフェース

本節では、第2期試作システムの特徴である知識獲得および、次善情報提供のためのインタフェースのイメージを、知識との関連を示しながら解説する。

2. 3. 1 知識獲得のためのインタフェース

第2期試作システムでは、IRS汎用分類（2.4節で詳しく解説）を使って、ユーザに新規用語の意味を指定させる。知識ベースのなかでは、類義語辞書がIRS汎用分類に従って作成されており、これを使って分類構造を示すことになる。

以下にインタフェースのイメージを操作手順に従って示す。

(1) 入力画面レイアウト

入力画面は、図2-7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。

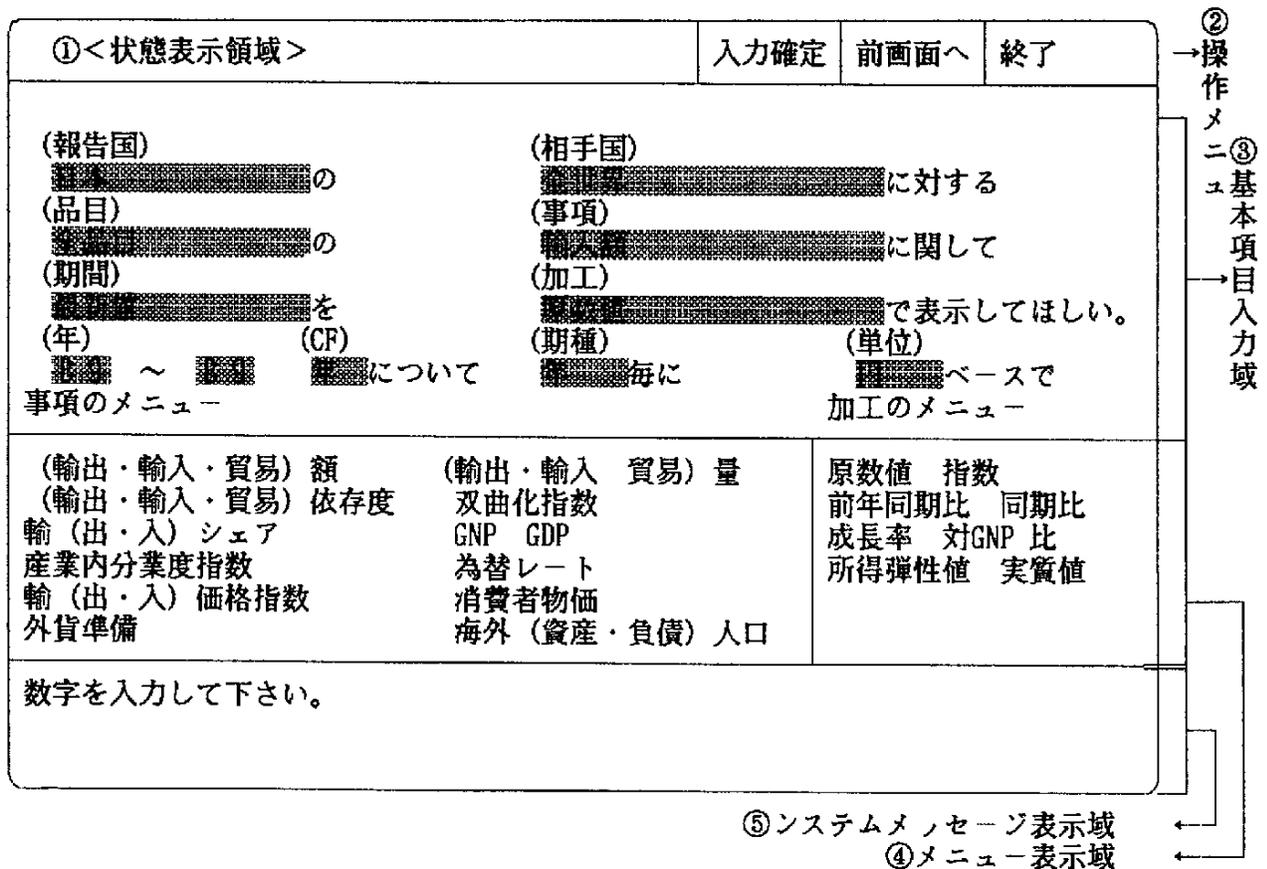


図2-7 画面レイアウト

(2) 検索要求の入力

ユーザは下図の画面から入力を行う。システムは、ユーザがどの属性入力枠に入力を行ったかを認識し、国名の場合は国名に関する類義語辞書、品目の場合は品目の類義語辞書を検索して、ユーザの入力に対応する基本名詞を探す。

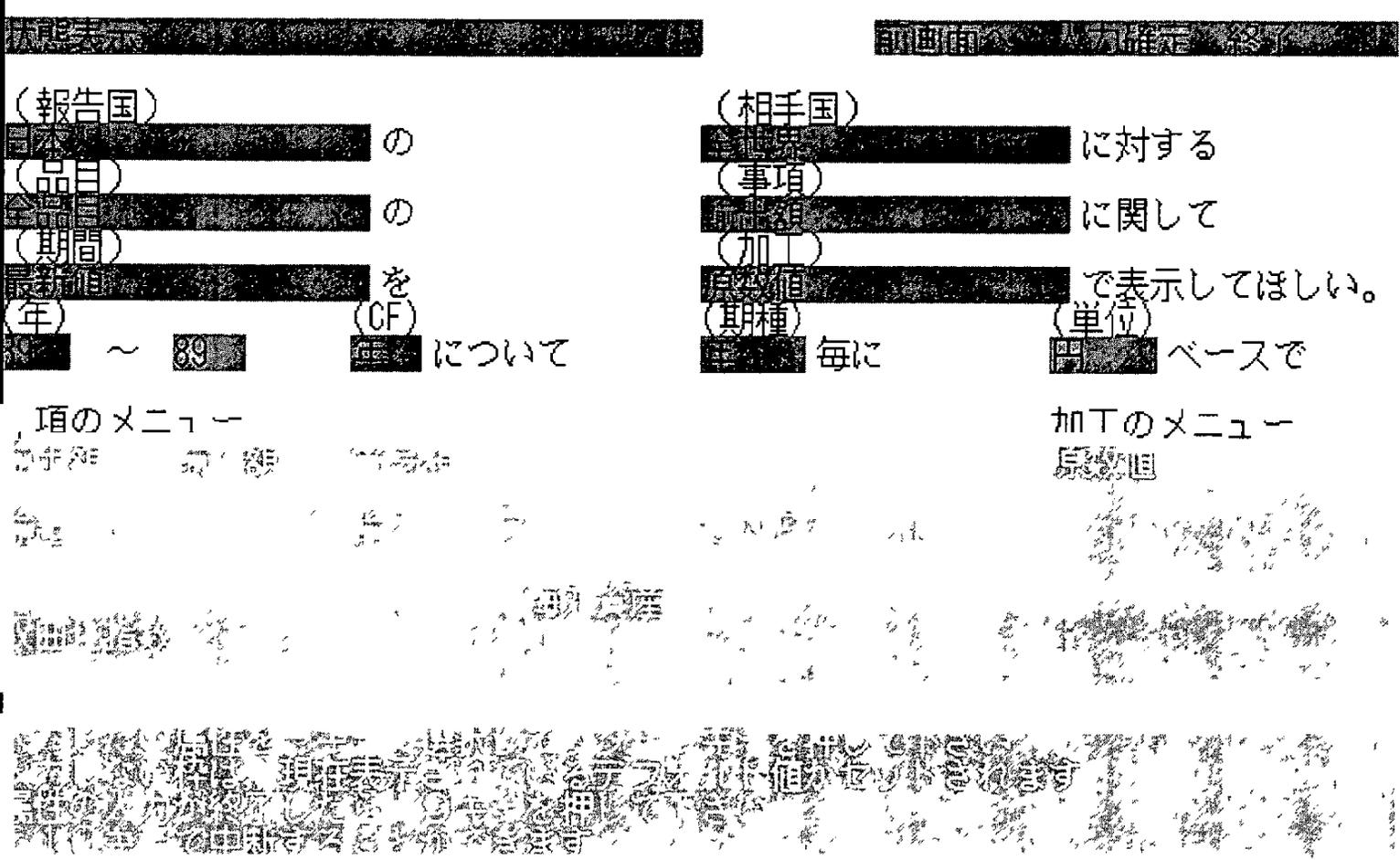


図 2 - 8 検索要求の入力 (相手国の入力)

(3) 新規類義語の登録

ユーザの入力した用語が類義語辞書に見当たらない場合は、下記の画面が表示されて、ユーザの基本名詞指定を促す。ここでは、例えば相手国に米国と入力した場合を想定している。

ユーザは、このページに必要な基本名詞がない場合は、'次ページ'の指示に従って、ファンクションキーのf8を押すと次のページが表示される。

表示の順序は類義語辞書の掲載順になる。この場合、国名は地域順、品目名はIRS汎用準分類順となる。

ユーザは必要な場合、複数の項目を指定できるようにインタフェースを用意した。(現在は、後処理の機能が不十分なので、実際の複数指定は不可能である)

品目の場合は、分類上の階層を類義語辞書から読み取り、階層に合わせて字下げを行うので、基本名詞は分類木状に表示される。

基本名詞の指定が終了したらf9キーを押すことにより、新規類義語の登録処理を終了する。

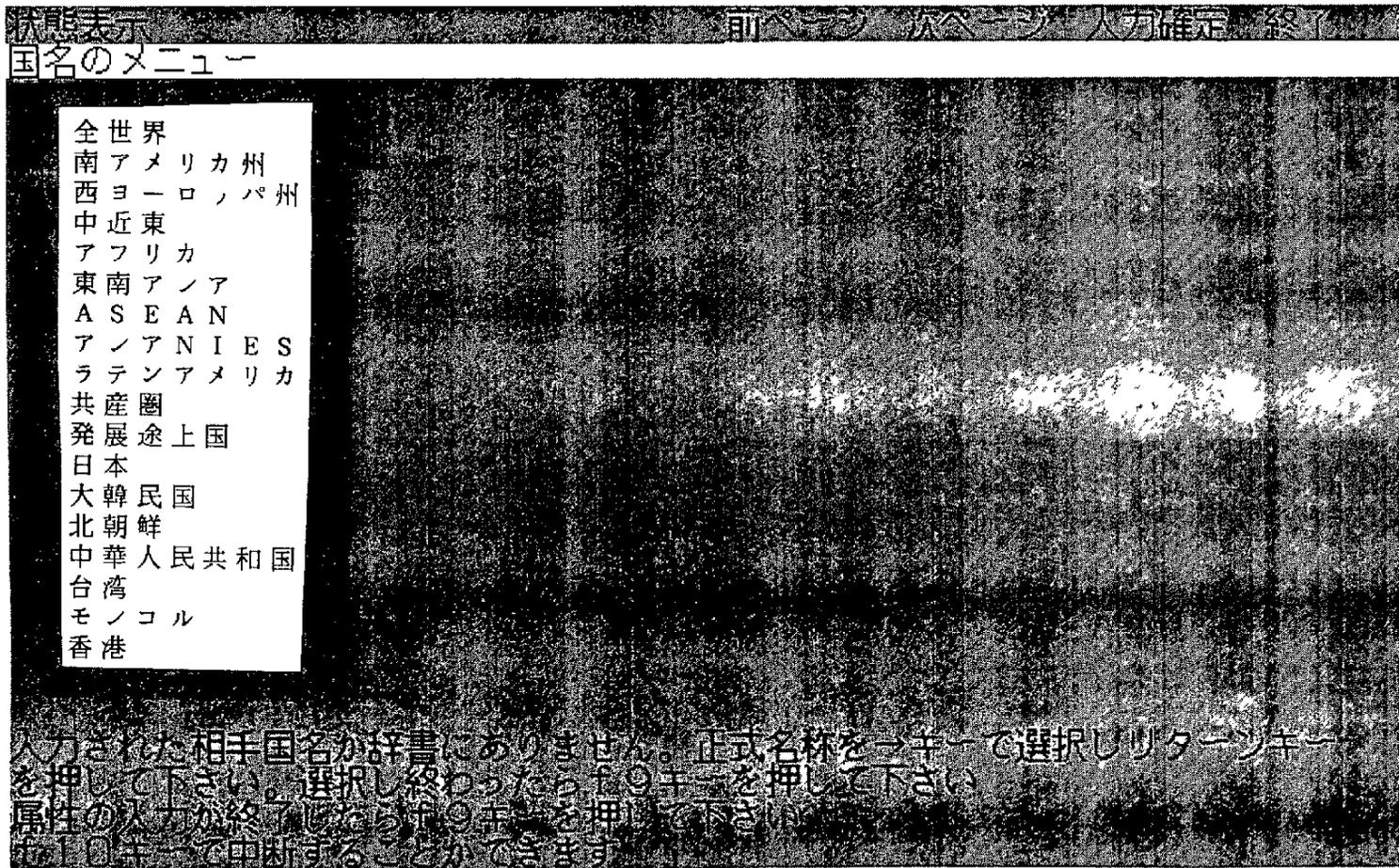


図2-9 新規類義語の登録

(4) 登録終了

登録が終了すると元の属性入力画面が現れるが、登録した基本名詞が、属性入力枠に表示された状態になっている。(この場合、相手国に'アメリカ'が入力されている) さらに検索のための処理が続き、検索が終了した時点で、アメリカに対する米国という類義語が類義語辞書の先頭に追加登録される。

以降の検索では、アメリカに関する他の類義語より優先して、米国が探索されるようになる。



図2-10 登録終了

2. 3. 2 次善情報提供のためのインタフェース

第2期試作システムでは、次善情報はシステムが自動的に生成する。ユーザへはメッセージを表示して次善情報を採用したことを示す。

品目に卓上扇風機、品目に貿易額を入力した場合と、品目厨房機器を指定した場合、および、事項にGNPを指定した場合のインタフェースを示す。

(1) 品目の入力を行う

この場合' takujyousenpuuki' と入力する。類義語辞書を探索して、対応する基本名詞を発見し、入力枠を基本名詞で書き換える。

類義語辞書を探索すると、卓上扇風機に対応するコードが999（対応するコードが存在しないことを示す）であるため、導出辞書で下位項目の積み上げが可能か否かを調べるが、可能ではないのでこのまま次の処理に移る。

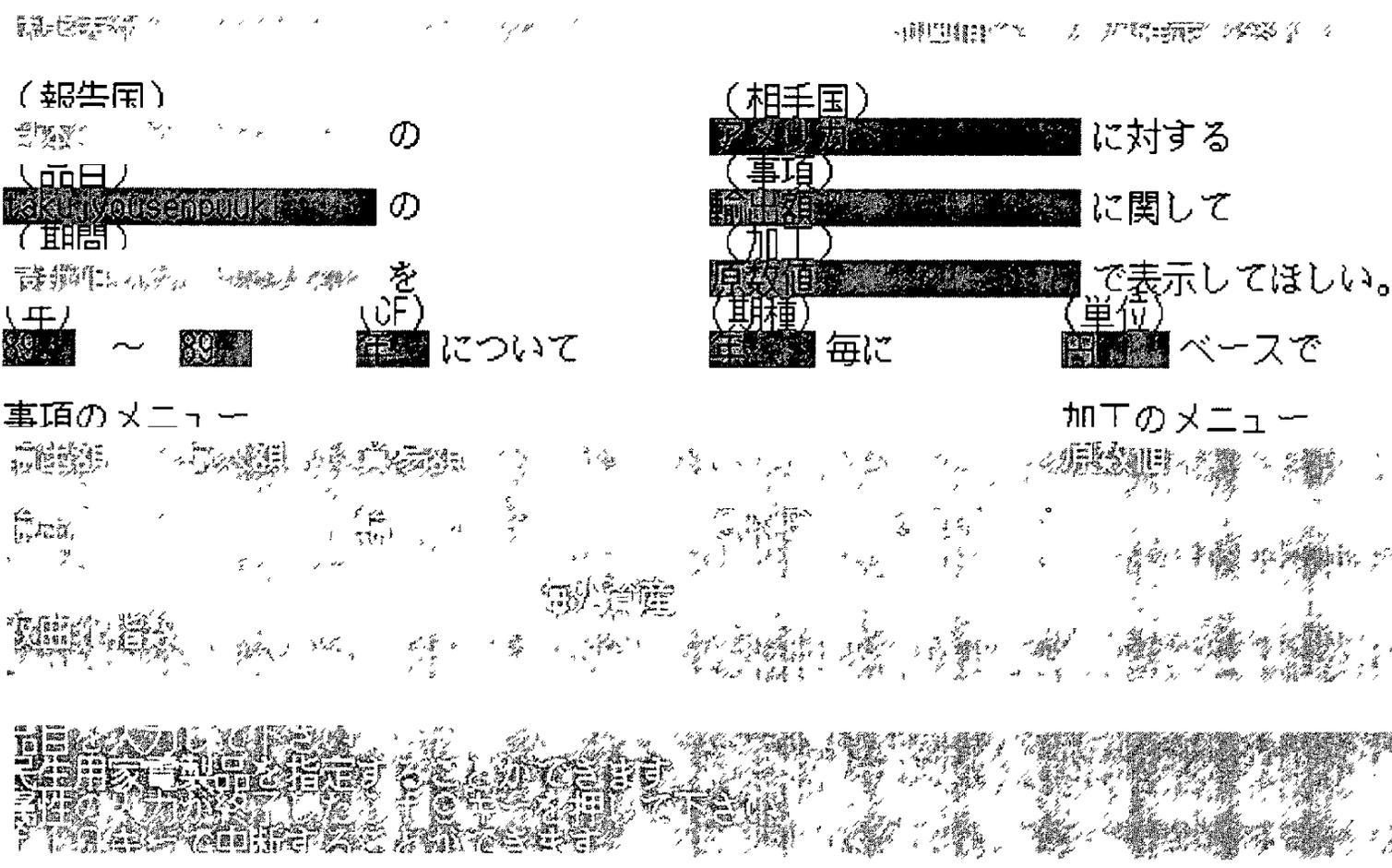


図2-11 品目の入力

(2) 事項の入力を行う

この場合、事項のメニューから貿易額を選択する。

システムは、導出辞書を探索し、貿易額に対応する導出項目を探索する。

導出辞書のなかから貿易額を構成する、下位項目のリストと、メッセージを探し、ユーザへのメッセージとして、画面に表示する。

システムは貿易額の代わりに、輸出額と輸入額を検索することになる。

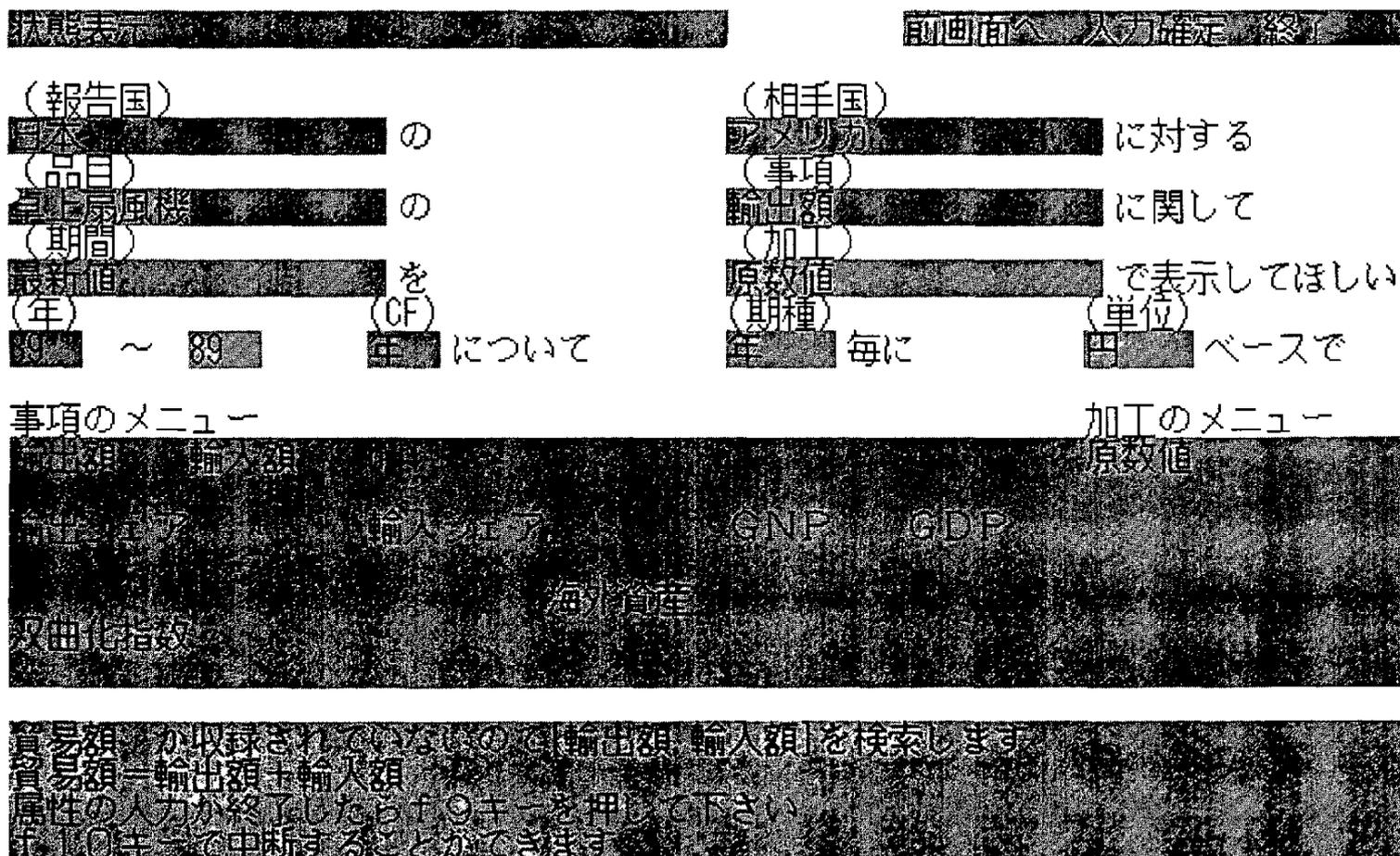


図2-12 下位項目の積み上げによる次善情報の生成

(3) 検索コマンドの生成に伴うメッセージ表示

検索コマンドを生成する時点で、卓上扇風機に関する上位項目を導出辞書により探索し、扇風機という上位項目とそのコードを得る。

ユーザに次善情報を利用する旨のメッセージを表示する。

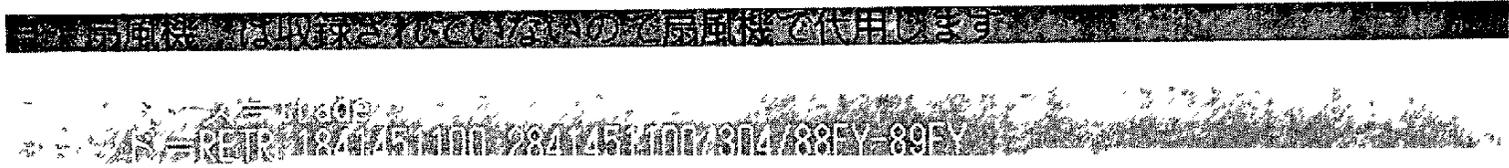


図 2-13 上位項目による次善情報の生成および検索コマンド表示

(4) 品目に厨房機器を指定した場合

類義語辞書を探索し、対応するコードが存在しないので導出辞書を探索し、下位項目として電子レンジ、電気釜・・・を得る。

ユーザに改善情報を利用した旨のメッセージを表示する。

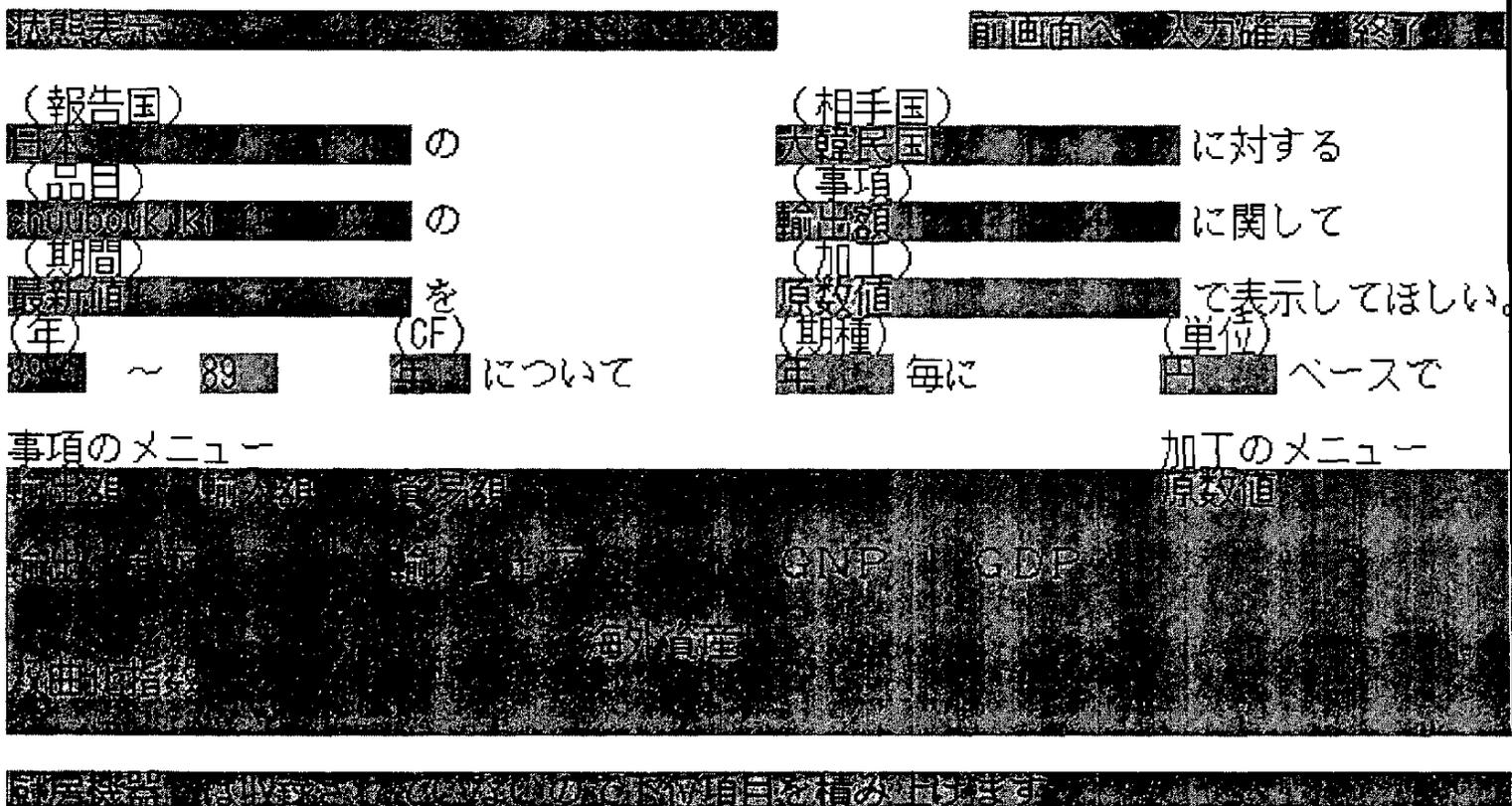


図2-14 下位項目の積み上げによる改善情報生成

(5) 事項にGNPを指定した場合

事項にGNPが指定された時点で、代替データ辞書を探索し、GDPを発見する。

ユーザに次善情報を利用する旨メッセージを表示し、GNPとGDPの両方を検索する検索コマンドを生成する。

(この種のデータに関しては、欠測が多く、検索失敗の場合、通信機能の関係で、ホストとの対話が困難な為、予め可能な代替情報も検索する方針をとった)

(相手国) 世界 に対する

(事項) 輸出額 に関して

(加工) 原数値 で表示してほしい。

(期間種) 年 毎に

(単位) 千ベースで

(報告用) の

(項目) の

(期間) を

(年) 89 について

事項のメニュー

加工のメニュー

検索結果

相手国	事項	加工	期間種	単位
世界	輸出額	原数値	年	千
			1989	

図 2 - 1 5 概念的に近い情報による次善情報の生成

2. 4 知識ベースの構造と働き

2. 4. 1 IRS汎用分類

ユーザが判りやすい言葉で検索要求を定義するために、統計上の分類とは別に'IRS汎用分類'を定義する。

IRS汎用分類は、統計の種類に関係ない標準的な分類として用意されるもので、将来取り扱う統計の種類を増やす場合は、IRS汎用分類と新規追加される統計との対応テーブルを追加する形で、システムの拡張を行うことになる。

IRSのユーザは、取り扱う統計やデータベースかどのように変化しても、要求入力の際に、意識する分類はIRS汎用分類1種類だけでよいことになる。

個々のユーザが独自に自分の呼び方や分類を定義したい場合は、ユーザ類義語辞書（IRS汎用分類の各ノードに自分の呼び名をつける場合）や、ユーザ導出辞書（IRS汎用分類と異なる下位構成を持つノードを定義したい場合）を使って、独自の分類体系の世界を定義することができる。本年度のIRS汎用分類は、生産動態統計を基に作成した。

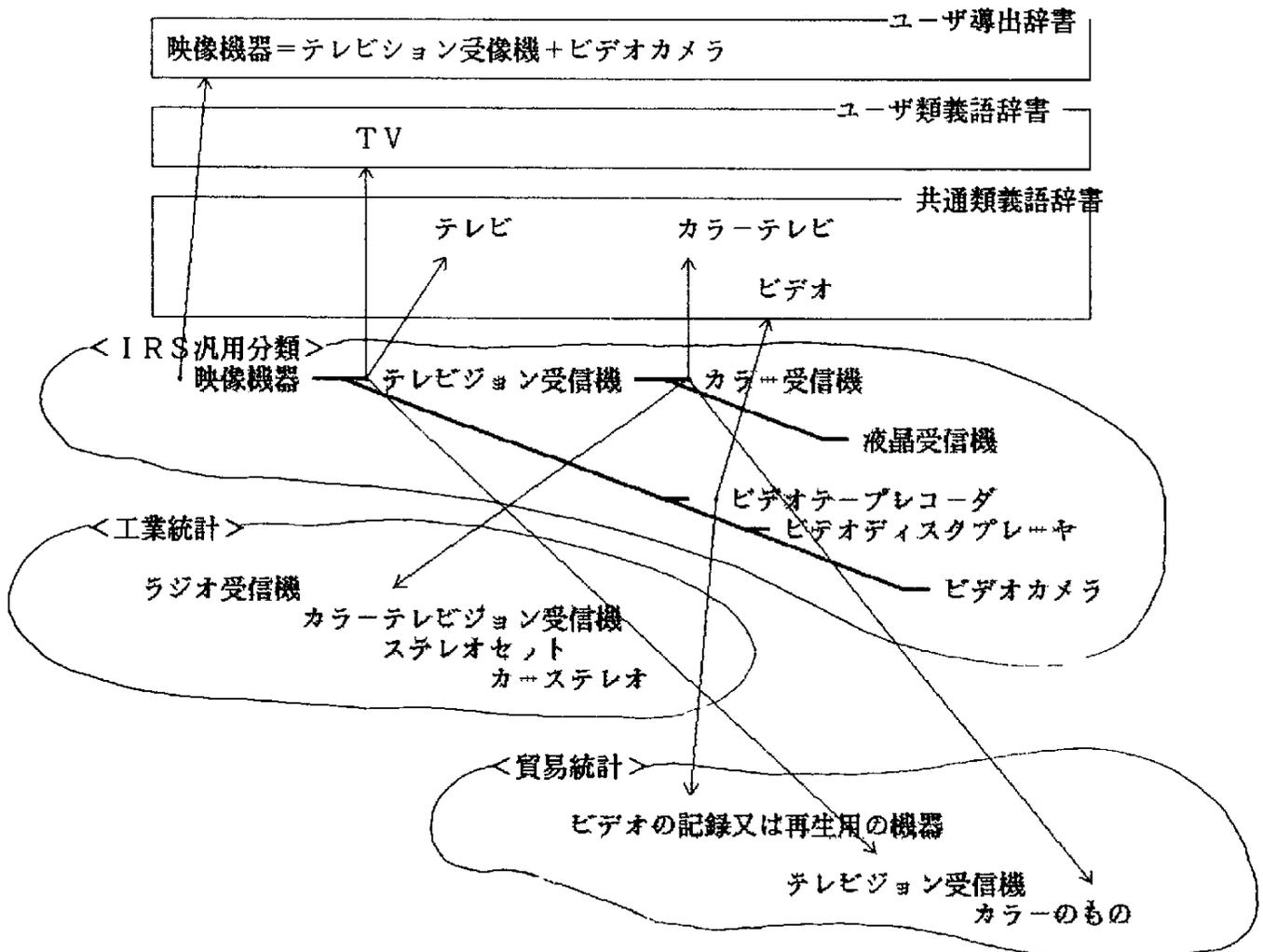


図2-16 IRS汎用分類と統計分類との関係および各辞書との関係

2. 4. 2 類義語辞書

類義語辞書は、国名と品目に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。

辞書はPROLOGのデータベースの形式に蓄えられている。

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)
 - 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)

例

```
dic(r_c,[シンガポール],[shingaporu],1,999).
dic(r_c,[中近東],[kankoku],1,999).
dic(r_c,全世界,[world],1,[1,1,800]).
dic(r_c,アジア州,[アジア,asia],1,[505,505,801]).
dic(r_c,ヨーロッパ州,[ヨーロッパ,europe],1,[170,170,802]).
dic(r_c,北アメリカ州,[北アメリカ],1,[*,*,803]).
dic(r_c,南アメリカ州,[南アメリカ],1,[*,*,804]).
dic(r_c,西ヨーロッパ州,[西ヨーロッパ],1,[*,*,805]).
dic(r_c,中近東,[],1,[405,405,811]).
dic(r_c,アフリカ,[],1,[605,605,812]).
dic(r_c,東南アジア,[],1,[*,*,813]).
dic(r_c,A S E A N,[],1,[*,*,814]).
dic(r_c,アジアN I E S,[],1,[*,*,815]).
dic(r_c,ラテンアメリカ,[],1,[210,*,816]).
dic(r_c,共産圏,[],1,[978,978,817]).
dic(r_c,発展途上国,[],1,[200,*,818]).
dic(r_c,日本,[japan,nihon],1,[158,158,*]).
dic(r_c,大韓民国,[韓国,korea,r_korea,daikanminnkoku],1,[542,542,103]).
dic(r_c,北朝鮮,[n_korea,kitachousen],1,[*,954,104]).
```

```
dic(goods,[電気温水機],[onsuiki],1,999).
dic(goods,電気暖房器,[danbouki,denkidanbouki],1,999).
dic(goods,電気ストーブ,[sutobu,onpuuki,denkisutobu],2,[$1851629100$, $28516
]).
dic(goods,電気こたつ,[denkikotatsu],2,999).
dic(goods,電気毛布,[denkimouhu],2,[$1630110$, $2630110$]).
dic(goods,電気カーペット,[denkikapetto],2,999).
dic(goods,厨房機器,[daidokoroyouhin,daidokoroyoukiki,chuuboukiki],1,999).
dic(goods,電子レンジ,[denshirenji],2,[$1851650$, $2851050$]).
dic(goods,電磁調理器,[denjichouriki],2,999).
dic(goods,電気こんろ,[denkikonro],2,999).
dic(goods,電気がま,[suihanki,suihanjya,denkigama],2,[$1851679100$, $2851679
```

2. 4. 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。

例

```
red(rezouko,[電気冷蔵庫],$$).
red(onsuiki,[電気温水機],$$).
red(sauji,[サウジアラビア],$$).
red(電気暖房器,[電気ストーブ,電気こたつ,電気毛布,電気カーペット],$$)
red(厨房機器,[電子レンジ,電気かま,トースター,電気オーブン,電気冷蔵庫],$$)
red(空調・住宅関連機器,[扇風機,換気扇,電気温水器,家庭用電気井戸ポンプ],$$)
red(扇風機,[卓上扇風機,その他の扇風機],$$)
red(衣料衛生関連機器,[電気洗濯機,洗濯物乾燥器,電気アイロン,電気掃除機,ハンドクリ-
red(理美容機器,[電気かみそり,電気マシ-ン器具,ヘアードライヤー],$$)
red(輸出依存度,[輸出額,GNP],$輸出依存度=輸出額/GNP*100$)
red(輸入依存度,[輸入額,GNP],$輸入依存度=輸入額/GNP*100$)
red(貿易依存度,[貿易収支,GNP],$貿易依存度=貿易収支/GNP*100$)
red(貿易量,[輸出量,輸入量],$貿易量=輸出量+輸入量$)
red(貿易額,[輸出額,輸入額],$貿易額=輸出額+輸入額$)
red(貿易収支,[輸出額,輸入額],$貿易収支=輸出額-輸入額$)
```

2. 4 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) .

2. 4. 5 データベース選択規則

データベース選択規則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

ch12 (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。

例

```
db_sel([日本,_,_,[輸出額,輸入額,輸出量,輸入量],_,_),same,trade,
'tradeで検索できます')
db_sel([_,_,全品目,[輸出額,輸入額,輸出ノエア,輸入ノエア,双極化指数],_,_),sam
dot,
'dotで検索できます')
db_sel([_,全世界,全品目,[GNP,GDP,為替レート,消費者物価,外貨準備],_,_),
same,ifs,
'多分ifsで検索できます')
db_sel([_,日本,_,輸出額,_,_),
ch12(輸入額),trade,
'報告国と相手国を入れ換えるとtradeで検索できます')
db_sel([_,日本,_,輸入額,_,_),
ch12(輸出額),trade,
'報告国と相手国を入れ換えるとtradeで検索できます')
```

2. 4. 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。

例

```
dbd(trade,[r_c([日本]),d_c,cond,goods,kind],  
     [輸出額,輸入額],[60,88],[年,年度],[年,月],[円,ドル]).  
dbd(dot,[r_c,d_c,cond,goods([全品目]),kind],  
     [双極化指数],[70,89],[年],[年,四半期,月],[ ]).  
dbd(ifs,[r_c,d_c([全世界]),cond,goods([全品目]),kind],  
     [GNP,為替レート],[55,89],[年],[年,四半期,月],[ ]).
```

2. 4. 7 事例ベース概要

(1) 事例の形式

試作システムは、PROLOG を使って開発し、事例は、PROLOG のデータベースとして下記の例に示す形式で蓄えられる。

```
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 ],  
      4  
      [ $RETR 317010503/304/80FY-82FY$, $DISP V3$ ],  
      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：3
- 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：2
- 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。

(2) 事例ベースの参照と更新の仕組み

① 事例の適用

事例の適用は次の3段階になっている。

- i データベースに関する知識を参照するまえに、事例ベースを検索し、同じ検索要求が登録されていれば、事例として蓄えられた検索コマンドを利用する。
- ii 同じ検索要求がないとき、「若干の修正」(次節参照)を加えれば適用できる事例を検索し、これを適用する。
- iii 「若干の修正」による適用が不可能な場合、部分的に別の事例を合成すれば適用できるかどうかを検証し、できればこれを適用する。

② 事例の修正

①で述べた事例の修正には、次のものがある。

i 「若干の修正」

品目コード、国コード、系列コードを類義語辞書を使って、検索要求に揃える。

ii 複数の事例の合成

類義語の範囲で、検索要求と一致するものが無い場合、属性指標から基礎となる事例を決定し、他の事例のなかから類義語に相当するものを発見し適用する。

(例)

検索要求 米国から西側先進国への一般機械の輸出額

事例1 米国から日本への電気製品の輸出額 (検索要求と属性指標が一致する事例)

事例2 西側先進国のGNP

(西側先進国という用語を使った事例)

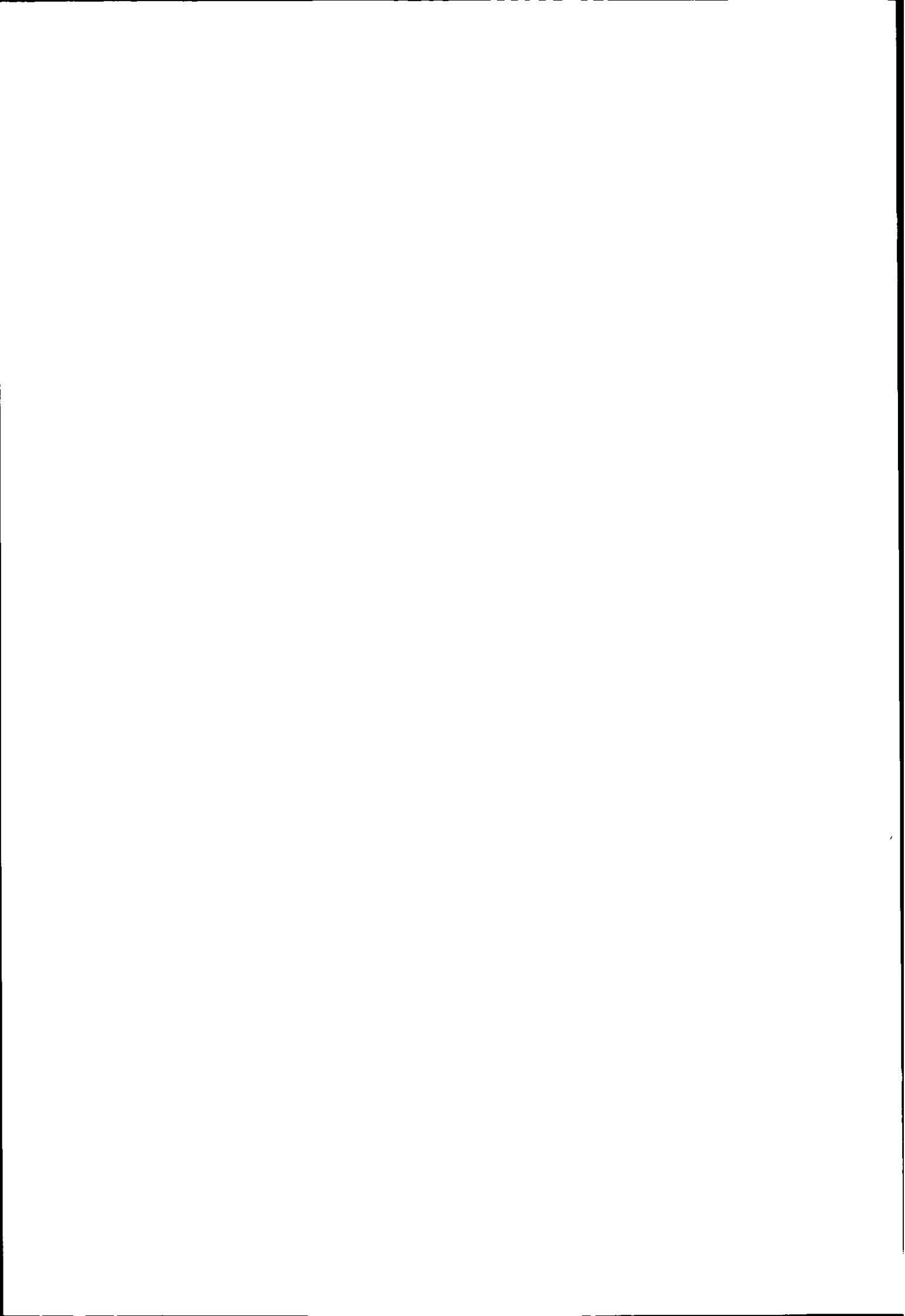
→ 類義語辞書には、'西側先進国' という用語は、存在しないが、'先進国' という用語を'西側先進国' の代わりに使って検索を成功させた事例

③ 事例の登録

事例の登録は、下記のタイミングでおこなわれる。

- I 事例の修正によって検索が成功した場合は、新たな事例としての登録を自動的に行う。
- II 事例の適用が失敗した場合は、一般的な問題解決の方法により、検索コマンドの生成を試み、成功すれば事例として登録する。
- III 一般的な問題解決の方法でも失敗した場合は、ユーザに適切な検索コマンドを入力させこれを事例として登録する。

第3章 第2期試作システムの評価



第3章 第2期試作システムの評価

本年度は3年間の調査研究プロジェクトの最終年度に当たる。本節では、本年度に作成した試作システムを評価する。また、試作システムの開発によって明らかになった、実用化のための検討項目について述べる。

評価にあたっては、初年度に設定した調査研究の目的を念頭におき、3.1では研究目的ならびに要求仕様を再確認し、3.2では、試作システムについて述べる。次いで3.3では、システムの評価を行い、最後に3.4で特に知識ベースに関してその有効性を評価する。

3.1 システムに対する要求とその背景

要求仕様と実現されたシステムを比較する前に、要求仕様の背景を再確認するために、初年度の報告書より、調査研究の目的に相当する部分を整理して再録する。

3.1.1 一般的なデータベース利用の現状と課題

先ずシステムに対する要件の背景としては、一般的なデータベース全体に対する要件と経済統計データベースに特有な要件が存在する。

以下にデータベース一般の問題点、経済統計データベースの問題点を明らかにし、本調査研究の対象システムに関する要件を整理する。

(1) データベース構築と利用の経緯

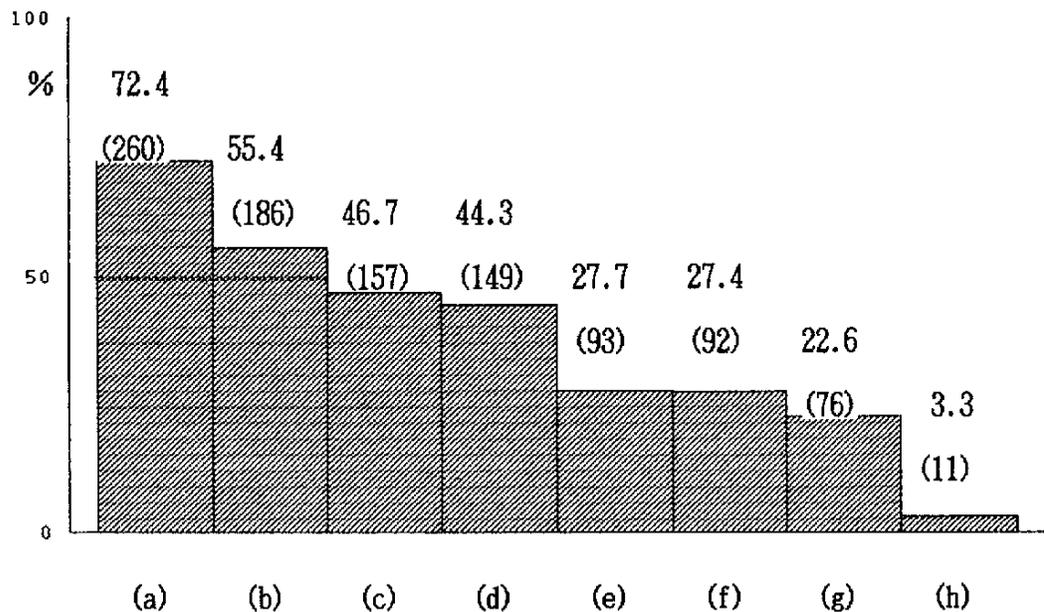
我が国のデータベースの構築とその利用は、1950年代より、文献情報を中心として開始され、現在では統計情報（数値情報）の利用もかなり一般化してきている。しかし、さまざまなデータベースが構築されるにつれ、その利用における問題点が多く発生してきていることも事実である。

データベースシステム自体は、データを追加更新しながら、長期間に渡って運用されるものであるのに対して、データの利用の状況は、より短期間に流動的に変化しているという事実が、これらの問題の多くを引き起こしていると思われる。

今後、データベースの利用を推進してゆくためには、データベースシステム自体の再検討と共にその利用技術の改善が重要となってきている。

(2) 一般的なデータベース利用における課題

図3-1は、データベースの利用上の問題点について、特に機能・操作性に関する指摘について示したものだが、この図からわかるように、指摘が一番多かったのは、「シソーラスやキーワードの標準化や拡充不足」が全体の77.4%、また「検索機能が不十分」についても55.4%と高い比率を示している。



- (a) シソーラスやキーワードの標準化や充実の不足
- (b) 検索機能が不十分
- (c) 検索速度が遅い
- (d) 接続手順が煩わしい
- (e) ファンクションキーが不足
- (f) 交換機能 (カナ漢字等) が不足
- (g) キーボード配列が悪い
- (h) その他

図3-1 機能・操作性に関する指摘 (N=336, 複数回答)

出典: DPC「データベース・サービスに関するユーザ意識調査」昭和63年3月

また、データベースの利用上の問題点としては、「検索のコマンドがシステムごとに、異なり、統一されていないために利用しにくい」といった点について、以下のような調査結果がでている。

- ・マニュアルを見て不自由ながら検索している・・・66.1%(225社)
- ・不自由は感しない・・・15.4%(57社)

- ・検索する気がしない・・・・・・・・・・・・・・・・・・10.8%(40社)
- ・その他・・・・・・・・・・・・・・・・・・7.6%(28社)

出典：同上

この調査からわかるように、「データベースを検索する気がしない」と「マニュアルを見て不自由ながら検索している」を合わせると、76.9%のユーザがコマンドの不統一に対して何らかの不満を持っているといえる。一方、「不自由は感じられない」が15%という結果を示しているが、これは、この問題が検索の専門家にとってはそれ程大きな問題になっていないためである。しかし、この比率は、ユーザ層が広がることによって、低くなる可能性を内包している。

本調査は一般的な調査ではないかユーザから見た利用上の問題点が反映されていると思われる。

(3) 一般的なデータベース利用に対する要求

このような利用上の不満が生してきた背景として、以下のような理由が挙げられる。

- ①データベースの利用者層が広がり、データベースシステムに対する要件が、構築時に想定した範囲を越えて拡張している。
- ②利用者が必要とする情報の詳細性のレベルと利用者のデータベースに関する知識のレベルに差がある。

①への現在の対応は、人海戦術的なものであり、「どのようなデータが必要で、どのようなものをデータベース化すればよいか」に関する検討はされても、利用者の常識のレベル、目的の違いなどの付随情報に関しては対応できていない。

また、②についても、特に経済統計データベースに見られるように、利用者の「常識」は、千差万別であり、初心者利用に耐えられるようなデータベースの構築が難しいのが現状である。また、専門家による各種データベースに跨がる横断的な利用など、高度な利用にも耐えられない。

このような状況を打開するひとつの方法として、データベースの利用者をサポートするシステムの構築が考えられる。即ち、利用者に不足している「知識」をシステム側で補うことによって、データベースの利用価値を高めようとする考え方である。この「知識」のシステム化についての技術的な実現方法として、今日、様々な分野で利用が進められている人工知能技術の適用が考えられる。特に、知識ベース (Knowledge Base) による「専門

的知識」のシステム化を可能にするエキスパートシステム (Expert System)の適用が期待される。

3. 1. 2 経済統計データベースの特徴と人工知能技術への期待

データベース利用の高度化について、本調査研究では、特に「経済統計データベース」の利用の高度化についての検討を行った。ここで、経済統計データベースを取り上げた理由としては、以下の2点が挙げられる。

①経済統計データベースの利用は、他のデータベースの利用と比較し、利用の適否に関する判断を要する場面が多く、適切な情報検索を行うためには、広範な専門的知識が必要とされている。

②経済統計及び経済情報に関するデータベースの利用が拡大され、その利用者層が多様化することによって、専門的な知識を持たない利用者に対するデータの所在や出力されるデータの意味等に関する知識のサポートが重要となってきた。

経済情報に関するデータベースの課題について以下に詳述する。

(1) 経済統計データベースの課題

経済統計データベースの特徴として

①利用可能な統計データの案内については、若干の印刷物があるものの極めて不備であり、類似統計、関連統計、相互の相違等の比較は専ら利用者の経験に依存していること。

②統計データに関する書誌情報や定義情報などのいわゆるメタデータが、データベース化の際に、統計データ自体と分離されている為に、データの選択にあたり困難が生ずること。

③統計データに不可欠な時系列データとクロスセクションデータを総合的に管理する必要があること。

④種々の統計を横断的及び学際的に利用するニーズが高いこと等が挙げられる。

このように経済統計データベースの利用には、統計及びデータベースに関する専門的な知識が必要であり、習熟に時間がかかる。これらの特徴をふまえ、蓄積された統計情報を活用促進するためには、以上の知識を整理し、専門知識のない利用者を支援する機能を構築することか課題である。

(2)経済統計データベース検索支援への人工知能技術適用の期待

経済統計データベース利用の課題として、「専門知識のシステム化」が対応策として挙げられたが、知識のシステム化を効率的に行うための手段として人工知能技術の適用を試

みることが本調査研究のアプローチである。

人工知能技術の利用に対する期待として次の各項目が挙げられた。

- ①過去に利用されたノウハウを知識ベースとして活用したい。
- ②データベースのデータは、数値の羅列であり、その意味を読み取ることは難しいことから、本にあるような庄積的なものがデータベースにも欲しい。
- ③データの分析のためのエコノミストとしての常識（基礎知識）がデータベースにはっていると便利である。
- ④データベースを利用する際、使用例が利用情報として参照できるようなシステムがあると良い。
- ⑤利用者が、システムからのコンサルテーションを受けながら利用できるデータベースがあると良い。

3. 1. 3 経済統計データベース検索支援システムへの要求

3.1.2 で述べた経済統計データベース利用の特徴をふまえ、具体的な現状分析を行った結果は以下の通りである。

(1). 経済統計データベース利用の現状

ここでは、経済統計データベースの利用者の傾向について触れ、更に人工知能技術の適用範囲について説明を行う。

①経済統計データベース利用の現状

データベース利用の傾向

- ・データベース及び経済統計に関して、素人が利用するケースが増えている。
(例. 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)
- ・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。
(例. データのグラフ化、外部情報と自社情報との統合加工等の要求が増えた)

このように、データベースの利用の様相が、普及と高度化の2つの方向に向かっていることがわかる。

次ページに挙げたデータベース利用の問題点にも、これらの傾向が反映されている。

データベース利用上の問題点

- ・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、検索と内容

理解がしにくい。

- ・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）
- ・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。
- ・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。
- ・欠測データに関する情報が判りにくい。
- ・異なるデータヘースからの情報の合成等高度な加工がしにくい。

②A I 技術の適用性について

上記のようなデータベース利用上の問題点を解決する方法の一つとして、A I 技術の適用可能性についての検討を行った。データヘース検索へのA I 利用の例が初年度報告書第3章に報告されている。

これらをもとにA I の適用可能局面をまとめると下記のようになる。

A I の適用可能局面

- ・ユーザの要求に対する適切なアドバイス機能
- ・経済統計の専門的知識の知識ヘース化による検索システムの高度化
- ・統計スキーマの知識ベース化による柔軟なインターフェイスの構築

更に、A I 技術の適用範囲を明確化するために、利用者が必要とする知識を分離し、図式化すると次頁のような3つの世界が想定される。利用者は、「ユーザーの世界」から「データベースの世界」に直接アクセスして、データを取り出しているように見えるが、「ユーザーの世界」での情報表現を「統計の世界」での情報表現に替えるための知識、「統計の世界」の情報表現を「データベースの世界」での情報表現に替えるための知識がなければ正確な情報検索はできない。

本調査研究では、特に「ユーザーの世界」から「統計の世界」を通して「データベースの世界」への橋渡しをスムーズに行う為に、A I 技術の利用を試みる事が中心的検討課題となった。

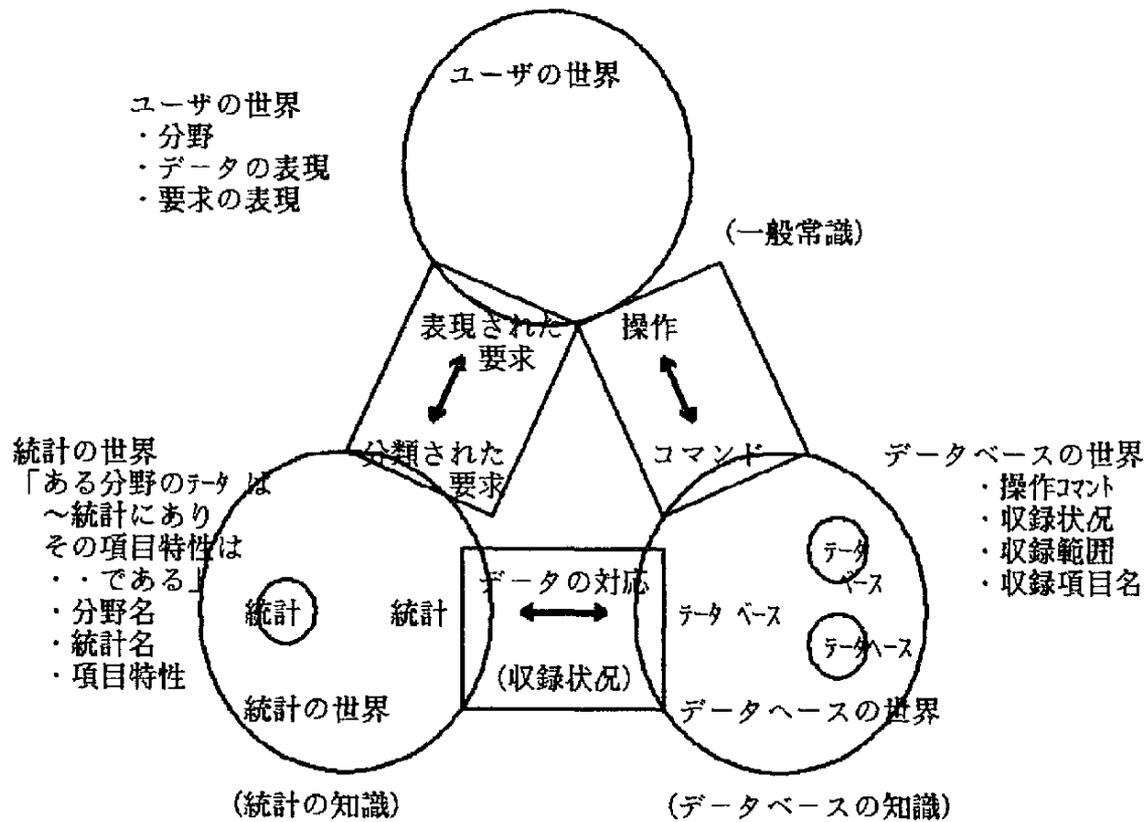


図3-2 経済統計DBとAIの関係図

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表現される。

「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。

「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。

ユーザが、独力で（専門家の手助けなしに）的確にデータベースを検索するためには、上述の知識をAI技術を使って知識ベース化し、ユーザの世界から直接データベースをアクセスするシステムを開発しなければならない。

3. 2 第2期試作システムの限界とその要因

本節では、3. 1 に説明したシステムに対する要件に対して、試作システムが、どの程度の出来ばえであったかを概観する。また、実現されなかった要件に対しては、その原因を分析する。

本節では、要求の背景をふまえた上流レベルでの要求に対して、第2期試作システムの限界を技術的に分析する。

3. 2. 1 要件の整理

先ず、3. 1 で述べた検索支援システムに対する要求を、背景、要件、人工知能技術への期待の3項目によって整理する。

表3-1 検索支援システムに対する要求

	背景	要件	人工知能技術への期待
一般的な問題	<p>利用者側に立ったデータベースシステム作りが難しい</p> <p>利用者が必要とする情報の詳細性のレベルと利用者のデータベースに関する知識のレベルに差がある。</p>	<p>利用者に不足する「知識」のシステム化。</p> <p>ユーザの入力に対する柔軟な対応。 説明機能の充実。</p>	<p>データ利用に関するコンサルテーション機能の付加。</p>
経済統計データベースの問題として	<p>所在情報や関連統計、類似統計に関する情報がない。</p> <p>書誌情報や定義情報がデータベースには収録されない</p> <p>時系列データとクロスセクションデータの総合管理が難しい。</p> <p>統計情報の横断的な利用に対するニーズが高い。</p> <p>欠測データに関する情報が判りにくい。</p> <p>統計データの更新に伴い同種の検索と引用を反復することが多い。</p> <p>データベース検索コマンドを覚える必要がある。</p>	<p>経済統計に関するガイダンス機能（データベースマニュアルの逆引き機能・・・欲しい情報に関して、収録データベースや類似情報の案内を行う機能） データに関する利用上の注意の案内。 複数系列の総合的な利用（異なるデータベースからのデータも含めた複数系列）</p> <p>次善の情報の提供。</p> <p>検索手続きの簡便化。</p> <p>コマンドを知らなくて済むユーザインタフェース</p>	<p>データベース選定に関する情報の知識ベース化</p> <p>データに関する注釈的な情報の知識ベース化。</p> <p>データ分析のための知識を知識ベース化したい。</p> <p>レポート作成過程で、過去に利用されたノウハウを知識ベースとして活用したい。</p> <p>従来の使用例を知識ベース化したい</p> <p>コマンド生成機能</p>

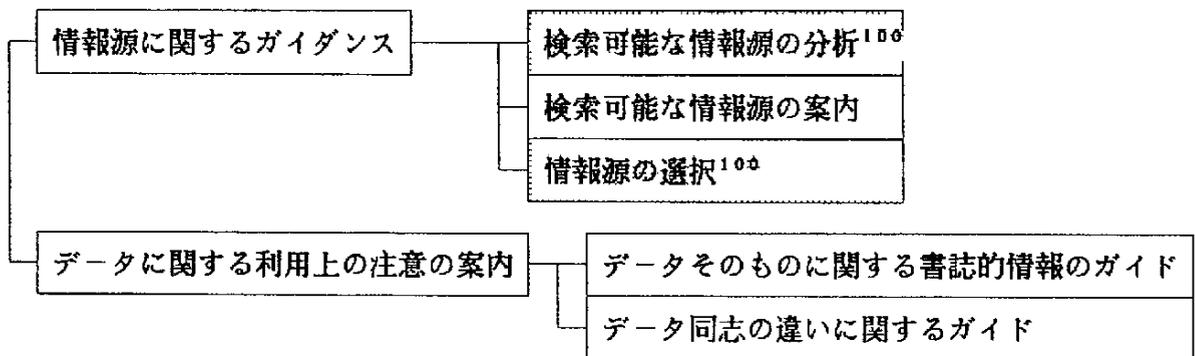
各種の分類体系が設定されている。	分類コードを調べなくて済む ユーザインタフェース	分類体系を知識ベース化したい。
------------------	-----------------------------	-----------------

表3-1より、一般的な問題として挙げられた項目は、経済統計データベースの問題を包含する'情報検索'の問題となっている。経済統計データベースに関する問題は、数値データベースであり、数値の意味に関する情報と切り離されてデータベース化されている点や、情報源を異にする類似データが多い点、データが時系列として収録されている点等経済統計上の特徴に起因する問題であると解釈できる。また人工知能技術に期待する事の本質が、コンサルティングの機能であり専門家によるアドバイスと同等の支援を各種の知識ベースを使って行うシステムの実現であることを認識して、次節以降に経済統計データベースの問題に対するシステム化の要件として挙げられた各項目について、実験システムの評価を行う。

3. 2. 2 経済統計に関するガイダンス機能及びデータに関する利用上の注意に関する案内について

この項目は、専門家の広範囲な知識を使った支援機能に関わる部分である。専門家のアドバイス機能のうちどの範囲をサポートすることができたか、また実用化のための技術的な問題点がどこにあるかを論じる。

(1) 要件の範囲とその実現の範囲の関係



注. 網かけ部分は試作システムで実現した機能
右肩の数字は実現した機能に関する充実の度合い (%)

試作システムでは、検索要求に対して、検索可能なデータベースを探索し、複数のデータベースが検索可能な場合は、ユーザにデータベース名を入力させて検索すべきデータベースを決定している。

実際は、専門家でないユーザは、少なくともデータベース間の違いが示されなければ、複数の候補の中から希望のデータベースを選択することはできないと思われる。

(2) 実用化のための技術的な問題

試作システムでは、データの書誌情報のなかで、データベースの選択上必要最低限の情報のみをシステム化（知識ベース化）した。

現在はデータベースが3種類なので、それほど多くの知識を必要としないで、データベースの選択をおこなうことができる。検索対象データベースの種類が増えれば、検索可能なデータベースに優先順位をつけるためさらに多くの知識が必要になると思われるがデータベースのふり分け自体は実用化に際しても問題はない。

データベースに関する利用上の注意という意味での書誌情報に関しては、実用化の際の問題が大きいと思われる。

書誌情報のなかには、次のような情報がある。

①統計単位の情報・・・調査方法、収録期間、調査対象、推定方法等

②個々のデータあるいは系列に関する情報・・・データあるいは系列に関する特記情報（独特の集計をしている、ある時点から分類方法や作成方法が変わった等）

①も②もノウハウというよりは辞書に近い知識であるが、①は情報量としては限られた量であり、比較的継続性が保証されている。

これに対して②は、国や品目のレベルの情報量であり、データベース本体に匹敵する大きさである。しかも、継続性のないものが多く、内容も統一的な記述が不可能である（文章による説明にならざるを得ない）。

②の情報に関しては、データベース化も知識ベース化も困難な情報であるといわざるを得ない。敢えてシステム化しようとする、統計の解説書の注釈をデータ化することになるが、試作システムにおける端末機の画面の制約から、解説書を読むよりも便利になるかどうかの保証が難しい。

実際は、②のなかの重要な項目に関してはデータベース側でユーザにメッセージを表示

し、後の情報は、解説書に頼っているのが現状である。

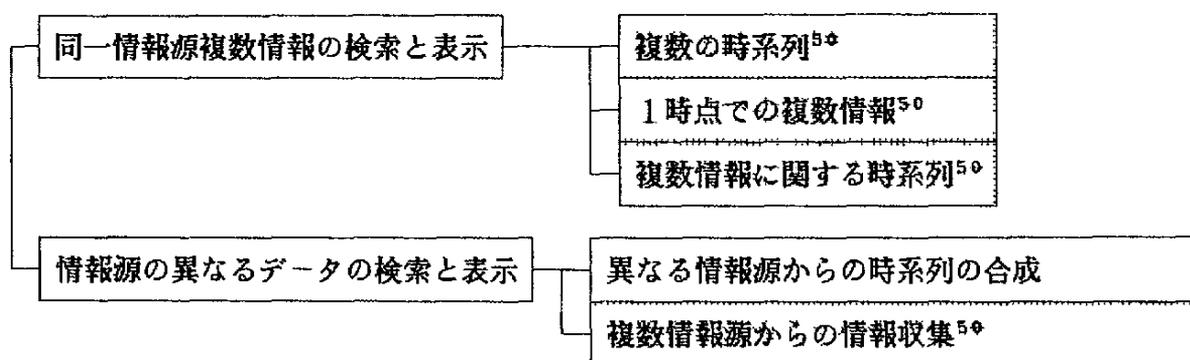
実用化の為の方策を検討すると、②の情報の全てをデータ化することは避け、利用事例または、利用事例とともに記録されたユーザのコメントとして、過去に利用されたものに関する対処の仕方をデータベース化することが考えられる。

また①の情報に関しても、各統計に関する辞書的な知識の他に、複数の統計の比較に関するノウハウ的な知識を専門家から抽出することがユーザの支援に役立つものと思われる。

3. 2. 3 複数系列の総合的な利用について

この項目は、特に経済統計データベース利用の特徴を示す時系列とクロスセクションデータの比較や合成に関わる支援である。支援機能としての意味は、①同一インタフェースから複数の情報源にアクセスができること、②複数の時系列やクロスセクションの比較が可能であること、③複数の情報源を使って新たな情報を合成すること 等である。

(1) 要件の範囲とその実現の範囲の関係



注. 網かけ部分は試作システムで実現した機能
右肩の数字は実現した機能に関する充実の度合い (%)

試作システムでは、検索したデータの表示がホストの機能に任されているため、複数の情報を検索すること自体はできても、同一画面で、数値を比較することができない。(履歴ファイルを使って別処理でレポートを作成する必要がある) ホストの機能により、同一情報源ならば、同一画面上に表形式で表示することはできるが、異なる情報源は、異なるページに表示されてしまう。

また、比較に際しては、データの比較可能性に関する吟味が必要であるが、試作システムは比較可能性に関する知識を持っていない。このため異なる情報源からの時系列を組み合わせ、新たな時系列を合成すること(データに適切な補正を加える必要がある)は

できない。

同一時点で、複数の情報源からの情報を比較する場合も、情報源の違いによる分類体系の違いなどを考慮して、ユーザにアドバイスを行う必要があるが、いまのところ、この種の知識は考慮されていない。(検索対象としている3件のデータベースのうち、商品分類が問題になるものは1件しかないため取りあえず問題になっていない)

(2) 実用化のための技術的な問題

技術的な問題としては、

- ① 検索した情報を同一画面上に見やすく表示するための表示技術
- ② 異なる分類体系に関する対応表の作成
- ③ 異なる統計間の類似情報に関して、数値の違いを説明する技術

の3つが挙げられる。

この内①は、マルチウィンドウの利用による複数系列の同時表示や、マルチタスク機能による通信処理とその他のデータ処理の同時実行により、ハードウェア的な問題として解決すべき(また可能な)問題であると思われる。

②については、国名と商品の分類が対象になるが、問題となるのは商品分類である。試作システムの解決策としては、産業連関分類を標準分類とした対応表の作成と、このシステム化が考えられる。

産業連関分類は、国内の経済活動に関して、網羅的な情報として作成されており、すでに、貿易統計、工業統計をはじめ各種統計との対応表が作成されているためである。

但し、産業連関分類は、商品のレベルではかなり集計度の高い分類になっているため、詳細対応表を別途持つ必要がある。また、対応表が作成できたとしても、各統計の分類精神の違いを把握した上で、これを利用する必要がある。(例えば産業連関分類は、同一企業内の生産物を産業別に分類して集計しているのに対して、工業統計は、全生産高をその企業を代表する産業の生産高として集計している) システムとしては、分類精神の違いによる分類項目の違いを、ユーザに伝える機能が必要になる。

③は、目的とする時系列を得るために、複数の情報源からのデータを合成する場合を想定している。目的とする時系列に商品分類が関係する場合補正は難しく、不可能な場合が多い。輸出総額のような商品に関係ない情報の場合は、補正係数が算定可能な場合が多い。

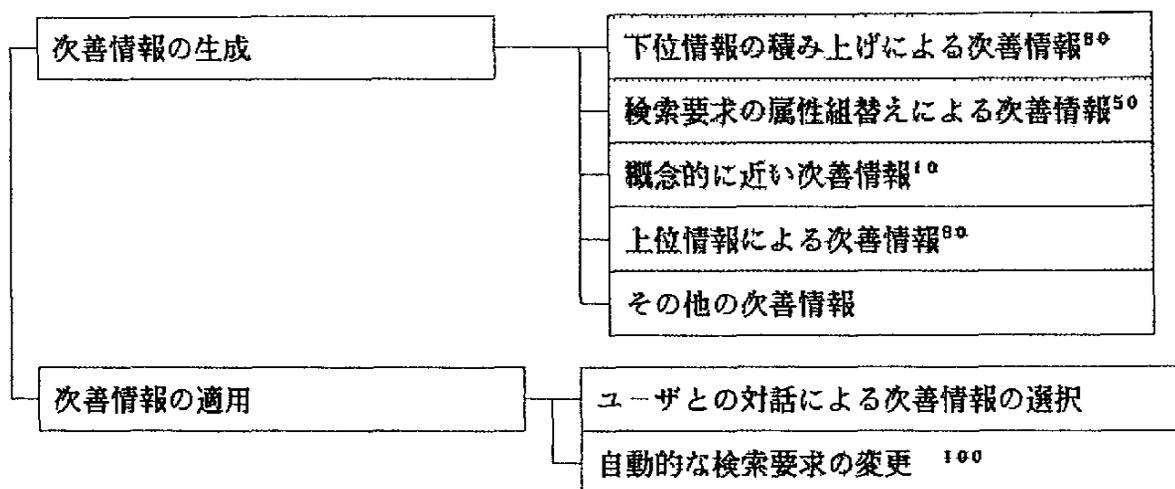
白書等での引用の実例を見ると、複数情報源からの情報を合成して作成したグラフや表

には、その旨の注記表示してあり、異なる情報源であることを明確に示せば、補正しない情報でも利用可能な場合が多い。

3 2. 4 次善情報の提供について

この項目は、ユーザの要求通りのデータが存在しない場合の代替データの提供に関する機能である。利用目的をほぼ達成することのできるデータを極力ユーザに提供するために、与えられた検索要求に対して何種類かの次善情報の提供を試みた。

(1) 要件の範囲とその実現の範囲の関係



注. 網かけ部分は試作システムで実現した機能
右肩の数字は実現した機能に関する充実の度合い (%)

試作システムでは、次善情報はシステムが自動的に生成した。実際はユーザに問い合わせを行うべきであろう。また概念的に近い情報に関しては殆どデータが集まらず、仕組みだけを整えた結果になった。

分類体系に関する次善情報については、商品分類に関して下位のを積み上げる次善情報と、上位のもので代替する次善情報の仕組みができた。商品分類自体が限定された範囲のものであるため、必要な知識ベースの規模は少なくて済んでいる。

また下位のを積み上げる次善情報としては、分類体系の他に、貿易額=輸出額+輸入額のように演算上の下位情報の積み上げも一部実現した。(実際は、演算まで行う機能

を検討していたが、ホストとの通信の関係で検索したデータを用いた演算が難しく、やむを得ず次善情報によって処理することになった)

検索要求の属性組替えによる次善情報は、報告国と相手国の入替えによるもの(例えばある国の輸出を輸出国側のデータで置き換える)のみを扱った。実際は、期種や期間、単位等についても次善情報の生成が可能である。

(2) 実用化のための技術的な問題

次善情報の種類として考えつくものを挙げてシステム化を試したが、次善情報の性質は流動的であり、利用とともに変化するものと思われる。事例ベースの利用等によって新たな次善情報を取り込む仕掛けを考慮する必要がある。

実用化に際して知識ベースの大きさという面では、下位項目の積み上げや、上位情報による次善情報は、品目分類に対応する部分を含んでいるので、品目分類中の階層関係に対応して知識ベースを作成する必要があるが、品目分類を記述した類義語辞書を利用することにより、辞書の大きさと更新の手順を低減することが可能である。

数値的に近いものに関する辞書にデータが集まらなかった理由は、近さの判断が利用目的によって異なるためと思われる。この種の知識に関しては事例ベース等の利用のほうが有効であろうと思われる。

次善情報適用の際のユーザの対する問い合わせは、まだ十分に検討が行われていないがユーザにとって次善情報が役に立つ場合とそうでない場合が考えられる。次善情報を適用して検索要求を検索可能な状態に変換していくプロセスのなかで、ユーザからの応答をどのように反映させるか(ユーザが否定した場合にすぐ検索を中止することは易しいが、次の候補をユーザに示す等の戦略を考慮すると、知識ベースの構造自体を再検討する必要性が生じる)、実際のユーザによる試用を繰り返して、ユーザにとって負担の少ない適切なインタフェースを考案することが重要である。

3. 3 要求仕様に対する実現方法の妥当性及び実現された機能の有効性

3. 3. 1 要求仕様に対する実現方法と評価

本節では、初年度に検討した理想的な検索支援システムに関する要求仕様をもとにシステムのレベルで評価を行う。

下記に要求仕様に対する実現方法及び評価の対応表を掲げる。

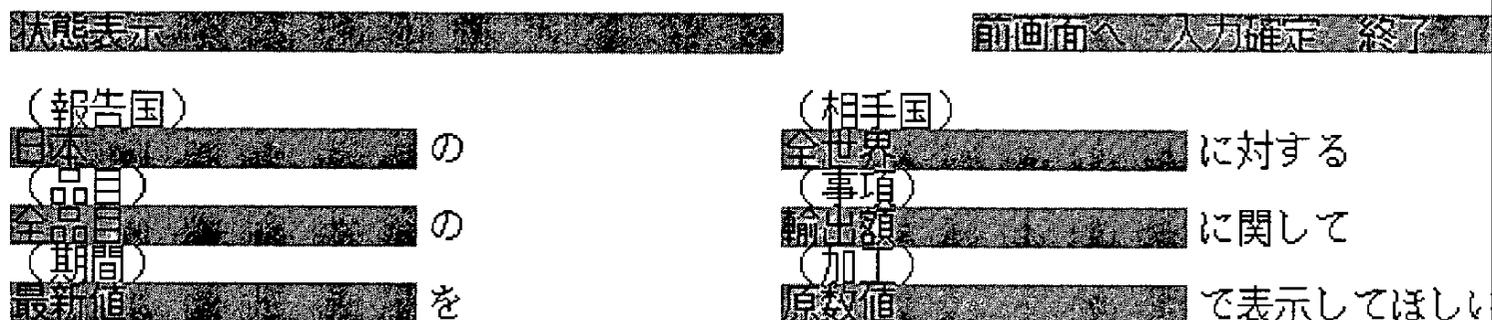
	要求仕様	実現方法	評価		
1	日本語による検索要求を受け付ける機能	テンプレート型入力画面 (マウスによる選択方法を、マウスを使わない方法に変更)	①検索要求入力 <small>の柔軟性</small> という面から捕らえると、検索コマンド入力と文章入力 <small>の中間に位置する方法</small> といえる。ユーザにとっての <input/> し易さについて、各方法を比較すると下表のようになる。 <他の方法との比較>		
			方法	柔軟性	ユーザ適性と問題点
			DB指定+コマンド入力	小	DBに習熟したユーザは正確にデータを入手できる。よりよいデータを見逃す危険がある。
			テンプレート	↓	テンプレートは、検索可能範囲に関するガイドの働きをしており、入力の負担も少ない。GNP等の場合多少の違和感がある。(相手国、品目等の属性が残るため)
			文章入力	大	検索可能範囲に入るように検索要求を考え、キーボード入力するのはユーザにとって負担が大きい。検索範囲が広くシステムが充分に、DB選択の機能を持てば、入力の制限はなくなるが、DBに詳しいユーザにとって、欲しいデータとシステムが選んだデータの乖離が気になることが予想される。(文章の多義性)
			②マウスを使うか否かは、好みの問題に属し、本質的ではないと思われる。		
2	不足情報の指摘や、検索条件等のガイダンス機能	テンプレート型入力画面	①ユーザの入力は、必ずデフォルト値によって補足される。またユーザにはデフォルト値がテンプレート上に見えているので、入力しない場合になが補足されるかが判る。 ②属性別にメニューによって指定できる範囲が示されるので、検索不可能な検索条件を入力することはない。		

3	要求した情報が存在しないときに代替の情報を提供する機能	類義語辞書 導出辞書 代替データ辞書 データベース選択規則	<p>①品目や国名に関して、ユーザの入力と同義の基本名詞を選択することができる。(類義語辞書)</p> <p>②品目や国名に関して、下位コードを積み上げることで代替情報を提供することができる。(導出辞書)</p> <p>③品目や国名に関して、上位コードを代替情報として提供することができる。(導出辞書)</p> <p>④数値的に近い情報を代替情報として提供することができる(代替データ辞書)</p> <p>⑤入力された属性を組み換えることによって、検索可能なデータベースを発見し、代替情報として提供することができる。(データベース選択規則)</p> <p>☆試作システムでは各辞書の内容が不足しているが、次善情報を生成する枠組み自体は専門家の助言に近いものになった。 ☆類義語辞書、導出辞書は、品目分類や、国名分類に対応するので大規模になる。他の辞書は小規模である。</p>
4	統計やデータベースの自動選定機能	データヘース選択規則 データベース辞書	<p>①複数のデータヘースが利用可能の場合、実験システムでは、ユーザに問い合わせを行う。(3種類なので頻繁にはおこならない)</p> <p>☆複数利用可能の場合、最適データを自動的に選定するだけの根拠は、抽出することが難しく、ユーザに決めさせるほうが現実的だと思われる。</p>
5	データベースの自動検索能	事例ベース コード生成プログラム 通信ソフト及び通信パラメータファイル	<p>①検索コードの生成には、事例ベースか、コード生成プログラムを使う。事例ベースの利用により、効率をあげている。</p> <p>②ホストとの交信は、自動的に行われるが、対話型の通信イメージにはなっていない。パソコンのメモリの制約や、通信方式が無手順であることが原因となっている。</p> <p>☆ホストがデータを持ち、端末機が入出力に関わる処理を受け持つという、分担方式は妥当だと思われる。端末機の選択、ホスト側プログラムの改造等が可能であれば、通信の効率向上は可能。</p>
6	検索履歴の保存と、連続検索支援機能	交信記録ファイルの保存 事例ベース	<p>①検索履歴は、画面に表示された検索結果をパソコン側のファイルに保存すること、また、事例ベースに検索要求と、生成されたコマンドを登録するという2種類の方法で実現している。</p> <p>☆検索結果を加工するためのデータは、保存していない。(ホストで、結果ファイルを作成し、端末側にダウンロードすることは可能)</p> <p>②連続検索は可能だが、前回の検索結果と今回の検索結果を比較することは不可能。</p> <p>☆端末機をワークステーションにすれば、分割ウィンドウに表示させたり、グラフ表示することが案にできるとと思われる)</p>
7	類義語の定義に関してユーザ定義を受け付ける機能	類義語の知識獲得機能	<p>品目と国名に関して、未登録の用語をユーザが入力した場合、必ず類義語辞書に登録される。</p> <p>☆運用上注意が必要となる。また全てを登録するよりもユーザに問い合わせが必要な場合のみ登録したほうが現実的。</p>

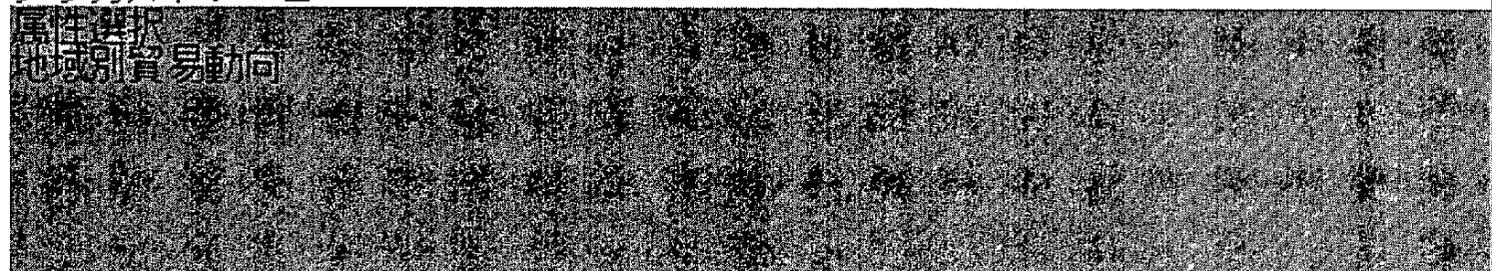
3. 3. 2 各機能の実現形態と実現方法に関する再検討

本節では、3. 3. 1 で、評価した要求仕様の各項目について、実現形態と実現方法をさらに詳しく解説し、要求仕様自体の見直しも含めて、その有効性を再検討する。

(1) 日本語による検索要求を受け付ける機能



アラカルトメニュー



アラカルトメニューを選択します。→キーで選択してリターンキーを押して下さい
国名や品目名、事項等を選択したい場合は、「属性選択」を指定して下さい
F10キーで中断することができます

図3-3 日本語による検索要求を受け付ける機能

背景には、ユーザの目的により近い形での入力を受け付けたいという要望がある。試作システムは、図3-3に示すような属性別の入力とアラカルトメニューによる入力の2種類の方法を用意した。これによって当初の要望を満たす枠組み自体は出来上がったが、属性入力で発生した検索要求を必要に応じてアラカルトメニューに登録するなど両者の連携を考慮する必要がある。

(2) 不足情報の指摘や検索条件等のガイダンス機能

(1)に示した属性別入力画面を使うことにより、ユーザが指定しない条件は、テンプレート上のデフォルト値が採用されることになり、自動的に不足情報の一部はカバーされるが、対話イメージでのガイダンスの機能が弱い。ユーザに負担をかけずに、且つユーザが不安を感じない対話の方式により、ユーザの要求を確認する機能の検討が望まれる。

(3) 要求した情報が存在しないときに代替の情報を提供する機能

次善情報生成機能として、2章に実現形態を説明した。

いまのところ6種類の次善情報（代替情報）を順次試すという単純な代替情報生成戦略しか持っていないが、ユーザに対する問い合わせ機能を充実させて、ユーザの反応に合わせた対応ができるようにする必要がある。

(4) 統計やデータベースの自動選定機能

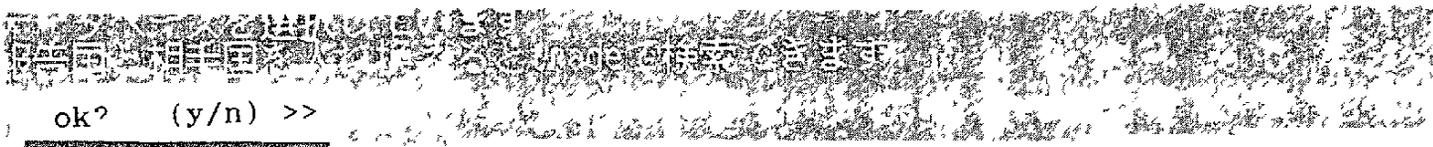


図3-4 データベース選定画面

試作システムでは、図3-4に示す画面で、検索可能なデータベースが複数存在する場合ユーザに問い合わせをおこなっている。選択基準となる知識の充実と共に、ユーザに複数の情報源の違いを判りやすく示す機能の充実が望まれる。

(5) データベースの自動検索機能

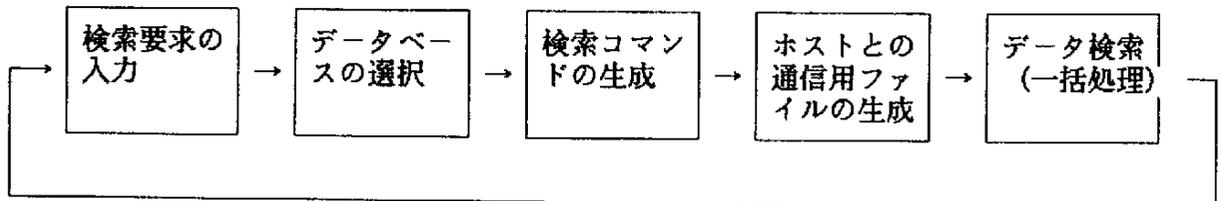


図3-5 データベースの自動検索処理の流れ

試作システムでは、図3-5に示す処理手順で、データベースの検索をおこなっている。通信手順に関する制約と、ホスト上での応答時間の問題があるためデータ検索は一括処理になっているが、欠測等検索実行後に次善情報が必要になる場合に対応するために、検索とコマンド生成の間にもフィードバックが必要である。

(6) 検索履歴の保存と連続検索支援機能

検索要求と成功した検索コマンドは、事例ベースを使って保存され、以後同様の検索要求が入力されると、優先的に過去の検索コマンドがユーザに示される。しかしながら、過去の事例を使いたくない場合のあり、事例ベースのメンテナンスのための機能が必要である。

また、入力された検索要求は、次の検索要求入力画面でデフォルト値としてユーザに示されるが、初期画面から入力しなおすという選択肢も必要であり、ファンクションキーを使った操作メニューをさらに充実させる必要がある。

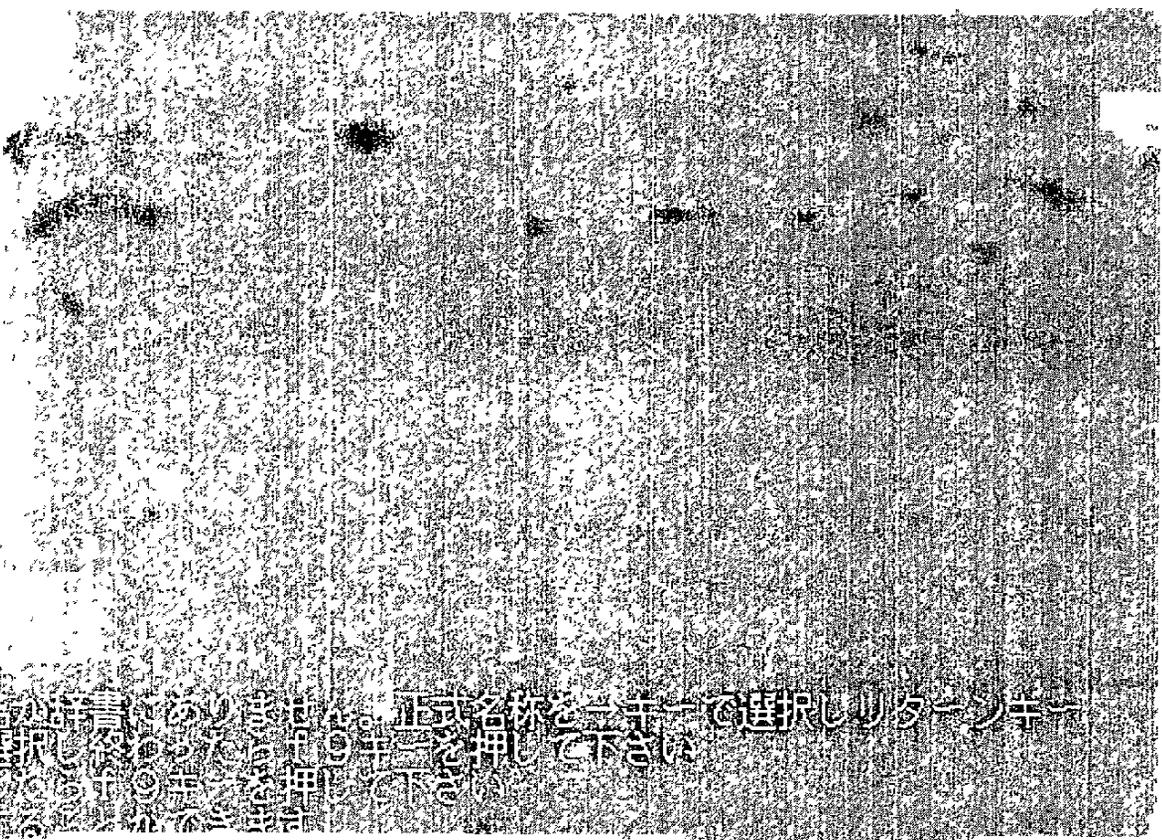
(7) 類義語の定義に関してユーザ定義を受け付ける機能

図3-6に試作システムでのユーザ定義画面を示す。

標準分類を介した指定の場合ユーザにとって、正しい定義ができたかどうか不安を残す場合が考えられる。また辞書の規模が多くなると、どれが自分の欲しい品目かを判断することが難しくなる。辞書の分類や、確認用のコード表示等の機能が必要であろう。

用語のメニュー

- 全世界
- 南アメリカ州
- 西ヨーロッパ州
- 中近東
- アフリカ
- 東南アジア
- ASEAN
- アジア
- ラテンアメリカ
- 生産圏
- 発展途上国
- 日本
- 大韓民国
- 北朝鮮
- 中華人民共和国
- 台湾
- モンゴル
- 香港



この画面は、用語の定義や検索に関する操作を行うためのメニュー画面です。各項目をクリックすると、詳細な定義や関連するデータが表示されます。また、検索機能を利用して特定の用語を迅速に見つけることも可能です。

図3-6 類義語に関するユーザ定義画面

3. 3. 3 ノシステム各機能の連携状況の妥当性

各機能は下図のイメージで連結されている。
 試作システムでは、事例ベースはコマンド生成の補助的な機能としての役割を果たしているにすぎないか、さらに、知識獲得の機能を強力にするためには、事例ベースを効果的に利用することか必要であろうと思われる。

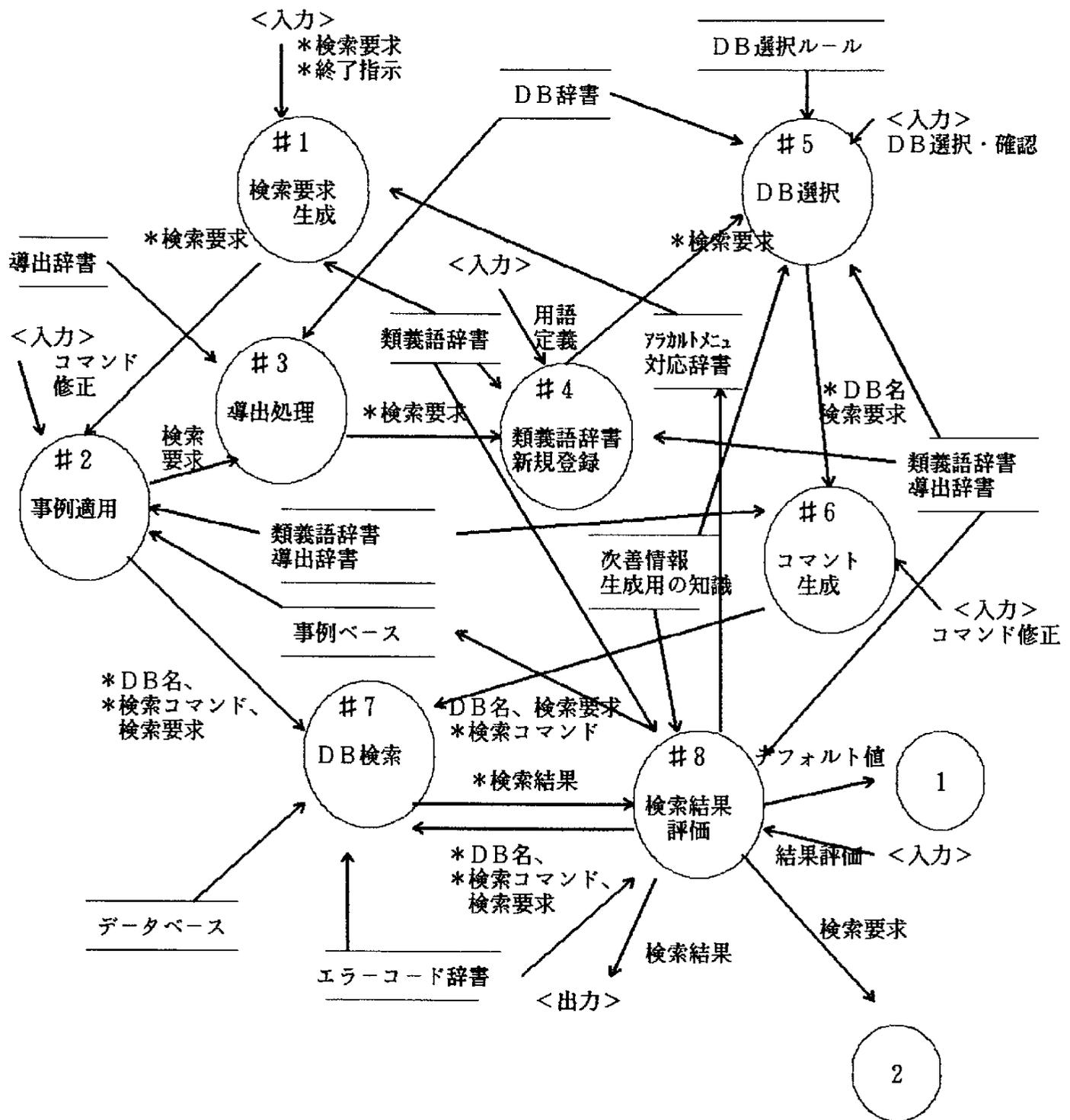


図 3-7 機能関連図

3. 4 知識ベース構成の妥当性と有効性

本節では、試作システムの知識ベースに関して、その構造の妥当性と有効性を評価する

3. 4. 1 知識ベースの役割

人工知能技術を応用したシステムにおける知識ベースの役割は、専門家の知識に相当する情報を蓄え、推論機能によりその情報を利用して問題解決を行うことである。また、稼働中に新たな知識を取り込み性能を向上させることができる。

知識ベースを利用することの利点は、次のように考えられる。

①判断的な処理の一部をデータとして取り扱うことができる。即ち、追加更新し易く、プロトタイプ的な開発に向く。

②判断のための条件を読みやすい形式に蓄えることができる。即ち、専門家と開発者のどちらにとっても読みやすく、処理結果の妥当性や因果関係が理解し易い。

3. 4. 2 本調査研究における知識ベースへの期待

3. 2 節の表3-1の人工知能技術への期待を参照すると人工知能技術への期待として次の各項目が挙げられている。

- ①データ利用に関するコンサルテーション機能の付加
- ②データベース選定に関する情報の知識ベース化
- ③データに関する注釈的な情報の知識ベース化
- ④データ分析のための知識の知識ベース化
- ⑤過去に利用されたノウハウの知識ベース化
- ⑥従来の使用例の知識ベース化
- ⑦検索コマンド生成手順の知識ベース化
- ⑧分類体系の知識ベース化

①は、②以下の各要求を包含し、専門家（データベースの専門家と、統計の専門家）のアドバイスをシステムを介して実現するための知識ベースの構築が、要求の本質であると考えられる。

3. 4. 3 第2期試作システムの知識ベース構成とその妥当性

(1) 人工知能技術への期待と対応する知識ベースの関係

前節の人工知能技術への期待に対応させて、試作システムの知識ベース構成を示す。

<データベースの選定に関する知識>

①データベース辞書

・データベースに収録されたデータの性質に関する知識

〔データの属性構成、収録期間、期種、年・年度の区別、単位〕

②データベース選択規則

・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知識

〔データベースが受け付けることのできる検索要求と必要な補正〕

<データに関する注釈的な知識>

①類義語辞書

・商品名及び国名に関する類義語の辞書

〔商品名：IRS (Intelligent Retrieval System) 汎用分類の各項目に対する類義語及び検索コード〕

〔国名：日本貿易統計で使用されている国名に対する類義語及び検索コード〕

②代替データ辞書

・概念的に近いデータに関する辞書

〔事項に属する用語間の代替可能な組合せ〕

③導出辞書

・データとデータの論理的な関係に関する辞書

〔商品名：分類上の上位下位関係〕

〔事項：データの算出式〕

④(データベース辞書)

<データ分析のための知識>

①(類義語辞書)

②(代替データ辞書)

<過去に利用されたノウハウ・前任者の使用例>

①事例ヘース

・過去の検索事例

[検索要求と選択データベース、検索コマンド、検索要求の特徴]

②(類義語辞書)

()は既出の知識ベース

[学習機能により、新たな類義語を追加登録]

(2) 知識ベースの構造上の関連

試作システムで作成された知識ヘースは構造的に次のように分類される。

— <データベースに対応する知識> —
データベース辞書

— <商品、国名、事項に対応する辞書> —
類義語辞書、導出辞書、代替データ辞書

— <検索要求に対応する知識> —
データベース選択辞書
事例ベース

(3) データベースの選定に関する知識の妥当性

当初の要求や背景を考慮して、これらの知識ベースの妥当性を検索する。

データベース選択のための知識としては、データベース辞書への収録項目が少ない。試作システムでは、検索対象のデータベースが3種類なので選択が可能であるが、検索可能

なデータヘースが増えると、検索要求に対しても、複数のデータヘースが検索可能になり、優劣を判断するために、さらに、調査方法、調査対象等の情報が必要になる。データベース選択規則に関しては、試作システムの内容ではほぼ必要な機能を満たしていると思われるが、今後ヒューリスティックなデータベースの選択方法が追加登録されると知識ヘースとしては、データというより、プログラムに近い形式にならざるを得ない。(例えば、相手国と報告国を入替えて、輸出、輸入の関係を入れ換えると検索可能・・・というような知識は、ヒューリスティックな知識といえる)

(4)データに関する注釈的な知識に関する妥当性

データに関する注釈的な知識としては、まず、知識の使い方に関して、注釈そのものをユーザに見せる方法と、システムがそれを使って内部的に処理する方法の2通りがあり、試作システムの場合後者の使われ方を意識した知識ヘースの集め方になっている点に注意する必要がある。機能としても注釈をユーザに示す機能は弱い。

データの意味を理解して、内部的に適切な次善情報の生成機能を果たすための知識として知識構成を考えた場合、知識の枠組みとしては十分なものができているといえるが、代替データ辞書に関してはシステム機能との連携をさらに考える必要があると思われる。代替データ辞書は、数値的に近いデータを代替情報として提供するための知識であるが、今のところ、数値的な近さの程度や修正係数等に関する情報は持っていない。さらに必要な系列を作成するために、複数の系列から合成する機能等を考慮すると、知識ヘース側としては、「データAとデータBが近い」という知識ではなく「データAとデータBは～の操作を行えば、～の場合に限り、同等のものとして扱うことができる」という形の知識にしておくべきであろう。

ユーザにデータに関する注釈を示すための知識として見た場合、データの調査法や、調査対象等、統計解説書に掲載されているような情報が必要な場合が多いが、統計調査単位での詳細な情報が不足している。これらはデータベース辞書の収録項目として記述されるべきであるが、利用上の問題としては、知識をそのまま画面に表示するのではユーザにとって理解しづらいものになる。また、複数データの違いを知りたい場合も多い。知識の形式としては解説書の内容をそのままつものではなく、ユーザの要望によって表示させることが出来るように、知識の各項目に何らかのインデックスを付す必要がある。

また類義語辞書を注釈的な知識と見た場合、類義語は品目や国名に関する基本的な名称

に対する一般的な別称の集まりと捉えるべきであるが、実際はローマ字入力レベルでの表記上の違いを主に収録したものになっている。

この点に関しては、日本語による表現のみを許し、表記上の違いは統一的な規則を示すことによって、取り込まないようにすることが実用的であるとの結論を得た。

(5) データ分析のための知識に関する妥当性

データの意味理解を支援する知識として、データに関する注釈的な知識と重複するものを挙げたが、データや系列の操作に関する知識や加工方法に関する知識等は、試作システムでは考慮されていない。

これは試作システムの機能が基本的には、単独の系列を求める検索要求を対象に考えられているためと思われる。

データの加工も含めたマクロ的な検索要求（日米の貿易不均衡の状況等）は、アラカルトメニューとして登録可能になっている。現在は、まだパターン化された検索を行うだけの機能にすぎないが、将来的には、この種の知識はアラカルトメニュー機能の拡張のなかで実現されるべきものと考えられる。

(6) 過去に利用されたノウハウ、前任者の利用例に関する知識

試作システムの開発過程で、過去の検索事例が重要であることは明らかになった。

現在試作システムでは、事例ヘースを使って、過去の事例を蓄積し、それに似た検索要求には効率よく対処できるようになっている。但し、データの表示等きめこまかなレベルでの過去の事例利用はまた難しい。また事例と辞書の連携関係についても今後検討を重ねていく必要がある。

3. 5 実用化の可能性に関する評価

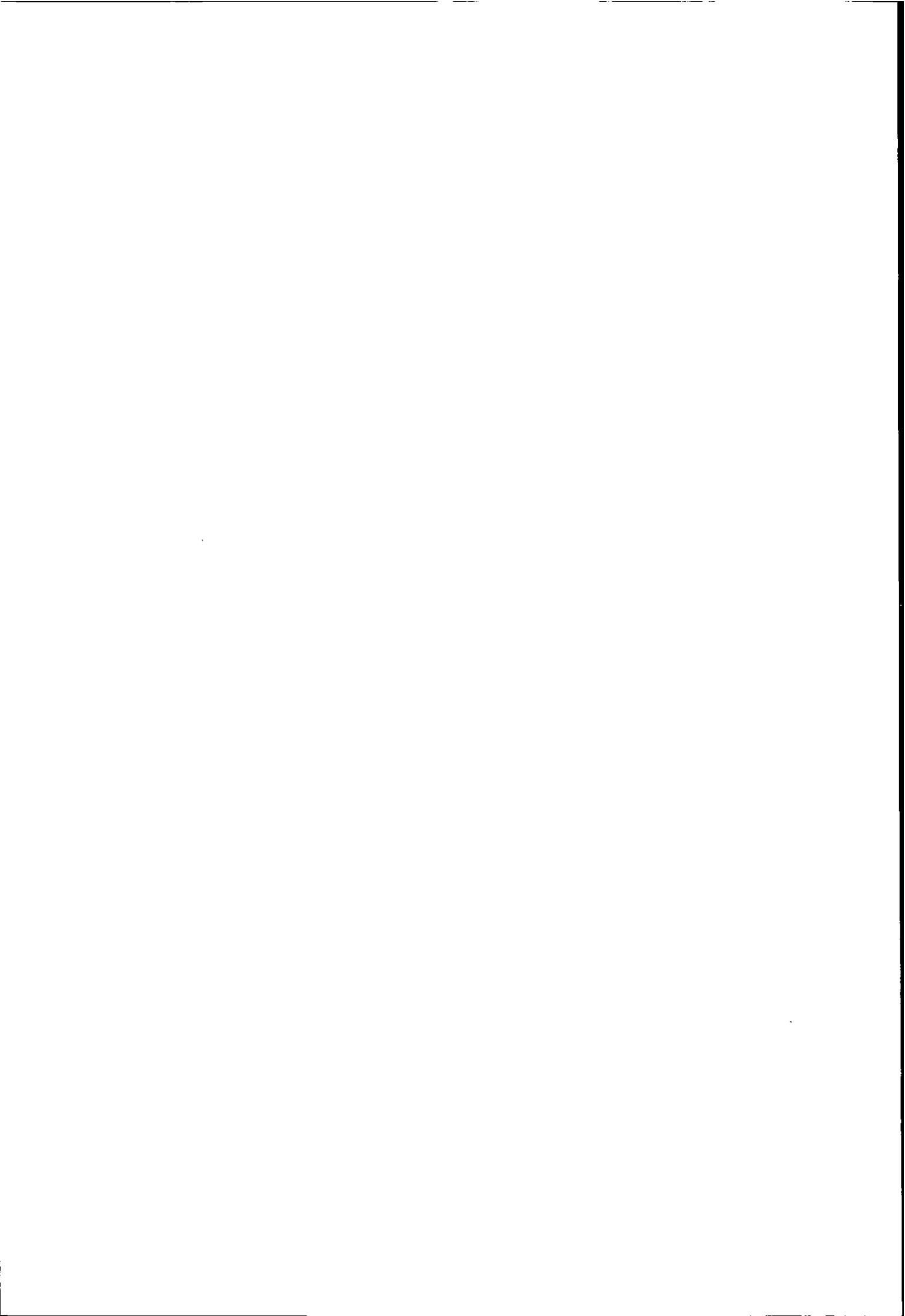
実験システムの稼働状態から、実用化の可能性について、稼働環境、運用方法の2点より評価した結果を次に示す。

(1). 実用的な稼働環境に関する評価

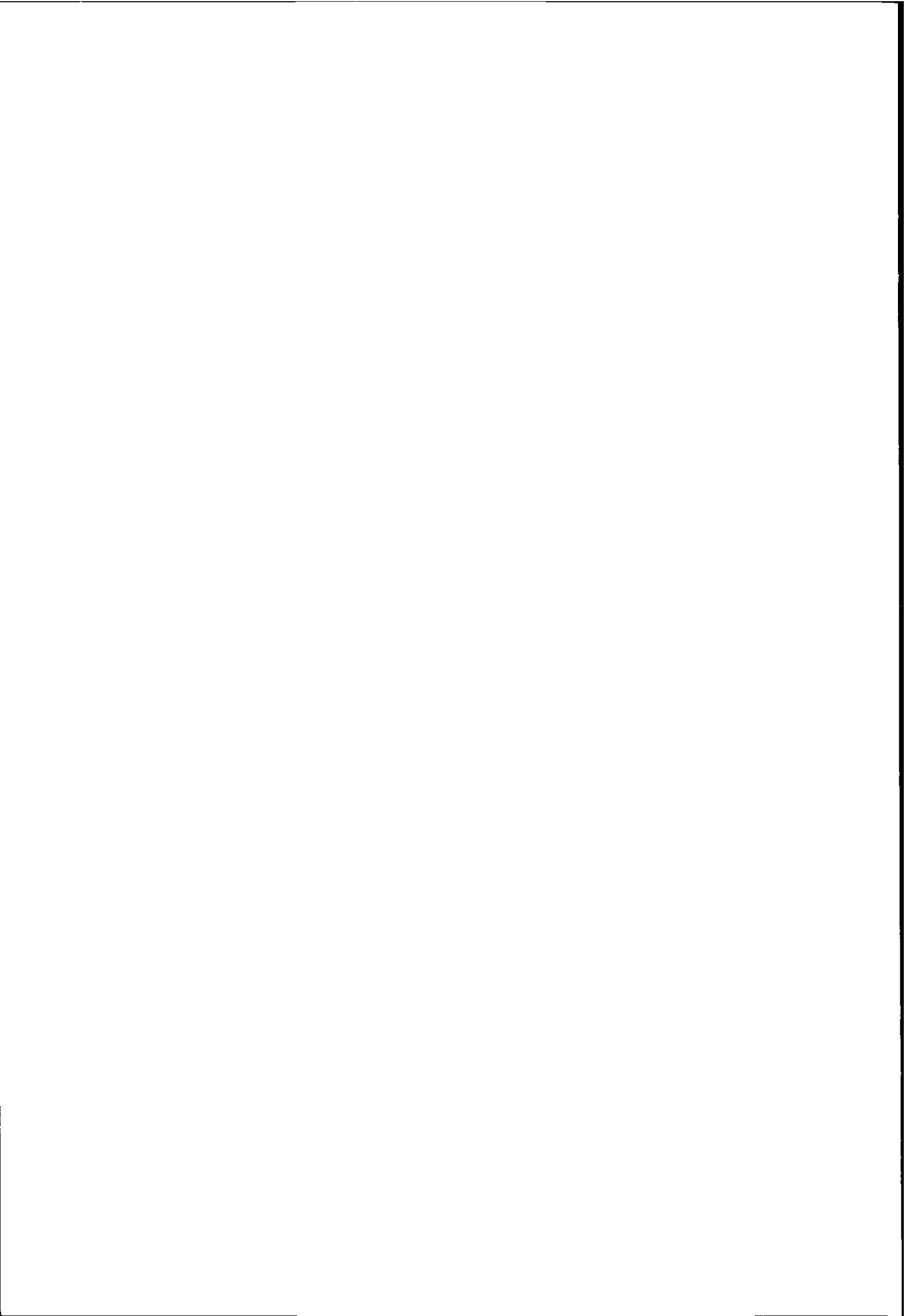
項目	実験システム	実用化可能性
端末機 メモリ専有量 知識ベース プログラム	(パソコン) 類義語辞書 (コード含む) 約 5 KB その他の辞書 (画面定義等も含む) 約 20 KB アプリケーション部分 約 120 KB 言語部分 約 200 KB 通信ソフト 約 100 KB	<p>☆現在の収録は、国数56か国、品目数25件程度である。</p> <p>☆標準分類の数に依存して類義語辞書が大きくなる。</p> <p>実用化のためには、最低1000分類程度また1語当たりの類義語の数も2・3倍になると思われる。類義語辞書の専有メモリサイズは、少なくとも200K程度は必要である。</p> <p>データの圧縮等は可能であるが、処理時間とのトレードオフの問題になる</p> <p>☆プログラム領域に関しては、現在の仕様の範囲で多少の改良を加えても、アプリケーションが1.5倍程度と見積もることができる。</p> <p>☆現在パソコンでは、メインメモリが640Kという制約がある為、日本語のフロントエンドプロセッサを組み込むことが難しいが、EMS等の利用により、この問題は間もなく解決すると思われる。</p> <p>☆類義語辞書に関しては、参照頻度を考慮すると、メンテナンスの問題はあるが、ホスト側ではなく、端末側にしたほうが実用的であると思われる</p> <p>ワークステーションを端末機とし、分類変更等の場合には、ホスト側の変更情報から、ワークステーション上の知識ベースを自動的に修正する機能を持たせることが望ましい。</p>
処理速度 ホスト機との 接続法	応答時間 端末側 3秒から 7秒 ホスト側 2秒から 50秒 1件当たりの処理時間 240秒から480秒 無手順方式 ホストからの応答を判定することが難しい。 通信の精度に不安がある。 ホストの応答が遅い場合の処理等通信部分のプログラム量が多くなり過ぎる。	<p>☆現在の規模では、端末側の応答時間はホスト側の応答時間より短い。</p> <p>検索式ができてから、ホストが検索処理を開始するまでの時間が長い。これは通信方式の改善によって大幅に短縮できる。</p> <p>☆ホストとの通信は有手順で行うべきである。その場合、ホスト側ソフトにも改造を加える必要がある。</p>

(2). 実用的な運用方法に関する評価

項目	実験システム	実用化可能性
知識の収集	ワーキンググループからのヒヤリング、経済用語関係の書物、百科事典、省内のコード対応表より事務局が作成。	<p>☆国名の類義語は、磁気テープで購入することも可能（購入後編集は必要）</p> <p>☆品目の類義語は、基本名詞のみ自動的に作成し、知識収集期間に、専門家が利用しながら作成する方法が現実的。一度に全体を作成することは、作業量が大きく、無駄なデータを収集し易い。</p> <p>☆知識ベースに登録するための基準、知識を利用する際の基準は、いままで未着手であったが重要である。ワーキンググループを組織して設定する必要がある。</p> <p>例 共通類義語辞書への登録基準 複数の基本名詞に対応する類義語に関する処理等</p>
知識のメンテナンス	知識獲得による自動的なメンテナンスのみ。	<p>☆一般性のある知識ベースを作成するためには、人手で定期的な知識のメンテナンスを行うことが必要。この際上述の基準が明確であれば作業は可能である。</p> <p>☆運用中に知識を収集する方式は実用的であるが、知識が希薄な状態の時は、一般的なユーザはかなり使い難い思いをされると思われる。一定期間の知識獲得段階を経て専門家がある程度の知識を投入してから公開するべきであろう。</p>



第4章 実用化に向けた今後の課題



第4章 実用化に向けた今後の課題

本章では、実験システムである第2期試作システムの成果をもとに、実用化に向けた今後の課題について論じる。4.1では利用目的毎にシステムに対する要件を分類し、多目的のユーザに対応するように拡張した場合のシステム構成を提案する。4.2では試作システムの知識ベース構成に対して実用化の観点から過不足の検討を行い、追加すべき知識について表現方法を考察し、実用的な運用のための留意点を挙げる。

次に第2期試作システムの範囲を離れて、広くデータベース検索の立場から、今後調査検討を進めるべき点について言及する。4.3では検索対象分野を拡張する際の問題点として、特に複数系列の比較についてシステム化のための知識表現を検討する。さらに、4.4章で複数系列の比較の問題への対策として、統計調査の企画段階での標準化を提言する。最後に4.5で、データベースシステムの今後の稼働環境について考察し、将来の稼働環境での、検索支援システムの問題点を確認する。

4.1 対象ユーザの利用形態と必要な機能

昨年度に調査したユーザの目的別利用形態には下記のものがある。

本年度の試作システムで対象としたものは、調査業務（通常業務）であるが、昨年度に挙げた利用形態全体に関して、試作システムの機能の過不足の関係を検討してみる。

対象ユーザによって考慮すべきシステムの要件を下記に挙げ、試作システムと各利用目的の関係を考察する。

4.1.1 利用目的の種類

前年度に検討した利用形態は次の4種類であった。（平成元年度報告書付録6）

(1) 研究業務

特徴

パイロットスタディを数回実施して実際の検索の方針を決定する。
データを眺めながら繰り返し複数系列間の計算作業を伴う分析作業を行う。
1系列のみで検索を終了する場合はまれ。
レスポンスタイムよりも、何が出来るか・・・が重要。

↓

必要な機能（支援）

データの特徴に関する情報の提供。
検索データのAvailabilityに関する支援。
検索データを加工し易い形式で、SAS またはLOTUS 用のファイルとして提供する。

(2) 調査業務 (定例資料作成)

特徴

毎回同一処理を行う。
前回と同じ処理をおこなえばよいかの比較が必要。
前回と比べ、結果がどのように変化したかの判断が必要。
報告書としての体裁を整える必要がある。



必要な機能 (支援)

処理手順の登録による操作の簡便化。
データの属性等の変化に関する管理。
加工結果の前回結果との変化管理。
報告書として体裁の整った結果出力。

(2) 調査業務 (通常業務)

特徴

定期的報告書の作成。
毎回同一のフォーマットで最新値を使用した作表。
担当範囲の監視。
自分の頼りとする指標について最新値を整理してファイリングする。
使用する統計の範囲はそう頻繁には変わらない。



必要な機能 (支援)

自動化。
担当者自身が起動をかけなくても帳表類を自動的に出力する。
体裁の編集支援。
上記の定期出力帳表類の体裁編集支援。(統計データを見ながら体裁を試行錯誤的にチェックしつつ編集する。ただし、頻度は極めて少ない。帳表は、そのまま報告書に使用できる体裁であること)
警告/強調等、傾向の変化に関する自動検出。(ただし、重要性は乏しい)
統計の属性変更に対する監視/チェック/及び対応の支援。(ただし、担当者のほうが詳しい場合もあるので、発生の警告のみでも充分)
担当範囲の管理機能。(自分が使用している統計に関する管理機能)
自分流の分類体系、インデックス等の追加的付加等。
定義済帳表の管理(ネーミング/配布先管理/出力実績管理等)

調査業務の定例資料作成と、通常業務はほぼ同一の仕事と考えてよいが、定例資料作成の方は、毎回同一の資料を作成することを目的とするのに対し、通常業務は、毎回同一のテーマを調査することを意味し、後者のほうがやや広い範囲の検索になる。

(3) 調査臨時業務

特徴

求める情報の所在の所在調査からおこなう。
データの抽出・検索方法の決定を行う。
複数のデータベースからとった情報を統合する。
データの比較を行う。
試行錯誤的な検索。
与えられたテーマに関しての調査。
テーマは、データベースの分類体系と一致しないことがある。
書誌的な情報が必要になることがある。



必要な機能（支援）

所在情報の提供。
データ検索のカイド。
情報統合上の注意等支援。
比較の正当性支援。
複数分野に渡る連続的な検索のための支援。
テーマと統計やDBの分類体系との対応付け。

(4) 企画業務

特徴

企画立案の背景を捉える。
企画立案の結果を検証する。
特殊なデータベースを利用する場合がある。
意見、考え方等を裏付けるために検索をおこなうことがある。
既存の雛形に改良を加えたい場合がある。
国際比較等を念頭において、情報収集する。



必要な機能（支援）

目的とする調査結果は、どのように評価したらよいか。
評価基準としては、なにを使えばよいか。
広範な領域に関するアラカルトメニューの用意。
過去の事例のガイド。
データの定義等が変わっていないか確認。
比較に関する支援。
標準的な分類基準、慣例の参照。

4. 1 2 システムへの要件と利用目的の関係

以上の利用目的を比較検討し、それぞれの共有点、相違点をシステムの要件として整理すると次のようになる。

(1) ユーザとの対話機能

試行錯誤的なシステムとの対話を必要とするものは、研究業務、次いで調査臨時業務である。通常調査と企画業務では、データの性質等に関する解説が重要になる。定例資料作成業務に関しては、手順の自動化が重要で対話の回数は少ないほうが望ましいと思われる。

このことから、次善情報利用の場合の問い合わせに関しても、研究業務や、臨時調査業務では利用の是非はユーザに確認する必要があるが、その他の場合は自動的に利用したほうがよい場合が多いと予想される。

(2) ホストとの通信機能

定例資料作成の場合は、検索すべきデータベースが決まっていると思われるので、ホストへのアクセスは一度だけで済む場合が多いと思われる。

他の場合、次善情報生成の必然性からも、通信時間の損失が問題にならないければ、ホスト上のデータベースへのアクセスは、必要な情報が得られるまで繰り返すほうが望ましい。

(3)データの加工機能

データの加工のなかには、グラフ化や作表等報告書作成用の機能と統計処理等の分析用の機能がある。

研究業務および調査臨時業務ではデータ分析のための機能が必要であり、その他の業務に関してはレポート作成のための機能が重要である。

いずれにせよ、これらの業務に関しては、加工の処理も含めて業務が完了するという性質があり便宜を考慮する必要はある。

また、データの加工に伴って、複数列の比較や相関分析等、複数列を問題にする処理が必要になる。

(4)データの性質や、データの変化に関する説明機能

定例的資料作成や通常調査の場合は、継続的に検索しているデータの性質（調査対象や集計方法等）が変化した場合は変更前後でどのようにデータを解釈すべきかの判断が必要である。この機能はどちらかといえばホストのデータベース自体が持つべき機能であるが、変更情報をもとに判断を行う場合は、検索支援システム側でも持つ必要がある。但し、この種の知識は‘特定の国に関する特定の品目に関するデータの性質’ということができ、他の知識との結びつきの少ない、‘点’的なデータである。これらは、膨大な書誌的データとして統計の解説書に記録されている。効率よくシステム化することの難しい知識であろうと思われる。

(5)データの値に関する監視機能

特に定例資料作成や通常調査の場合に、値の変化に関する監視（アラーム）が重要である。但し、値が異常であるか否かの判断は、いくつかの方法が存在する。

具体的な方策としては、絶対値に対するしきい値を与える、前期比に対するしきい値を与える、前年同期比に対するしきい値を与える 等のうちのいくつかを使って、異常状態を表現し（段階的な表現も可能である）自動的な検索処理のなかで、異常データに関するメッセージ表示を行うこと等が実現可能なものとして考えられる。

(6) システムの自動運行に関する機能

特に定例資料作成や通常調査の場合自動化が望まれる。これに対して、その他の業務では毎回異なる検索を行うため自動化のメリットが少ない。

4. 1. 3 多目的利用への対応方法

実際の検索作業には、第2期試作システムの機能の他に、データの加工、データの比較、時系列解析等の統計処理機能、データの監視等の機能が必要になるが、4. 1. 2 で説明したように、利用目的によって互いに矛盾する要求があるため、全ての利用目的を一度に満足することは得策ではない。

より広い利用目的への対応を考慮してシステム構成を検討すると下図のような構成になる。

専門家でないユーザを対象とした場合、下図のようにIRSのインターフェイスに類した共通インターフェイスを持ち、ユーザのニーズに合わせて検索や加工の手順を自動的に決定し、処理を行うような構成が望ましい。

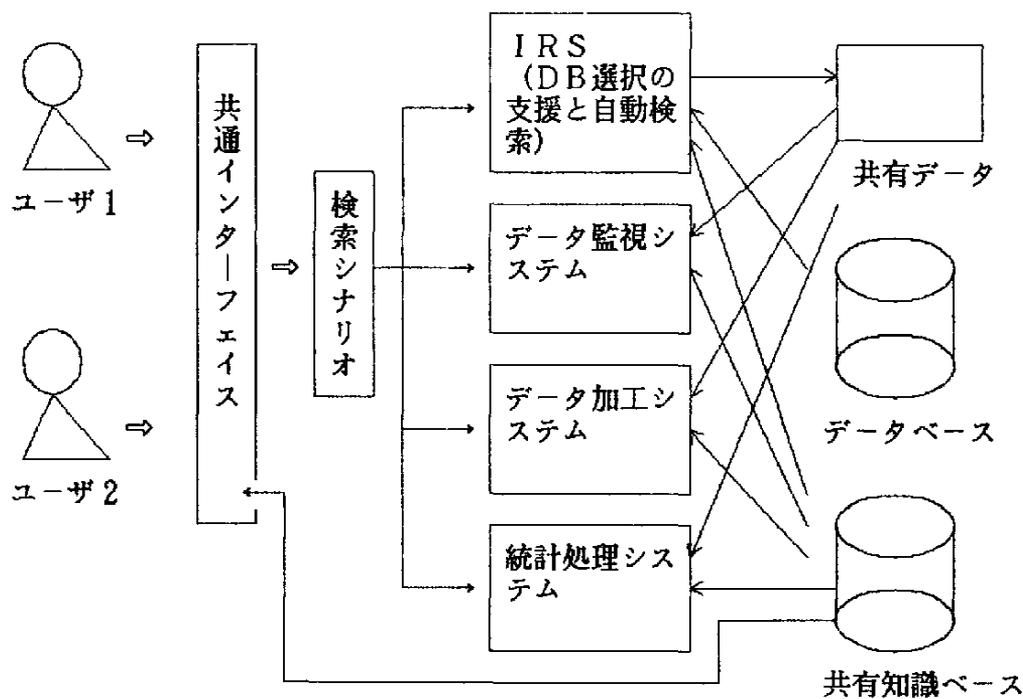


図4-1 複数の利用目的に対応するシステム構成

4. 2 知識ベースの構造と知識獲得の方法

試作システムの知識ベース構成に関しては、2章で説明したが、本節では今後の課題として、理想的な知識ベースの構成と知識獲得の方法に関して検討する。

試作システムでは、Prolog の述語として各知識ベースを蓄えた、また必要に応じて知識獲得の仕組みを組み込んだ。

4. 2. 1 必要な知識の構造

システム全体をみて必要な機能を知識の性質によって分類すると下図のようになる

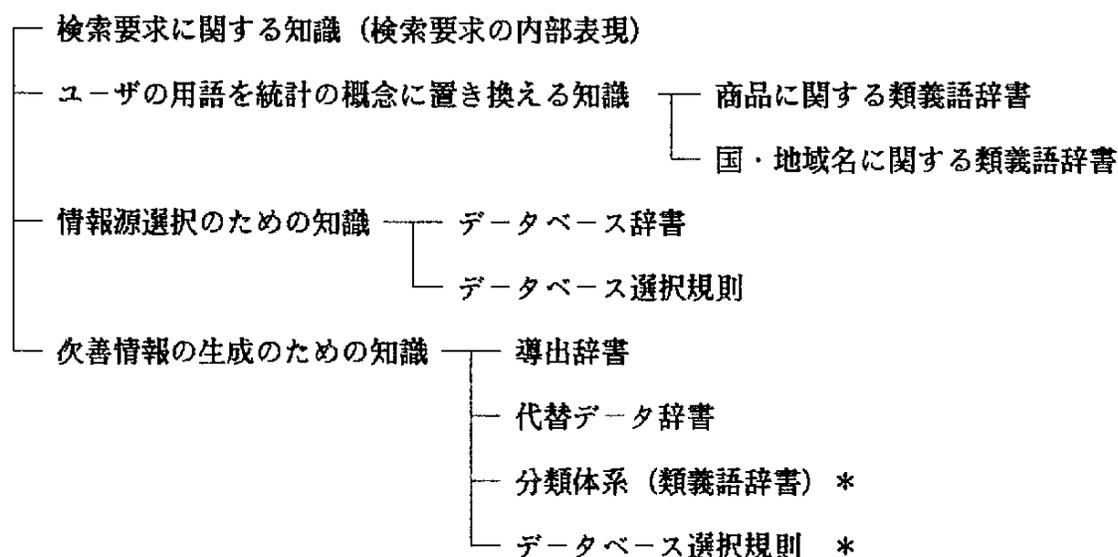


図 4-2 システムの知識ベース構成

上図の*印は、重複しているものである。

このように知識を整理すると知識の役割が明らかになる。また、知識のなかには属性レベルの知識と情報（属性の組合せ）レベルの知識があることが判る。

即ち、類義語辞書、導出辞書、代替データ辞書 は、属性レベルの知識であり、検索要求の内部表現、データベース辞書、データベース選択辞書は、属性の組としての情報に関

する知識であるということが判る。

4. 2. 2 書誌的情報の取扱

図4-2に含まれていないが実際の検索に必要な知識が存在する。それはデータの性質に関する書誌的な情報である。

例えば、国際通貨基金（IMF）のInternational Financial Statistics（世界金融統計）の日本の金融統計に関する注釈として、下に示すような情報が添付されている。

Date of Fund membership August 13, 1952
Standard Sources:
B: Bank of Japan, *Economic Statistics Monthly*
S: Bureau of Statistics, *Monthly Statistics of Japan*
Exchange Rates: Market Rate (End of Period and Period Average). Midpoint rate in the interbank foreign exchange market in Tokyo
International Liquidity: Prior to March 1978, IFS data on gold excluded amounts held on the government's precious metals special account. These holdings then were fully transferred to the Bank of Japan in March 1978
Gold (National Valuation) (line 1and) is the U.S. dollar value of official holdings of gold as reported in the country's standard sources. Gold was valued at U.S. dollars 35 per fine troy ounce until December 1971, at U.S. dollars 38 from January 1972 to January 1973; at U.S. dollars 42.222222 from February 1973 to July 1974. And after August 1974 gold is valued at SDR 35 per fine troy ounce and converted into U.S. dollars at the dollar/SDR rate on the IFS page line 5a for the United States.
Beginning December 1973, data for lines 7a,d and 7b,d include long-term foreign accounts and therefore are not the U.S. dollar equivalents of lines 21 and 26c, which comprise only short-term accounts, data are from the Bank for International Settlements' *Annual Report and Quarterly Press Release*
Monetary Authorities: Consolidates the Bank of Japan and monetary functions undertaken by the central government. The contra-entry to the Foreign Exchange Fund and Treasury IMF accounts is included in line 16d
Deposit Money Banks: Consolidates all banks listed in source B, plus Shinkin banks, the Norinchukin Bank, and the Shoko Chukin Bank. Data for lines 21 and 26c include the foreign assets and foreign liabilities of resident foreign banks which are reported in section 40. Bonds (lines 26ab, 36ab, and 56ab) refer to bank debentures issued by deposit money banks
Other Banking Institutions: Comprise specialized credit institutions, which cover resident foreign banks, financial institutions for small business, financial institutions for agriculture, forestry, and fishery, securities finance institutions and other private financial institutions, government financial institutions, the Trust Fund Bureau, Postal Savings and Postal Annuity. Data related to trust accounts of all banks are included. Demand and Time Deposits (line 45a) includes certificates of deposit
Nonbank Financial Institutions: Comprise life and non-life insurance companies, the National Mutual Insurance Federation of Agricultural Cooperatives, and mutual insurance federations of agricultural cooperatives
Interest Rates: All interest rate data are from source B
Discount Rate (End of Period) Rate at which the Bank of Japan discounts eligible commercial bills and loans secured by government bonds, specially designed securities, and eligible commercial bills. This rate is considered the key indicator of the Bank's discount policy

• • •

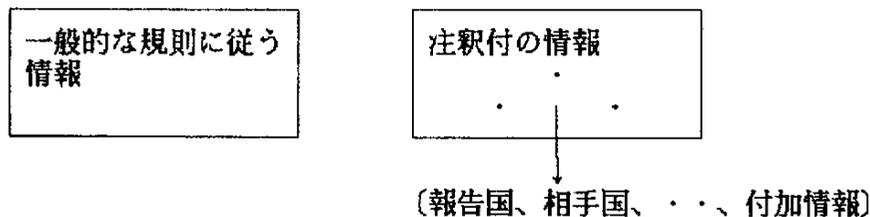
図4-3 書誌的情報

これらは、上記の情報レベルの知識、属性レベルの知識という観点から分類すると、情報レベルの知識にはなるが、グループとしての情報ではなく、個としての情報に付与された知識である。

この種の知識は、現在データベースにも収録されず、知識ベース化も難しいと思われる（膨大であり、一般性がない）が、実際にデータを利用する場合には、重要な情報である。このような知識をどう取り扱うべきかが、知識の取扱に関する今後の課題であるように思われる。

システム化はしないとする事も一つの考え方であるが、次のような方法でのアプローチが可能ではないかと思われる。

① 前述のような注釈の付いた情報の集まりに付加情報を付けて注釈知識ベースを構築する。



② 付加情報としてはフレームのように属性をきめない、キーワードの集まり（順序には、意味がない）のようなものとする。

③ キーワードとしては、

含む (16d)、測定 (年度中間点) 、・・・

のように述語論理的 (Prologの述語に近い形式) に表現し、柔軟性を持たせる。

(カッコのなかは、複数の情報が入ってかまわない)。

④ 複数の情報を比較する場合に、それぞれが注釈付データであるかどうかを判定し、注釈の内容として、比較不可能なものが含まれないか否かを判断するための知識ベースを構築する。

このような方法で、比較可能性の面から構築可能な程度の量の知識が抽出できるのではないかと考えられるが、実際の検証が必要である。

試作システムの開発を通して検索に必要な知識は上記以外にも存在すると思われるが、試行錯誤的なノウハウであったり、個別の事情にもとづく処置であったりして知識ベースとしての形式を定めることが難しいと思われる。

第2期試作システムで採用した事例ベース推論の仕組みは、まだその機能が充分には発揮されていないものの、今後、検索上の知識、特に経験に基づく知識を整理する上で役立つものと思われる。

4. 2. 3 過去の検索経験に関する知識

さらに図4-2に含まれない知識として過去の検索経験に関する知識がある。これは、図4-2の種々の役割のレベルには収まらない、直接的に、検索要求と検索コマンドを結び付ける知識である。

第2期試作システムでは、事例ベース推論の技術を適用して一部この機能を実現したか、現状ではコマンド生成の効率化のために使われているに過ぎない。

実際の検索場面では、理屈ではなく、'××白書で、このような使用例があったから、この検索式で検索する'といった検索のしかたがかなり見受けられる。

この種の他の知識ベースに収容するのが難しい知識を収録する機構として、今後事例ベースの利用が有望であると思われる。

4. 2. 4 知識獲得の方策

試作システムの特徴は、運用中に知識を獲得し、徐々に性能を上げる方式を取った点である。

知識の性質から、運用開始時にすべての関連知識をデータベース化することは、運用開始までの工数がかかり、また、無駄なデータの多い知識ベースを構築する危険を含む。

但し、運用中に知識獲得をするシステムの場合に、ユーザの利用実績を調査し、同じユーザ辞書を共有するユーザグループを想定することによって、どの程度知識獲得のための

運用を行えば、ユーザが負担を感じないで検索を行えるかを把握する必要がある。

4. 2. 5 知識の品質保証及び知識のメンテナンス

類義語辞書の維持に関して、収録の基準をとうするかが重要である。

収録の基準としては、例えば、

収録の基準に関しては、日本語のかな漢字で表現された用語で基本名詞に対する一般的な呼称を収集する。表記方法には基準を設ける。

(試作システムでローマ字表現された用語を収集し、ヘボン式や訓令式等の表記方法の違いによるものも収集する方針で辞書が作られた。)

といったものが考えられる。

また、ユーザ辞書に蓄えられた知識をどのように共通辞書化するかという点については、独特の言い方ではあるが、特定のグループでは、一般的に使われているような用語をどのように扱うべきかといった問題が考えられる。

この場合、ユーザのグループに応じて、複数の共通辞書を持つことが実用的であるが、その運用管理はデータベース管理者が行うべきであろう(知識ヘース管理者や、ユーザの立場での管理では、ユーザの利用に関する情報がデータベースシステムに反映されなくなる恐れがある)

以上より、検索支援に必要な知識の性質を、当初モデルとして検討していた、ユーザの世界、統計の世界、データベースの世界というフレームに立ち戻って考察すると、ユーザの世界での'常識'に相当する知識が、メンテナンスの面からは、取扱のむづかしい知識であることがうかがえる。しかもユーザの'常識'を適切にシステム化することは、データベースの利用効率を高めるうえでの大きな要件であると考えられる。

実用化の際の留意点として次のものが挙げられる。

①辞書の収録に関しては、管理可能な範囲を考慮し、可能な限り明文化した収録規則を作成すべきである。

②専門家でないユーザが誤った知識を投入することを避けるために、個人辞書の共通辞書化には専門家による正当性の判断という手順を踏む必要がある。（利用の件数が多いという理由で、共通辞書化することは危険である）

4 3 分野の拡張指針

4 1でも触れたが、試作システムの今後の拡張の方向として、検索範囲を現在対象としている貿易統計だけでなく生産、投資、物価等の範囲に拡張することか考えられる。

本節では、分野の拡張の際に留意すべき点について、試作システムをもとに言及する。

4 3 1 検索要求の受入れ機能の拡張

分野の拡張の為には、システムがさらに柔軟に要求を受け付けるように、システム機能や知識構造を検討する必要がある。

検索要求を表現するための属性が検索対象分野毎に異なることが予想される為、複数の検索対象分野でユーザの検索要求を表現するために十分な、属性構成を持つインタフェースと制御の仕組み。

入力テンプレート自体は属性の変更に耐えられるが、ユーザの入力から、当該分野を判定し、その分野に合わせて表示する属性欄を変更する等の柔軟な対応が必要である。

4. 3. 2 複数系列の比較に関する処理

異なる分野や情報源からのデータの比較可能性についてのガイドを工夫する必要がある。

データの比較可能性のなかには次の要素がある。

- ・調査方法の違い
- ・調査対象の違い
- 推計方法の違い等統計的な処理方法の違い
- 調査間隔や速報性の違い
- 指数等の算定式の違い
- ・集計度（産業、地域等）の違い

また、上記の各要素は異なる国の間で発生するばかりでなく、同じ国内で、異時点間で発生する場合があることに注意が必要である。

必要なガイドの要素としては、

- ・数値的な傾向（A法とB法では、A法が大きめに計上される等）
- ・調査対象の違いがデータにとのくらい影響をあたえるか。

（生産統計と貿易統計での'自動車'の調査対象の違いは、全体の何割か等）

等が考えられる。

但し、上記ガイド要素で支援可能な範囲は全体の一部であり（時系列的に安定した比率を得ることは困難だと思われる）、異なる系列の比較ノウハウを数量的に表現できない場合が多い事を認識すべきであろう。

検索支援システムの対応としては、

- Ⅰ. 調査方法、調査対象等の比較可能性の要素をデータとしてユーザに示す。
- Ⅱ. 2系列の比較の際に、数量的なガイドが可能な場合はユーザに示す。
- Ⅲ. 2系列の比較に関して、数量的なガイドか不可能な場合は比較可能性の程度を何らかの記号でユーザに示す。

例 A 比較不可能（殆ど相関がない）。

B マクロ的な傾向を比較するためになら使える。

C 比較的整合性があり、若干の修正を加えれば比較可能。

等が考えられる。

具体的にデータの比較可能性を検討するための要素として、総務庁の統計総覧に記載されている項目を以下に挙げる。（知識ベースとしての表現形式を検討するため）

各統計調査に関して、次の各項目が記載されている。

(1)統計調査名

(2)実施機関

(3)目的（調査目的及び利用目的）

(4)沿革（調査開始年月、その後の調査実施状況、調査事項、調査方法等の主な変更状況

- (5)調査の構成（調査票が複数の場合、その区別を示す）
- (6)集計（調査結果の集計が中央機関でおこなわれるか、地方機関でおこなわれるか、民間委託の有無、機械集計、人手集計の別）
- (7)公表（集計結果が公表される刊行物の名称、発行時期）
- (8)表章（調査実施者の確定結果公表の地域的表章単位）
- (8)備考

また、同統計調査総覧では、さらに調査票単位に次の各項目が掲載されている。

(1)調査対象

- ①地域（調査対象となり得るものの属する地域範囲、標本調査の場合は、母集団の地域的範囲）
- ②単位（調査対象の観察単位（調査単位））
- ③属性（調査対象となるものの地域を除く属性）
- ④抽出枠（調査客体の抽出に使用した母集団のフレーム名）

(2)調査方法

- ⑤選定（調査対象の選定方法を全数、無作為抽出、有意抽出の別で示したもの）
- ⑥客体数（客体数または、標本数、抽出調査の場合は、母集団の大きさ）
- ⑦配布・収集（調査法の配布・収集が調査員によるか、郵送によるか、併用か）
- ⑧記入（報告者が記入するか、調査員が記入するか両者の併用か）
- ⑨把握時（調査事項の把握時点または、把握期間）
- ⑩系統（調査実施機関から報告者にいたるまでの調査の系統と関係機関）

(3)周期・期日

- ⑪周期（調査の実施または、報告の徴収が一回限りか、月、四半期、年等の間隔で繰り返し行われるか）
- ⑫期日（調査表提出期日または、面接期日）

(4)⑬調査事項（調査票によって申告または報告をもとめる項目）

実際の統計調査は下表に示す例のように、記述されている。

例

統計調査の名称	種別	特性（調査対象、方法等）
国勢調査	人口	<p>地域：全国、単位：世帯及び個人、属性：本邦に居住する者、選定：全数、配布・収集：調査員、記入：自計・他計、把握時：昭和60年10月1日現在、系統：総務庁統計局→都道府県→市町村→調査員→報告者、周期：5年、期日：昭和60年10月、調査事項：所帯員の氏名、性別、続柄、出生年月、配偶者の有無・・・</p>
工業統計調査	鉱工業	<p>（調査票甲）地域：全国、単位：事業所、属性：従業員30人以上の製造業に属する事業所（製造、加工または、修理をおこなっていない本社または、支店を除く）、選定：全数、客体数：58,984、配布・集計：調査員、記入：自計、把握時：毎年12月31日現在、系統：通産省→都道府県→市町村→調査員→報告者、周期：年、期日：毎年12月から1月、調査事項：事業所の名称、所在地、資本金額または、出資金額、従業員数、現金給与総額、</p>
輸出入動向調査	商業サービス・貿易	<p>地域：全国、単位：企業、属性：貿易業者で、年間輸出入取扱高上位250社、抽出枠：貿易業態統計調査（指定統計第58号）、選定：有為抽出、客体数：250、配布・収集：郵送、記入：自計、把握時：四半期、系統：通産省→報告者、周期：四半期、期日：毎年2月、5月、8月、11月の各月の5日、調査事項：地域別商品輸出入成約高、通関の実績と見通し、品目別輸出入の成約の動向、外貨建込約比率と各期の換</p>

貿易業態調査 (調査票甲)	商業 サービス 貿易	算レート 地域：全国、単位：企業、属性：①輸出業者又は輸入業者②自己が生産、修理及び加工した貨物の輸出を業として行う者③自己の生産に直接使用する原材料又は燃料の輸入を業として行う者、抽出枠：事業所名簿選定・全数、客体数・17,000、配布・収集・郵送、記入・自計、把握時：3月31日現在、系統：通産省→通商産業局→報告者、周期 年、期日：毎年5月 調査事項：商号、本店の所在地、業種、事業所数及び従業員数、資産、資本及び負債の額、外国人の投資額、商品の手持ち額・取扱額・収益・費用、商品別取扱額及び、地域別輸出入額、主要取扱商品及び、主要取引相手国
------------------	------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

総務庁：統計調査総覧より

データの比較可能性に関して、以上のような項目の抽出は可能である。しかしながら、統計調査総覧では、各種統計の分類体系の違い等についての記述がない。そのためこれを統計検索上の手掛かりとすることはできない。統計総覧のような記述で統計調査の特徴を表現した場合、複数の統計間の違いは判りにくい。

さらに、統計の解説書等から統計調査に関する情報を抽出する必要がある。

データの解釈を行うためにも、ユーザに説明を行うためにも、データ比較の方策を検討する必要がある。

検索支援システムとしては、4. 2. 2 で述べた、述語論理とキーワード検索を組み合わせたようなアプローチ等も含め、書誌的情報の利用方法に関する技術的な方法論が求められている。

4. 4 統計分類に関する提言

4. 3. 2の複数系列の比較は、統計調査の段階で統一性がとれていれば、解決する問題である。これを受けて、本節では、統計分類の標準化に関する提言を行う。

4. 4. 1 現行の統計分類に関する問題点

特に商品に関する統計に関して実際にデータの比較可能性をデータベース化しようとする時、統計ごとの品目対応表を得ることか困難な場合が多い。

現在、日本標準商品分類とHS分類が商品に関する標準的分類となっているが、実証統計データとこれらの分類との対応をつけようとする場合つきのような問題がある。

① 新しい商品が登場してから分類項目に追加されるまでに時間がかかり、分類不可能な商品かてきる。分類不可能なものに関しては恣意的に分類せざるを得なくなり、統一的な分類対応表が作成困難となる。

② 日本標準商品分類は、分類基準等が利用者に容易に判る形で公表されていない。

(関税分類については、関税率表解説、関税分類例規集、輸入商品の分類実務等が公表されている。)

③ 国際的に統一された分類で全ての統計が作成されていないので、統計の国際比較が難しい。

例 日本標準商品分類とHS 分類の対応

分類番号	商 品 項 目 名	Description	HS
60 61	扇風機	Electric fans	841451
60 611	卓上扇風機	Desk fans	
60 612	床上（座敷）扇風機	Floor fans	
60 613	天井扇風機	Ceiling fans	
60 619	その他の扇風機	Other electric fans	
60 62	換気扇	Ventilating fans	841451
60 621	ウィンドファン	Window fans	
60 622	空調換気扇	Ventilating fans with thermo exchange system	
60 623	レンジフード換気扇	Ventilating fans for range hood use	
60 629	その他の換気扇	Other ventilating fans	
60 63	冷風扇（水冷式クーラ）	Fan coolers	847989
60 64	サーキュレータ	Circulators	
60 7	電気洗濯機（業務用を除く。）	Electric washing machines (home use only)	

日本標準商品分類、HS分類のそれぞれの分類について見直しの間隔を短縮し、商品の実情に合わせるとともに、分類の定義を明らかにし、これらの分類との分類対応表を磁気テープで供給する仕組みを作るなど、データ供給側からの取組を期待したい。

4. 4. 2 統計分類上の問題に対する取組

4. 4. 1 では、日本標準商品分類とHS分類に関して標準分類としての評価を行ったが、さらに他の多分野での横断的な情報検索を考慮すると、産業連関分類についても標準分類として利用価値を検討する必要があると思われる。

産業連関表には、産業連関分類と、工業統計や、貿易統計との対応表が添付されており複数分野への対応付けが考慮されておりコンバーターも用意されている。

また、産業連関表では、国内の生産に関しては網羅的に捉えられており、工業統計と貿易統計を直接比較するよりも、産業連関表を介したほうが整合的な情報が得られると思われる。

例えば、中小企業（従業員3人以下の事業所）で生産した商品に関しては、貿易統計には現れるが、工業統計には現れない。

また、工業統計では、企業が複数の分野に渡る商品を生産している場合、（機械、鉄、食品等）全ての生産高が主要商品の生産高（例えば機械の生産高）として計上されてしまう。これに対して産業連関表は、アクティビティ・ベースに組み換えられている。

産業連関表は、調査期間が5年間隔であるが、調査と調査の間を埋めるための延長表（

年次)も用意されており、年単位の利用には耐えると思われる。

但し、産業連関分類では、第2期試作システムで、取り上げた日本貿易統計のような商品レベルの分類に関しては集計度が高く、詳細な情報は不十分である。

産業連関表は標準分類として利用し、さらに、下位の詳細な分類への連結には各分野での標準的な分類を接続するような分類体系によって解決が可能であると思われるが、32でも述べたように、分類精神の違いを考慮した上で、適用することができるかどうか、さらに分析が必要である。

以上を踏まえ、情報化の進展や普及に則し、コンピュータによる処理を前提とした統計調査のあり方が再検討されることを期待する。

4. 5 稼働環境

4. 5. 1 必要な稼働環境

本年度の試作システムでは、ホストコンピュータ上のデータベースを、端末機上の支援システムを使って検索するという方式で、システムの機能を実現した。

この条件を前提として実用システムの環境を考察すると 実用的なシステム環境としては、

- ①類義語辞書のためのメモリ
 - ②データの加工、比較等のためのウィントウ機能
- の面で、実験環境よりも高い性能が要求される。

現状で必要とされるメモリサイズは、類義語辞書、言語、通信ソフト等を含めて、650から700Kバイト程になる。さらに日本語のフロントエンドプロセッサも必要とする場合、800から900Kバイト程度の主記憶容量が必要になる。(現在仮想ページ機能等、言語に含まれる補助記憶へのアクセス機能によって、少ない主記憶で、稼働させている状況である。)

4. 5. 2 パーソナルコンピュータの技術動向

パーソナルコンピュータの世界では、従来のMS-DOSと呼ばれるOS(オペレーティングシステム)に対して、OS/2という新しいOSが普及しつつある。

これによって、640Kバイトという主記憶の限界が拡張され、また、マルチタスク機能により、データ処理と、通信の2つの処理を平行して行うことに期待かもてるようになってきた。また、MS-DOS上でもEMSというメモリ拡張機能が実現された。

OS/2とMS-DOSの違いは、OS/2では、完全なマルチタスク機能が実現される点である。MS-DOSにも疑似マルチタスクと呼ばれるような、類似機能はあるが、MS-DOSはあくまでもスタンドアロンのシステムを前提にした環境であると言えよう。

まだOS/2が本来の機能を発揮できる機種は、限定されており、不満が多いのが現状であるが、近い将来、この点では、パーソナルコンピュータとワークステーションの差が少なくなると思われる。

また、マルチウィンドウ等、グラフィックスの機能に関しては、パソコン上のMS-Windowsというウィンドウ管理システムがリリースされているが、現在のパーソナルコンピュータのレベルでは、高機能のものを選択してもワークステーションとくらへ、ハード的にもまた、ウィンドウをコントロールするソフトの面からも劣っているのが実情である。

4. 5. 3 実用的なシステムのための稼働環境の選択

第2期試作システムのように、大型コンピュータにあるデータベースを検索することを前提とした場合は、OS/2が有望であると思われるが、今後データベースのあり方はさらに多様化し種々の環境での運用が可能になるとと思われる。

但し、CD-ROM化されたデータベースが販売されるようになり、データベースシステム自体が、ワークステーションやパソコン上で稼働することも可能になりつつある。

そのような環境が実現した場合に新たに問題となるのはデータベースそのものの保守の問題であろう。データの更新に合わせて、複数の環境で独立に稼働しているデータベースを更新するためのシステムや運用形態が、分散データベースシステムとして議論され、実用化され始めている。

このようなホストから離れたデータベースシステムが実現した場合の利点としては、ホストとの通信時間が不要になり、高速検索が期待できること、また、データを終始同一の環境のなかで扱えるため加工ソフトへのインタフェースが構築しやすくなること等が挙げられる。

第2期試作システム開発で得られたノウハウのなかで、上記のようにデータベース環境が変化しても、意義を持ち続け、さらに詳細な分析を要するものは、次善情報や、知識獲得を含む、知識処理の枠組みであろう。通信や、結果の表示等は環境に左右されるし、ユーザインタフェースの部分も環境によって条件が変化してくると思われる。

データベースシステムと支援システムが同じ環境で稼働する場合は、複数系列の取扱や検索後のフィードバック等は遙かに楽に構築できるものと思われる。

む す び

本報告書では、3ヵ年計画の最後の年に当たり、本年度の成果とともに、3年間の調査研究の成果をとりまとめた。

本年度は、昨年度に開発した試作システムを見直し、実用化のためにさらに改良を加え、実用的なデータベース検索支援システムの要件を抽出することが主な調査内容となった。

調査研究の結果、検索支援に必要な知識の枠組みを整理し、実際のシステム上でその効果を確認することにより、検索支援システムの要件を詳細なレベルで議論することができた。また、人工知能技術としては最近の技術である事例ベース推論の仕組みを知識ベースに採り入れ、人工知能技術の利用という面でも、先進的な試みを行うことができた。

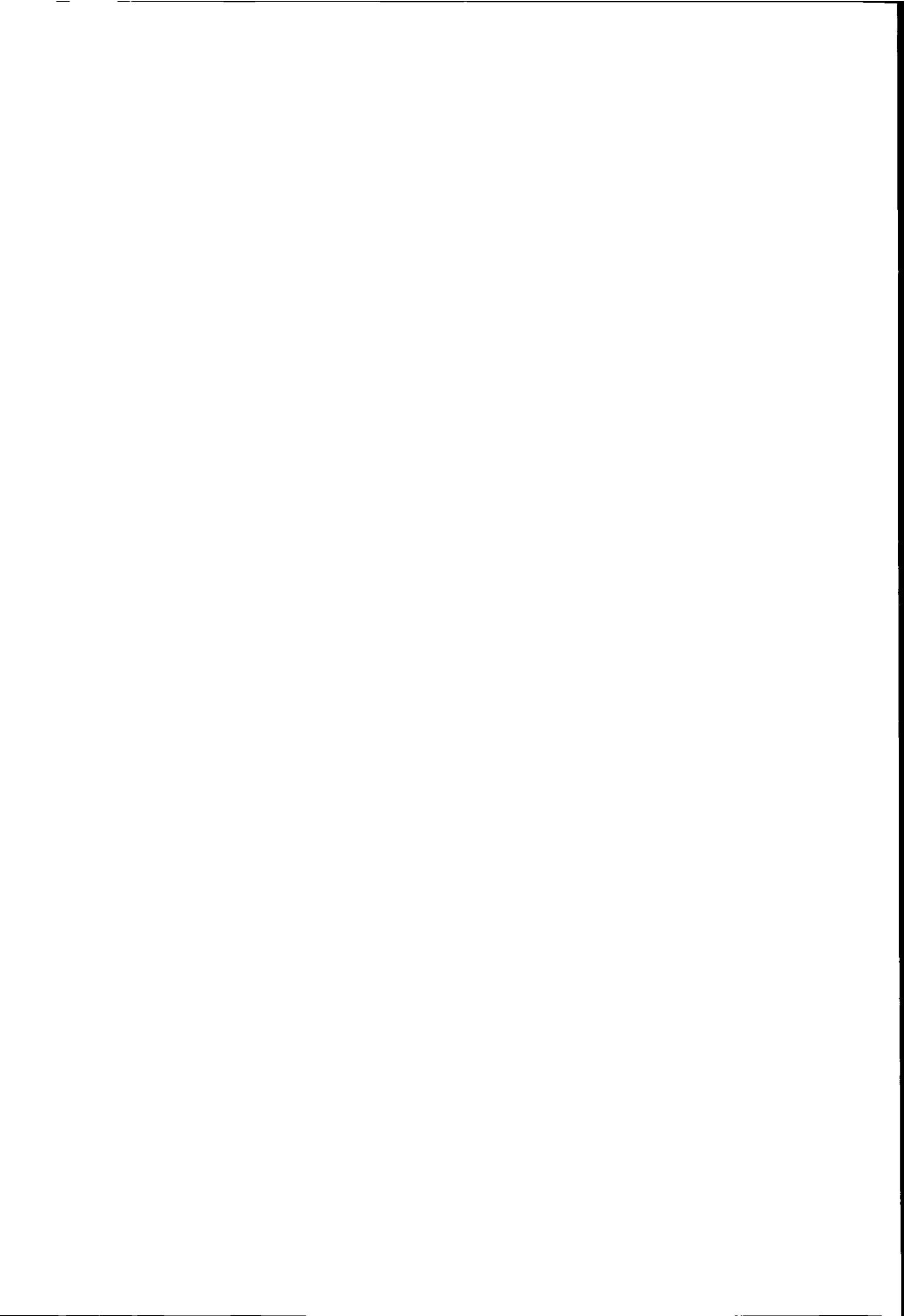
このような試みのなかで、新たに知識ベースのメンテナンスの問題や、データベースに関する書誌的な情報の取扱等実用化のための課題が明確になってきた。

3年間の調査研究を総括すると、経済統計データベースを対象とした検索支援の試みとして、具体的なシステム開発や海外での研究事例を含むまとまった報告書を作成することができた点に、本調査研究の意義があるものと思われる。

理想的な検索支援システムの実現には、尚一層の努力と関係各方面への働きかけが必要であるが、本調査研究が実用化への試みに対して何らかの示唆を与えるものとなれば幸いである。

3年間の調査研究の終了に当たり、調査研究に多大なご協力を戴いた委員各位に改めて謝意を表するとともに、今後のデータベース利用促進のために、引き続きご指導ご協力を賜るよう切にお願いしたい。

付録



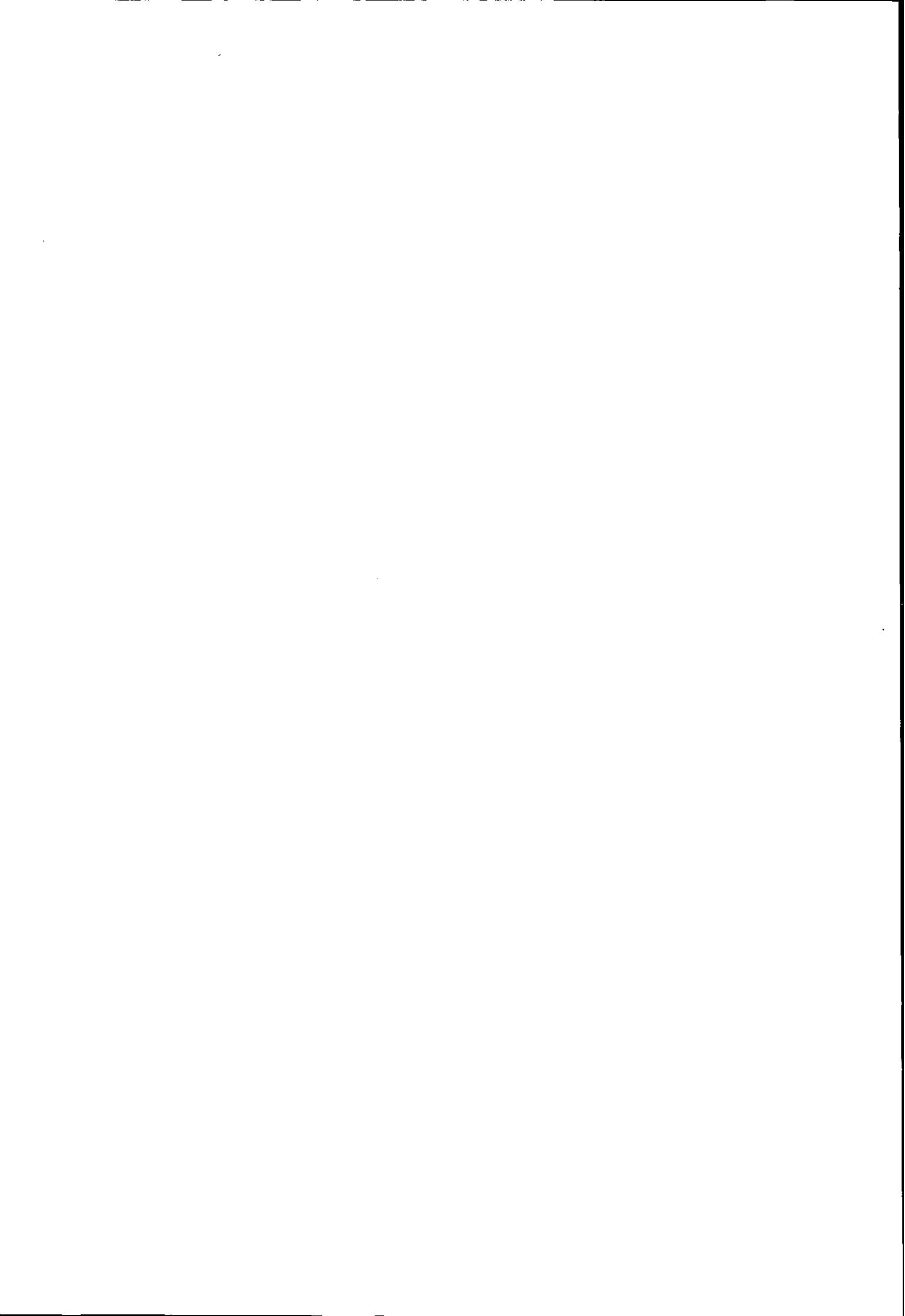
付録1 ノシステム拡充仕様書

付録2 ネットロにおける貿易統計データベースの工夫

付録3 分類体系の問題

付録4 Frame Work of COMPASS

付録5 システムテスト概要



付録1

システム拡充要求仕様書

1. システムの概要

(1) システムの目的

本システムは、経済統計のデータベースを、多くのユーザが容易に検索できることを目的として開発する検索支援システムである。

結果の分析やレポート作成に関する便宜は考慮するが、本システムの主たる目的は、複数のデータベースのなかからユーザの要求に合ったものを選択することであり、検索結果の加工分析に関しては、本システムの範囲外とする。

(2) 対象ユーザ

本システムの対象ユーザは、主に調査を目的として、貿易統計を中心とする経済統計データを検索したい人とする。あるテーマに関して、それを示す的確なデータを必要としている人々である。

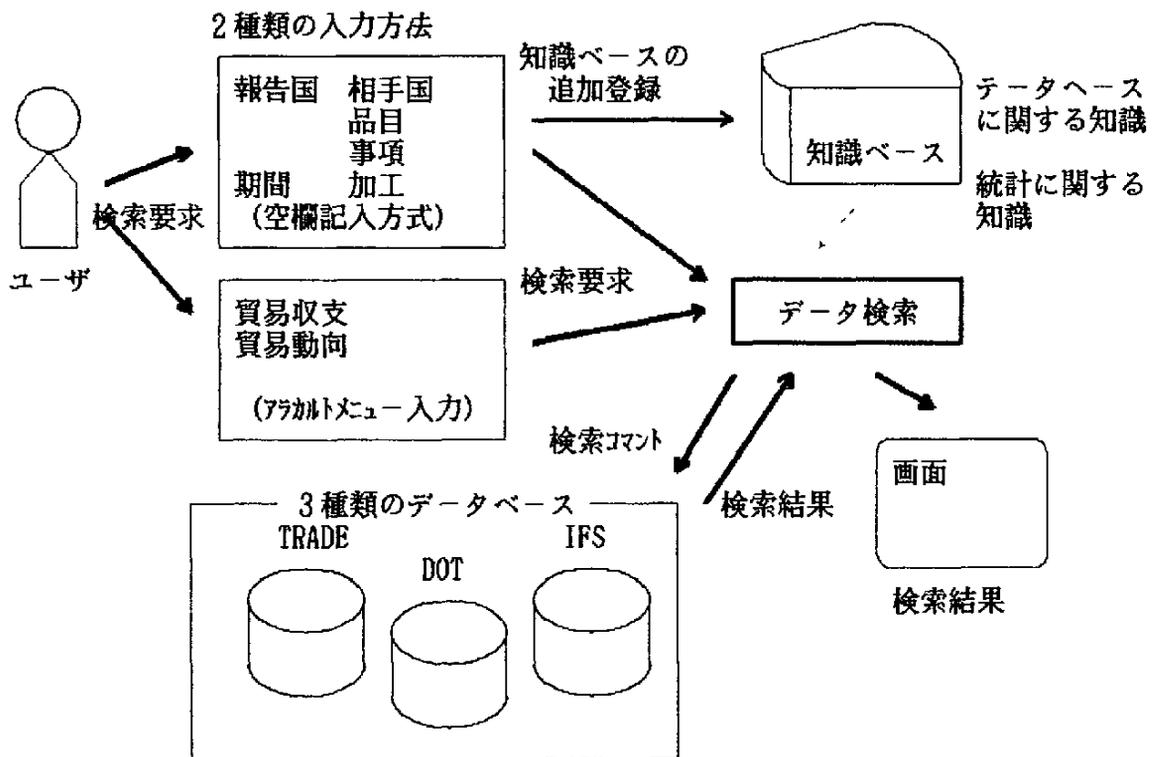
データの所在に関する知識は、ユーザが持っている必要はない、またデータベースの検索コマンドに関する知識も必要としない。

経済指標等の算出方法に関しても、システムがサポートするので、自分の調べたいテーマを、空欄記入方式と、アラカルトメニュー入力方式のどちらかの方法で、入力できればデータ検索ができる。

(3) システムの機能

本システムでは、ユーザの検索要求を入力するために、2種類の入力方式を用意する。

入力の後、データベース及び統計に関する知識を用いて検索要求を分析し、3種類のデータベースのなかから適切なものを選択して、検索コマンドを生成し、データベースへのアクセスを行い、検索結果をユーザに示す。(下図参照)



IRS概念図

(4) システムの特徴

本システムの特徴として下記の3点が挙げられる。

- ①ユーザが、データベースを意識せずに検索できること。
- ②システムが、データの所在に関する知識や、次善の情報の生成機能を持っているため検索失敗が少ない。
- ③運用中に新たな知識を獲得する機能を持つ。

2. 知識ベースの拡充

実用的なシステムへの課題を抽出するために、範囲を限定して、類義語や相互関連に関する知識を中心として、知識ベースの拡充を図る。

(1) 知識ベース拡充の狙い

- ・ 実用版に必要な類義語の量に関して、目安を得る。
- ・ 次善情報を生成するために必要な用語間の関連に関する知識を収集する。
- ・ データベースの選択に必要な知識に関して、複数のデータベースか検索可能である場合も含めて、検討を行う。
- ・ アラカルトメニューとして用意する検索要求の種類に関して、どの程度の種類があればユーザが満足するかについて検討を行う。
- ・ 生産や、金融、労働等、貿易統計以外の分野との関連性について、どのような定義をしておけば、対応が可能か検討を行う。

(2) 拡充する知識の範囲

品目
← 一般の商品 →

	電気 機器	約60品目
--	----------	-------

△1

国名

全世界

事項

輸出・輸入・貿易量
双曲化指数
輸出_{レ7} 輸入_{レ7}
GNP
GDP
輸出・輸入・貿易依存度
産業内分業度指数
為替レート
輸出・入価格指数

海外資産
海外負債
外貨準備
国民総生産 (GNP)
国民総支出
人口
以上15種類

(分野の拡張の検討をおこなうために、商品以外の国際取引に関する項目を追加する)

加工

原データ
指数
前年同期比
同期比
対輸入比
三項移動平均

成長率
対GNP比
所得弾性値
実質値
対輸出比

以上11種類

△2

期間

最新値
最新10年間の推移
指定時点の値
指定期間の推移

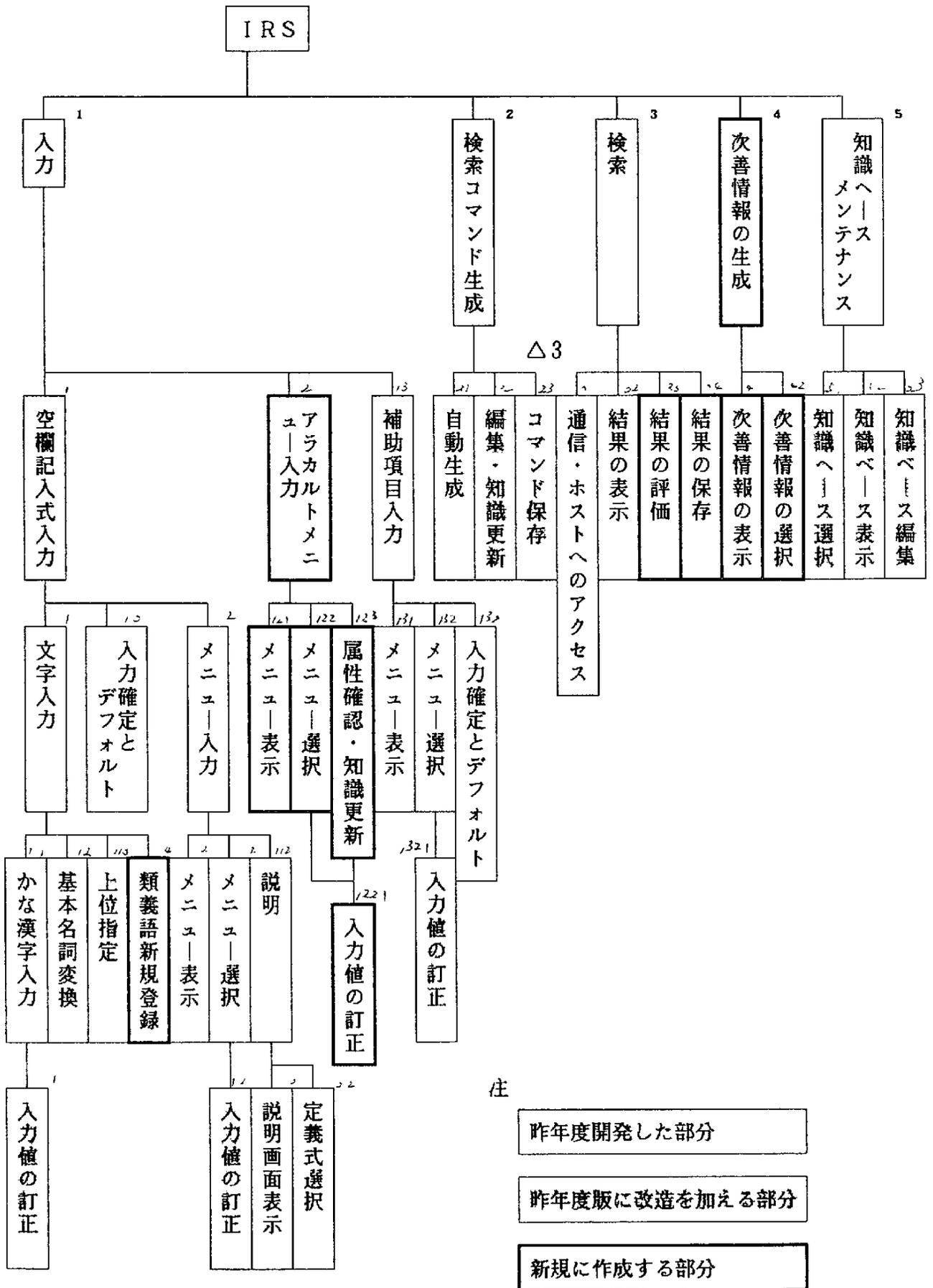
以上4種類

アラカルトメニュー

内容は今後検討するが、国際取引の範囲で最大20件を用意する。

3. 機能構成

IRSの機能構成は、下図のようになっている。



機能概要一覧表 (○:新規 △:改造)

番号	機能の名称	機能の概要
1	入力	検索要求の入力
11	空欄記入式入力	属性による入力(報告国、相手国・・・)
111	文字入力	報告国、相手国、品目の入力
1111	かな漢字入力	キーボードからのかな漢字による入力
1112	基本名詞変換	ユーザの入力を類義語辞書を使って基本名詞に変換
△ 1113	上位指定	上位10か国、上位10品目の指定
○ 1114	類義語辞書新規登録	ユーザが類義語辞書にない用語を入力した場合の、登録機能
△ 11111	入力値の訂正	ユーザがすでに入力した属性に対する再入力
112	メニュー入力	事項、期間、加工の入力
1121	メニュー表示	入力可能な用語一覧のメニュー形式表示
1122	メニュー選択	メニュー表示された用語のなかから、検索要求に合うものを選択する機能
△ 11221	入力値の訂正	メニュー選択した後に、選択を変更する機能
1123	説明	メニュー表示された用語の意味や用法に関する説明
11231	説明画面表示	説明分の画面への表示
12232	定義式選択	定義式や、収録条件の違いにより、複数のデータが該当する場合にユーザに選択させる機能
○ 113	入力確定とデフォルト	入力値の確定と、デフォルトの働き
○ 12	アラカルトメニュー入力	アラカルトメニューによる要求入力
○ 121	メニュー表示	アラカルトメニューの表示
○ 122	メニュー選択	アラカルトメニューの選択
○ 123	属性確認・知識更新	アラカルトメニュー選択による入力を、報告国、相手国・・・の属性表現になおしてユーザに示し、確認する。ユーザが修正した場合知識ベースを更新する。
○ 1221	入力値の訂正	メニュー選択の後、あるいは、属性確認の後、入力値の訂正を行う
13	補助項目入力	時点、期種等、補助的な検索要件の入力
131	メニュー表示	補助項目入力のためのメニュー表示
132	メニュー選択	補助項目入力のためのメニュー選択
△ 1321	入力値の訂正	補助項目の入力に対する入力訂正
133	入力確定とデフォルト	補助項目の入力確定とデフォルト値の働き
2	検索コマンド生成	ユーザの検索要求から検索コマンドを生成する機能
21	自動生成	検索コマンドの自動的な生成
△ 22	編集・知識更新	生成された検索コマンドのユーザによる修正と知識ベースの更新
△ 23	コマンド保存	生成された検索コマンドの保存
3	検索	データヘース検索
31	通信 ネットへのアクセス	パソコンとホスト間の通信とホスト側データベースへのアクセス
32	結果の表示	検索結果の画面表示
○ 33	結果の評価	検索結果に対して、①検索成功、②欠測 ③コマンド誤り、④ユーザの要求に合わない の判定をおこない必要な場合再検索を行う機能
○ 34	結果の保存	検索結果をユーザファイルに保存する機能
○ 4	次善情報の生成	何らかの原因で、ユーザの希望するデータがない時に次善のデータを提供する機能
○ 41	次善情報の表示	次善情報をユーザに示す機能(次善情報は複数)
○ 42	次善情報の選択	次善情報のなかからユーザの要求に近いものを選択する機能
5	知識ヘースメンテナンス	知識ヘースの管理のための機能
△ 51	知識ヘース選択	編集する知識ベースの選択
△ 52	知識ベース表示	知識ヘースの内容の表示
△ 53	知識ベース編集	知識ヘースの修正

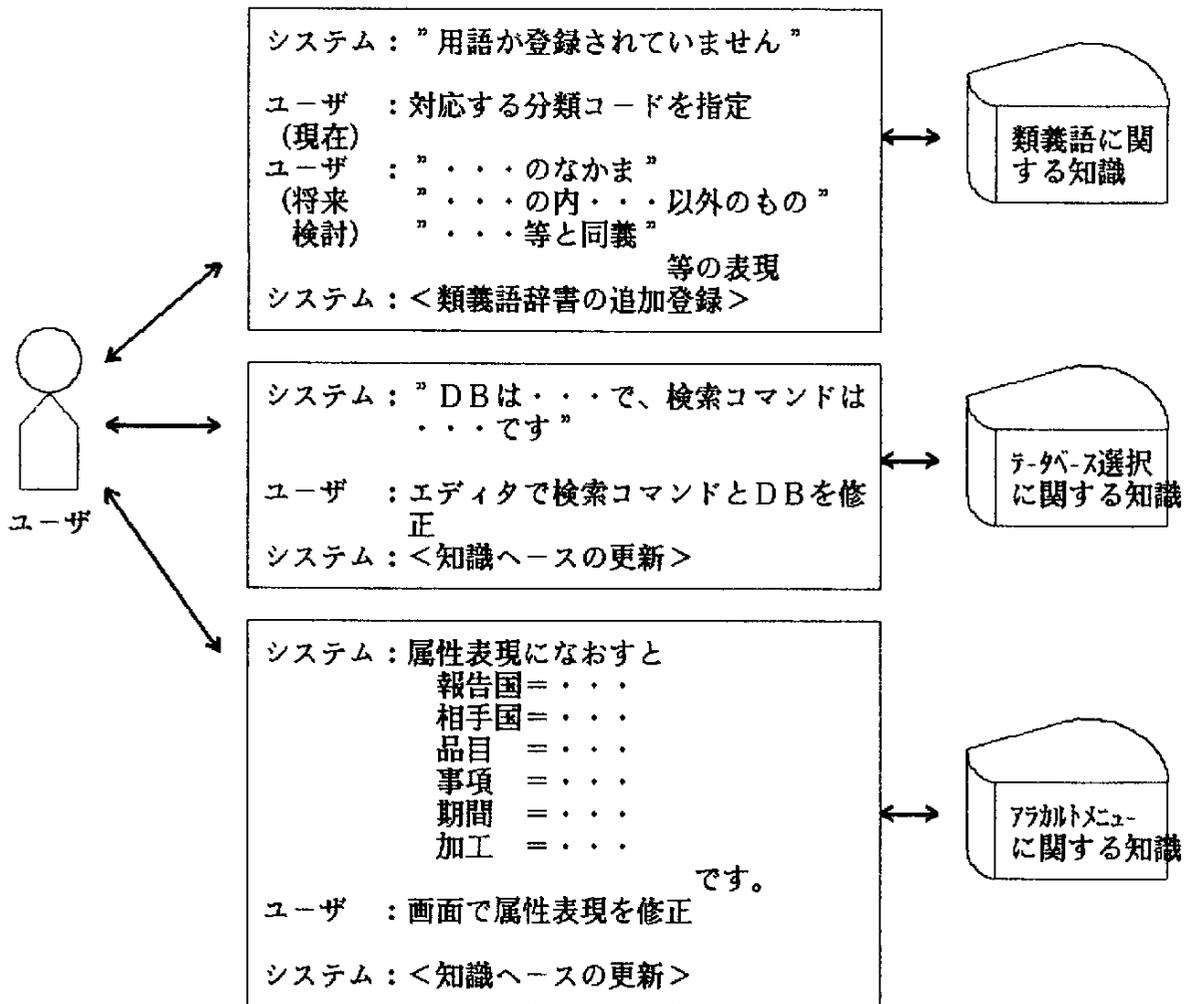
△ 4

4. 知識獲得の概要

知識獲得とは、ユーザからの指示をもとに新たな知識を知識ベースに追加する機能と、ユーザからの指示をもとに、既存の知識を修正する機能を指す。

知識獲得の対象となる知識は、①類義語に関する知識、②データベース選択に関する知識、③アラカルトメニューに関する知識の3種類である。

以下に、各知識毎の、知識獲得の概念図を示す。



知識獲得の概念図

5 1 知識獲得に関する問題点

知識ベースの管理の方法

知識ベース（類義語辞書等も含む）をどこに置くか（ホストか端末か）という実現方法の問題と、ユーザ管理の問題の2面がある。

実現方法に関しては、物理的制約に左右される面もあるので、ユーザ管理の方法に関する案を述べ、本年度の実現方法を説明する。

(1) 知識ベースの種類と分類

I R Sの知識ベースを更新の形態別に分類すると次のようになる。

（☆は、本年度あらたに作成する知識ベース、その他は昨年度の報告書で説明したもの）

①ユーザが運用中に更新するもの

- I 類義語辞書
- ☆II アラカルトメニュー対応辞書
- ☆III 意味階層辞書
- IV 導出辞書
- V 導出式辞書
- ☆VI 検索事例ベース（DBの選択とコマンドの生成に関する辞書）

注. I、II、VIについての更新イメージはp 9に示したとおり、

III、IV、Vについては更新方法は決定していないが、ユーザ更新の対象とする。

②知識ベース管理者が定期的に更新するもの

- I DB属性辞書
- ☆II 事項・加工対応辞書
- III 説明辞書
- IV DBコード対応辞書

③システムの拡張の場合に更新が必要になるもの

- I 属性定義辞書
- II 属性関連辞書
- III DB種別辞書
- IV 入力枠定義辞書
- V メニュー定義辞書
- VI ウィンドウ定義辞書

(2). 知識ベース管理の段階

①基本知識投入段階（全ての辞書に共通）

- ・知識ベースの基本的な内容を入力する段階。
- ・マニュアルや、仕様書の範囲で、確定している知識の入力。
- ・ユーザに対する事前調査で、確定した知識の入力。

②知識収集段階（類義語辞書のみ）

類義語辞書に関して、ユーザの共通的な知識を収集する。

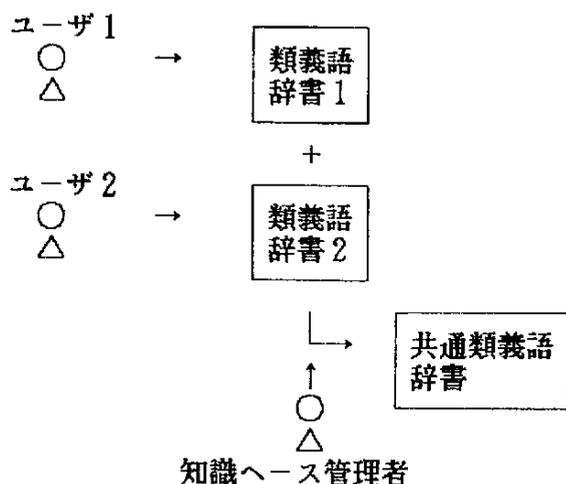
I 限定したユーザに、集中的に利用してもらい（あるいは直接基本名詞・類義語対応表を埋めてもらい）、類義語を収集する。

II ユーザ毎に収集された類義語の和集合をとる（自動的処理が可能）。

III IIの類義語の中から、異常なものを取り除く（知識ベース管理者が行う）。

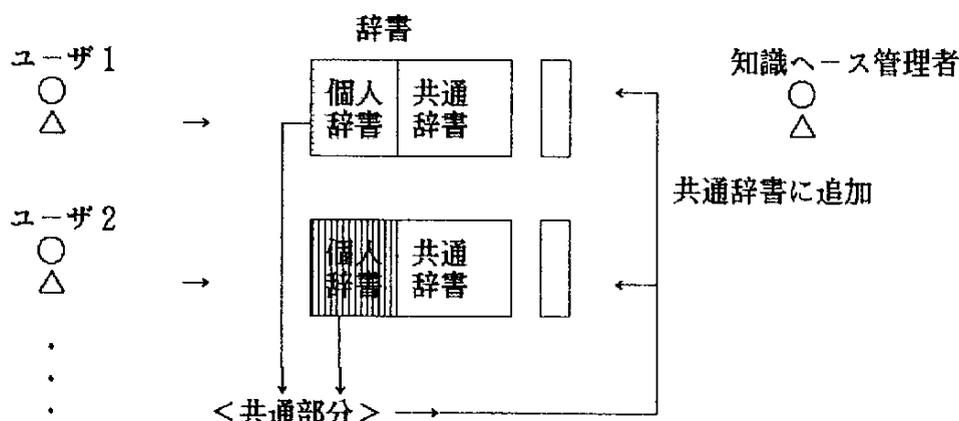
IV IIIの類義語の中から、複数の基本名詞に対応するものを抽出し、プライオリティ付けをする（知識ベース管理者が行う）。

次ページの図参照



③ 運用段階 (ユーザ管理段階) (ユーザが運用中に更新する知識のみ)

- ・運用に伴いユーザの便宜のために、ユーザ辞書を作成する。
- Ⅰ ユーザが共通辞書にない語を登録した場合、個人辞書として蓄積される。ユーザは、個人辞書と共通辞書を合わせた辞書を辞書として使うことになる。(検索は、個人辞書を先に、共通辞書を後に行う)
- Ⅱ 知識ベース管理者は、定期的にユーザ毎の個人辞書を収集し、共通部分を共通辞書に繰入れる作業を行う。



(3). 本年度の実現方法

- ・本年度の環境はホスト1台、端末1台
- ・個人辞書の部分をユーザがフロッピーディスクで管理する。
- ・知識ベース管理者は、当面政策情報システム部の委員が担当する。
- ・知識収集段階では、品目は、電気機器に限定し、ワーキンググループのメンバーをユーザとして、作業を行う。
- ・運用段階では基本的に全ての委員がユーザになる。(8月中旬以降利用可能)
- ・運用段階での知識ベース管理は、知識ベース管理者が、ユーザのフロッピーを管理し、定期的に回収し、共通部分を抽出し(システム化可能)共通辞書に追加するという手順で行う。

5. 2 次善情報に関する問題点

(1) 次善情報の定義に関する問題

次善情報の出し方として次の2種類がある。

① 導出辞書的次善情報

導出辞書に書かれた用語の定義に従って、ユーザが指定した項目が検索不可能な場合に下位の概念の組合せにより、次善情報を合成するものである。次善情報ではあるが、ほとんど同義の情報といえる。

②上位概念的次善情報

ユーザの検索要求とは異なるが、①ユーザの指定した項目を含んでいる、②ユーザの示した項目と同じ傾向を示す等の理由により代用品として認知される情報である。

IRSでは、①のアプローチで、対応できない場合は、②のアプローチで、次善情報を提供するのが望ましいと思われる。

例

国グループの呼称に関しては、ユーザが下位概念によって定義可能なら①で解決できる。(西側先進国は米国とドイツと・・・というように)

ユーザが下位概念によって表現できないときに、②が採用される。
(西側先進国の代わりに、先進工業国)

(2). 次善情報が一意的に決まらないという問題

一般的に上位のレベルに関しては用語の意味を決定する下位要素の組合せが複数考えられる。

本年度は複数の次善情報をユーザに示し、選択させる方法で、解決し、ユーザの満足度を評価する。(案)

(3). 統計分類の違いにより言葉の意味としては近くても集合が異なる場合

△8

カラーテレビの例のように、統計によって、対応する集合が違う場合に、次善情報として適切か否かを判断する必要がある。

5. 3 統計分類に変更があった場合の問題点

国名や品目分類等、基本名詞の定義に変更があった場合の類義語の位置付け
(ミャンマーの例に見られるようなケース)

基本名詞の変更は、DB更新時におこない、それまでは類義語扱いとする。
変更前後の時系列を連続して検索できるかどうかというDB側の問題もある。

5. 4 入出力方法の問題

昨年度の試作システムはマウスとキーボードの併用による入力をおこなっていたが、本年は、キーボードのみの入力を試みる。キーボードのみを使う場合、マウスのみを使う場合、両者を併用する場合の利害得失は下表のように予想される。

評価項目	キーボードのみ	マウスのみ	併用
初めて使う場合	キーボードを使う機会が多いので、あまり抵抗かない	使ったことのないユーザが多い。	操作が煩雑たという印象を持つ恐れあり。
習熟の早さ	キーが限定していれば早いですが、日本語変換等は時間がかかる。	早い。	時間がかかる。
日本語入力	自由に入力ができる。	メニュー選択方式でなら入力が可能	自由に入力ができる。
メニュー選択	指示したい項目に移動する場合、何回もキーを叩く必要がある。	好きな場所へ、簡単に移動ができる。	好きな場所へ、簡単に移動ができる。
ウィンドウ操作	ウィンドウの大きさや、位置の任意指定には不便。	大きさや位置の指定が、らくにできる。	大きさや位置の指定が、らくにできる。

システム拡充詳細仕様

1. 入力機能

ユーザの検索要求を入力するために、IRSでは、空欄記入方式とアラカルトメニュー入力の2種類の方法が用意する。

空欄記入式入力画面	
(報告国) の	(相手国) に対する
(状態) を	アラカルトメニュー画面
	地域別貿易動向 排気量別乗用車輸出 商品別貿易動向 台数増減率 貿易収支の推移 鉄鋼輸出の推移 財別貿易動向 部品輸出の推移 品目別輸出数量の推移
↑キーで入力方式を選択して下さい	

初期画面例

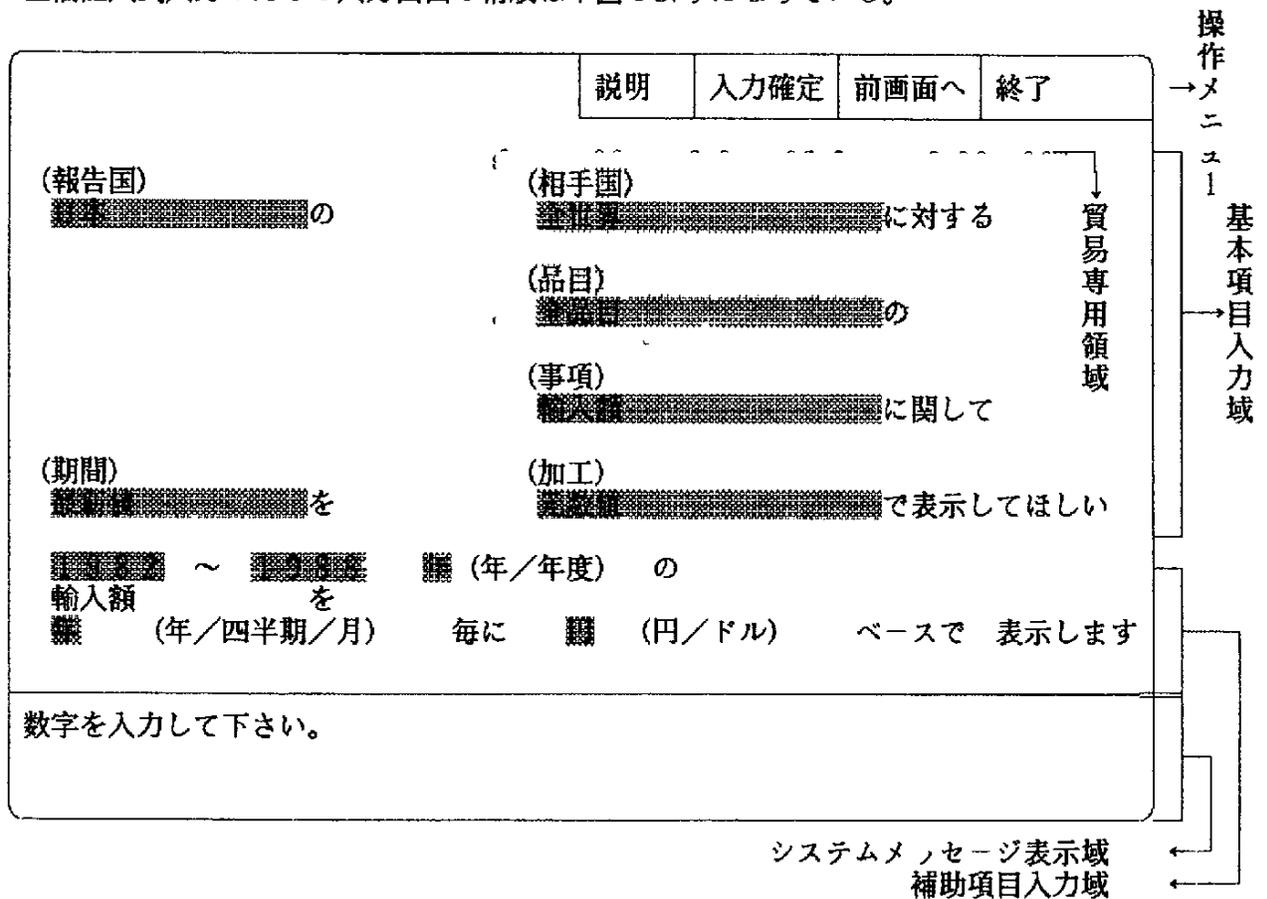
また、時点や、期種等の細かな項目を入力させるために、補助項目入力機能がある。

11. 空欄記入式入力機能

入力機能のなかで、報国情、相手国、品目、事項、期間、加工の各属性に関して、空欄記入方式で、入力を行う方法を、空欄記入式入力と呼ぶ。

(1) 入力画面

空欄記入式入力のための入力画面の構成は下図のようにになっている。



- ・操作メニューは、ユーザがシステムの操作を指示するために使う。
- ・基本項目入力域は、検索要求の基本項目である相手国、報告国、品目、事項、期間、状態の各属性を入力するために使う。
- ・基本項目入力域の中の、貿易専用領域は、特に貿易特有の入力の場合のみ使われる領域で、それ以外の場合は、マスクがかかった状態になっている。
- ・補助項目入力域は、基本項目の入力が終わった状態で、指定時点や、期種等の確認のためにユーザに表示される入力領域で、入力された基本項目から考えられる、値がセレクトされている。
- ・システムメッセージ表示域は、システムからの操作指示等を表示するための領域で、必要に応じて、種々のメッセージが表示される。

(2) 入力手順

ユーザは先ず基本項目入力域に基本的な検索要求を入力する。入力は全てキーボードによる。

報告国、相手国、品目の入力は、文字入力方式による。

事項、期間、加工は、メニュー選択方式による。

基本項目の入力が終了したら、入力確定の操作アイコンにカーソルを移し、実行キーを押す。入力がない場合はデフォルト値が入力値と見なされる。

(3) 属性入力の整合性チェック機能。

事項の性質により、各属性は自ずと制約条件を持つ。(たとえば、GNPに関する、検索要求の場合、相手国、品目の概念は無意味である)

矛盾した入力避けるために、

a. 事項が入力されている場合は事項が最優先

b. 事項以外の属性に関しては、入力が新しいものが優先

という規則に従って、入力の誘導を自動的におこなう。

111 文字入力

報告国、相手国、品目はキーボードからの文字入力により入力する。

1111 かな漢字入力

ユーザが、報告国、相手国、品目の各欄に、デフォルト以外の値を入力したいときは、カーソルをそれぞれの入力枠に移動し、かな漢字で入力する。

また、'、' または、'、' または、'、' をセパレータとして、複数項目の入力が可能である。(例 日・米・ソの 韓国 からの 製品輸入 等)

基本的にどの項目も複数入力が可能である。

事項として品目や相手国に意味のあるものとなないものが両方指定された場合は、ユーザの指定する品目や相手国は、それらが意味を持つ事項にかかるものと解釈される。

例

日・米 の 東ドイツへの 鉄鋼の 輸出額、GNP を知りたい

→ 日・米 の東ドイツへの 鉄鋼の輸出額 と
日・米 の GNP を知りたい。

11111 入力値の訂正

文字入力の機能によって入力した入力値を訂正したい場合は、再度カーソルを訂正したい場所へ移動して実行キーをおすことによって、訂正可能になる。

但し、先に入力した文字列の一部を修正することはできない。最初から入力をしなおすことになる。

1112 基本名詞変換

ユーザが入力した国名または品目は、類義語辞書を使って、基本名詞に置き換えられ、基本名詞が、ユーザ入力に代わって入力枠に表示される。

基本名詞とは、国名、品目の標準的な呼称のことで、基本名詞と、各データベースへのコード対応表が作成されている。

例 報国国の入力の際の画面の遷移

- ① の (デフォルトの状態)
- ② の (ユーザの入力)
- ③ の (基本名詞への変換)

1113 上位指定

国名と品目には、'上位10品目(か国)'の指定が可能である。
但し上位10品目(か国)の指定は、全体で、1度しか使えない。(即ち、報告国を上位10か国、品目を上位10品目とすることはできない)。ユーザかすでに上位・・・の指定をしているとき、さらに上位・・・の指定を行うと、先に指定したものが、デフォルト値に自動的に変換される。

1114 類義語新規登録

ユーザか、類義語辞書にない用語を入力した場合、システムは、ユーザに対応コードを入力させ、類義語への新規登録を行う。ユーザか、一端登録をおこなった用語は、類義語辞書に保存されるので、次回からは、通常の入力方法で、入力が可能になる。

例 報告国に新しい用語を使った場合

の (ミャンマーが類義語辞書にない)

'ミャンマー'は、辞書に登録されていません。TRADEの国名コードを入力してください。
>>

(システムメッセージとコード入力のためのプロンプト表示)

'ミャンマー'は、辞書に登録されていません。TRADEの国名コードを入力してください。
>>122

(ユーザが入力)

の (対応する基本名詞を表示)

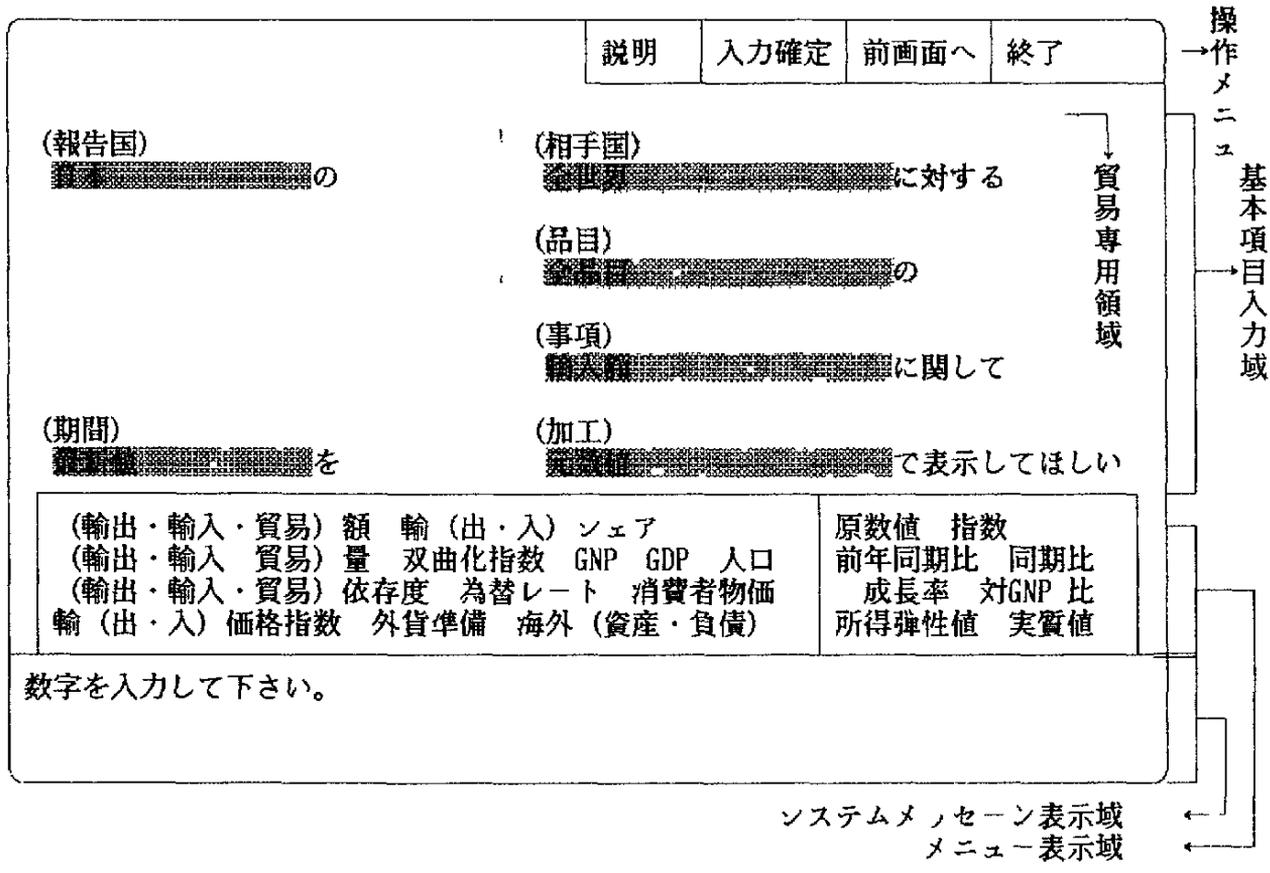
同様に品目に関しては、概況分類のコード指定により、新たなコードを登録する。

112 メニュー入力

事項、期間、加工は、メニュー選択方式で入力する。

1121 メニュー表示

メニューは、初期画面から表示されている。
指定は、カーソルの移動によっておこなう。



期間の選択肢は、最新値、最近10年の推移、指定時点の値、指定期間の値の4通りである。

事項と加工は、連動的なメニューになっており、ユーザが事項を指定することにより、選択可能な加工のみがユーザに示されるようになる。下記に事項と加工の対応表を示す。

事項	加工											備考 (DB名)
	原数値	指数	前年同期比	同期比	成長率	対GNP比	所得弾性値	実質値	対輸入比	対輸出比	三項移動平均	
輸出 輸入・貿易額	○	○	○	○	○	○	注				○	TRADE, DOT IFS
輸出・輸入・貿易量	○	○	○	○	○						○	TRADE
双曲化指数	○	○	○	○	○						○	DOT
輸出 ₇ 輸入 ₇	○	○	○	○	○						○	DOT
GNP	○	○	○	○	○			○			○	IFS
GDP	○	○	○	○	○			○			○	IFS
輸出輸入貿易依存度	○										○	IFS
産業内分業度指数	○										○	IFS
為替レート	○	○	○	○	○						○	IFS . AE.
輸出・入価格指数	○	○	○	○	○						○	IFS
消費者物価	○	○	○	○	○						○	IFS
外貨準備	○	○	○	○	○				○		○	IFS
海外資産	○	○	○	○	○						○	IFS
海外負債	○	○	○	○	○	○				○	○	IFS
人口	○	○	○	○	○						○	IFS

注. 輸入の所得弾力性のみ

1122 メニュー選択

表示されたメニューの中から、指定したいものにカーソルを移動し、実行キーで、選択する。

11221 入力値の訂正

メニュー選択の場合

メニュー選択による入力後、値を修正したい場合は、再度カーソルを入力枠に移動して、メニューのなかから選択することができる。

先に入力されていた値は、キャンセルされる。

複数入力の一部だけを入れ換えたい場合は、最初から入力をしなおす必要がある。

1123 説明機能

メニュー選択方式の入力の場合、ユーザは、'説明'メニューを指定することにより用語の意味の説明を得ることができる。

11231 説明画面表示

右上の説明メニューにカーソルを移動して実行キーを押すと、説明画面が表示され、用語に関する説明や、類義語が表示される。

説明画面は、実行キーを再度押すと、消去される。

11232 定義式選択

経済指標等の場合、定義式が複数表示される場合があります、この場合ユーザは、どちらの定義式を使うかの選択を行うことができる。(ユーザが選択しなければデフォルトの定義式が使われる)。

113 入力確定とデフォルト

画面右上の入力確定メニューにカーソルを移動して、実行キーを押すと、その時点での入力値によって、検索要求が確定する。

入力終了していない属性に対しては、画面に表示されているデフォルト値が入力値として採用される。

入力確定した直後に、'前画面へ'のメニューにカーソルを移動し、実行キーを押すと、確定は解除され、入力可能な状態にもどる。

デフォルト値は、初期はp10の値がセトされているが、2回目以降は前回利用時 △9の最終状態がセトされている。

12 アラカルトメニュー入力機能

空欄記入式以外に対して、使用頻度の高い定型的な検索要求を入力するための入力方法がアラカルトメニュー入力機能である。

121 メニュー表示

① アラカルトメニューは、空欄記入式入力画面と同時に表示され、ユーザは、どちらかの入力方式を選択することができる。

② アラカルトメニューは、属性による表現よりも、より、ユーザの要求に近い形式の入力を許すもので、文章の形式で、現されたメニューのなかから、ユーザが選択すると、システムが自動的に属性に展開する機能である。

③ アラカルトメニューとして用意されているものは、次の各項目である。

アラカルトメニュー

項目名	内容
地域別貿易動向	世界、先進国、アメリカ、日本、西ドイツ、発展途上国、共産圏の輸出、輸入額の前年比伸び率(%)を過去3年分表形式で出力する。
商品別貿易動向	全商品、農産物、鉱産物、工業製品に関して、輸出、輸入額の前年比伸び率(%)を過去3年分表形式で出力する。
貿易収支の推移	輸出額、輸入額、収支を過去10年分折れ線グラフで出力する。(報告国はユーザ入力)
財別貿易動向	輸出額、輸入額の伸び率及び、その財別内訳(資本財、耐久消費財、中間財、その他)を過去10年分棒グラフで出力する。
部品輸出の推移	輸出総額に占める部品輸出の割合を、アメリカ、EC、アジアNIE S、ASEAN諸国に関して過去10年分折れ線グラフで出力する。
日本の品目別輸出数量の推移	食料品、繊維及び同製品、化学製品、非金属鉱物製品、金属及び同製品、機械機器(一般機器、電気機器、輸送機器)、及び全体に関する輸出数量の伸び率を、過去3年分出力する。
排気量別乗用車輸出台数増減率	2千cc超、千~2千、千以下の乗用車の輸出台数増減率(%)を最新時点に関して棒グラフで出力
鉄鋼輸出の推移	輸出数量、ドルベース輸出価格、円ベース輸出価格、ドルベース輸出額、円ベース輸出額の前年同期比を、4半期毎に最近4年分グラフ表示する。

注。これらの項目は、通商白書を参照して作成した。

122 メニューの選択

指定したいタイトルにカーソルを移動して実行キーを押し選択する。

123 属性確認・知識更新

ユーザがアラカルトメニューを指定すると、属性の入力された空欄記入式入力画面が表示される。ユーザはこの画面から属性の変更を行うことができる。

ユーザが変更をおこなった場合、知識ベースは、変更後の属性の組合せに置き換えられる。

13 補助項目入力

昨年度の試作システムと同様

2. 検索コマンド生成機能

データベースの選択と検索コマンドの生成

対象とするデータヘースは、TRADE、DOT、IFSの3種類で、検索要求に合わせてシステムが自動的に適切な(検索可能な)データベースを選択する。

複数のデータヘースが検索可能な場合で、優劣のつけがたい場合は、ユーザへの問い合わせが行われ、データベースの選択を指示する。

生成された検索コマンドは、ユーザに示される。

ユーザは、それを見て、変更したいときには、エディタを使って修正することができる。ユーザの修正した検索コマンドで、検索に成功すれば、修正された検索コマンドが新たに登録され、次回からは、新しい方の検索コマンドがユーザに示される。

21. 自動生成

検索コマンドの自動生成機能とは、ユーザの入力に合わせて、システムが、予め用意されている、検索事例を使って、自動的に検索コマンドを生成する機能である。

ユーザが検索要求の入力を終了した時点で、データヘースの選択と、検索コマンドの生成が行われる。

データベースの選択は、次の手順でおこなわれる。

① ユーザの入力した属性の組合せから、検索可能なデータベースが一つだけ見つかる場合は、そのデータベースに決定する。

② ユーザの入力した属性の組合せから、検索可能なデータヘースが複数みつかる場合は、(輸出額等は、3つのデータベースで、検索が可能)

a. 最新値が求められる時は、収録時点の新しいものの優先とする。

b. a 以外の場合は、データの性質をユーザに示し、ユーザに選択させる。

22. 編集・知識更新

検索コマンドは、テキストファイルの状態にユーザに示される。

ユーザは、システムに組み込まれたエディタを使って修正することができる。

修正の結果、検索が成功した場合は、修正された検索コマンドが保存され、次回からの検索に使用される。

23. コマンド保存

生成された検索コマンドは、ユーザフロッピーに保存される。また、このファイルは、ACOSの自動実行ファイルになっているので、ユーザは後から、同じ検索手順で、再検索を行うことができる。

△7

3. 検索

ACOSのデータベースへのアクセスを行い、検索結果を画面上に表示する。

さらに、検索の結果が成功したか否かの判断をおこない、失敗の場合に次善情報を生成する等の措置をとる。

31. 結果の表示

ホストから送られる検索結果は、画面に表示される。

32. 結果に関して、次の各ケースが考えられる。

① 検索に成功した場合

② 検索に失敗した場合(該当データがない場合)

③ 検索に失敗した場合(②以外の理由によるもの)

④ 検索結果にユーザが不満足の場合

当面①、②、③、④の区別は、ユーザが判断して入力する。(エラーメッセージの判読が機械的に可能であれば、自動判定に移行する)。

②、④の場合システムは、すでに用意された次善情報をユーザに示し、再検索を行う。

- ③の場合は、開発途上でおきる可能性があるが、運用時には発生しないはずである。
- ①の場合は、検索を終了し、初期画面に戻る。

33. 結果の保存

検索結果は、パソコン側のフロッピーディスクに保存する。
 フォーマット等を考慮すれば、ロータス等の表計算ソフトでの利用が可能になるか、
 当面の課題としては、何らかの形式で、保存することを目標とする。

4. 次善情報の生成

検索結果で、再検索の必要が認められた場合、次善情報が生成される。
 次善情報は、システムが用意する、意味階層を使って、生成される。
 一般的には、複数の次善情報が生成され、ユーザに選択を促す。

意味階層は、次善情報を提供するための知識構造で、属性によって次のように定義した次善情報を、蓄える。

- ①国、品目に関しては、分類コード上の上位概念を次善情報とする。
- ②事項に関しては、意味的な関連概念を収集して、次善情報とする。

例

電卓の次善情報 → 一般機械 等

GNPの次善情報 → GDP 等

41. 次善情報の表示

下記の画面イメージで、ユーザに次善情報が示される。

	説明	入力確定	前画面へ	終了
<p>(報告国) [] の</p> <p>(期間) [] を</p>	<p>(相手国) [] に対する</p> <p>(品目) [] の</p> <p>(事項) [] に関して</p> <p>(加工) [] で表示してほしい</p>			↓ 貿易専用領域
<p>① 日本の 先進工業国に対する 電卓類の 輸入額に関して 最新値を 原数値で ② 日本の 先進国に対する 電卓類の 輸入額に関して 最新値を 原数値で ③ 日本の 全世界に対する 電卓類の 輸入額に関して 最新値を 原数値で</p>				
<p>次善情報を選択してください</p>				

42. 次善情報の選択

ユーザは、画面に示された、次善情報のなかから、自分の要求に最も近いものを選択し指示する。

さらに、その次に近いものを指示する。

システムは、先ず最も近いとされる次善情報で、検索を試み、再度失敗した場合に次の候補を検索する。

それでも失敗した場合は、今のところ、検索コマンドをユーザに指示させる方法で、知識を獲得する。

5. 知識ベースメンテナンス

IRSでは、知識ベースは、ユーザの入力によって次第に増加してゆく。

但しユーザ任せにした場合、知識ベースに矛盾が発生する場合や、冗長な知識が詰め込まれる恐れがある。

この事態を避けるために、知識ベース管理者が、知識ベースをメンテナンスするための機能を有する。

51. 知識ベース選択

初期画面から、知識ベース選択モードを指定すると、メンテナンス可能な知識の一覧が表示され、ユーザは、そのなかからメンテナンスしたい知識ベースを選択する。

52. 知識ベース表示

知識ベースをプリントアウトする場合に使用する機能である。

53. 知識ベース編集

知識ベースをエディタを使って編集するための機能である。

システムは、知識ベースの文法チェックのみを行う。

付録2 ジェトロにおける貿易統計データベースの工夫

2.4.4 ジェトロにおける貿易統計データベースの工夫

ジェトロにおける貿易統計データの利用状況とその対応については、既に2.4.3で述べたが、こうした個々の対策は、本調査研究の目的とかなり一致する点が多い。そこで、その具体的対応の例を最後に取り上げる。

(1) 独自の商品分類の作成

貿易統計データベースの特徴として、各国ばらばらのデータに対応してゆかなければならない点が指摘されてきたが、これへの対応として、分類を探す仕組みとして、一種の細分類に関して特徴を抽出し、基準を作成し、分類表を整理した。

この例として商品分類名（履き物の例）を図2-27に示す。

商品分類名	NBDC NO	I	E	数量単位	Description
e. 発行物 ← 雑貨品中の発行物	KC000	○	○	--	Books Newspapers & Journals
— 書籍・小冊子	KC010	○	○	--	Printed Books, Brochures
— 新聞、雑誌等の定期刊行物	KC020	○	○	NO	Newspapers, Journals and Periodicals
— 設計図・図案	KC030	○	○	KG	Plans and Drawings
— 広告、販賣用紙類	KC040	○	○	KG	Advertising Material, Commercial Catalog
— 絵画、デザイン、写真	KC050	○	○	KG	Pictures, Designs and Photographs
d. 履き物	KD000	○	○	--	Footwear
— スポーツ用の履き物	KD100	○	○	DZ	Sports Footwear
— スキー靴	KD101	○	○	DZ	Ski-boots
— その他のもの	KD102	○	○	DZ	Other Sports Footwear
— ゴム、プラスチック製 (テニス用を除く)	KD103	○	○	DZ(KG)	Other Sports Footwear of Rubber, Plastic
— 革製 (テニス、バスケット用を除く)	KD104	○	○	DZ(KG)	Other Sports Footwear of Leather
— テニス、バスケット、体操用のもの	KD105	○	×	DZ(KG)	Footwear for Tennis, Basket, Gymnastics
— 革製のもの	KD106	○	×	DZ(KG)	Footwear for Tennis, Basket of Leather
— 紡織用繊維製のもの	KD107	○	○	DZ(KG)	Footwear for Tennis, Basket of Textile
— キャンパスシューズ	KD108	○	×	DZ(KG)	Canvas Shoes for Tennis, Basket
— 履き物 (スポーツ用を除く)	KD200	○	○	DZ	Footwear (excl Sports Footwear)
— ゴム・プラスチック製のもの	KD201	○	○	DZ	Footwear of Rubber, Plastics
— 防水性長靴	KD202	○	○	DZ	Boots of Rubber, Plastics
— 短靴	KD203	○	×	DZ	Shoes of Rubber, Plastics
— サンドル	KD204	○	×	DZ	Sandals of Rubber, Plastics
— 革製のもの	KD205	○	○	DZ	Footwear of Leather
— ショーツが指にかかるとのもの	KD206	○	○	DZ	Footwear of Leather, Strapped (Sandals)
— 短靴	KD207	○	×	DZ	Shoes of Leather
— 紡織用繊維製のもの	KD208	○	○	DZ	Footwear of Textile
— キャンパスシューズ	KD209	○	×	DZ	Canvas Shoes
— 地下足袋	KD210	○	×	DZ	"Jaketabi"
— 中敷き、ヒール等の物品、部分品	KD300	○	○	KG	Parts of Footwear, In-soles, Heel Cushions

【注】 KD100, KD200 - 輸出入の定義が一部異なる。
 テニス、バスケット、体操用の革製の靴は輸入ではスポーツ用履き物に(KD100)に輸出ではスポーツ用を除く履き物(KD200)に分類されている。

図 2-27 商品分類名 (履き物の例)

また、こうした分類の定義をデータベースに追加してゆくためのユーザインタフェースとして、BDCシステムでは、分類定義画面が用意されている。図2-28に示した分類定義画面例から分かるように、履き物の定義に関して「PARENT」、「BROTHER」、「CHILDREN」といった概念により構造化を行っている。特に、分類に意味を持たせていないが、最初に分類を定義し、そこから構造化を行うアプローチをとっている。「PARENT」、「CHILDREN」といった分類関係は、例えば集合関係で含まれていれば、「子供」であり、また、相反する商品があり、その上に同じカテゴリーが存在すれば、「兄弟」と呼ぶといった考え方である。

```

R   IDLE   KBO           007           R5.40 OPEN フォウ
      ****   BDC-CODE DEFINITION TABLE (BY HS-CODE)   ****
      ***   JAPAN   **   IMPORT   ***

      KD200 : FOOTWARE, _____ EX.SPORTS FOOTWARE
      PARENT CODE : KD000           BDC-CODE : KD200

      BROTHER      CHILDREN      CLASSIFIED HS-CODE
      +-----+ +-----+ +-----+ +-----+
      | KD100 | |   | | 640110090 | ~ | 640191 |
      | KD200 | |   | | 640192090 | ~ | 640199 |
      | KD200 | |   | | 640220   | ~ | 640299 |
      |   | |   | | 640320   | ~ | 640351012 |
      |   | |   | | 640351022 | ~ | 640359019 |
      |   | |   | | 640359031 | ~ | 640359099 |
      |   | |   | | 640391012 | ~ | 640391019 |
      |   | |   | | 640391022 | ~ | 640391029 |
      |   | |   | | 640399012 | ~ | 640399029 |
      |   | |   | | 640419   | ~ | 6405   |
      +-----+ +-----+ +-----+ +-----+

      REPORT CTY 1 TYPE 1 PARTNER CTY 0000 ITEM KD200 VALUE
      ***** H.COMMODITY GUIDE R.RETRIEVAL M.MENU E.END
  
```

図2-28 BDCシステム 分類定義画面

このように、ジェトロの貿易統計データベースでは、貿易統計の特徴の一つである商品分類の複雑さを、ユーザの立場に立った分類定義を用意し、また、先に述べたように使いやすいメニューを用意することによって、使用者層の「素人化」に対応している。

以上、本節では、ジェトロの貿易統計データベースについて、特に、「JETRO BDC」と「JETRO ACE」を中心に報告したが、今後のデータベース利用への人工知能技術利用にとって貴重な示唆を提供するものと言えよう。

*** BDC SYSTEM MASTER (JAPAN) ***

IMPORT

FROM:KB000

TO:KC050

KB000	---TRAVEL GOODS, 4202	HANDBAGS & THE LIKE	KC000	---BOOKS NEWSPAPERS & 49	JOURNALS
KB100	----TRAVEL GOODS, BRIEF- 4202. 11.....	CASES, VANITY-CASES 4202. 19	KC010	----PRINTED BOOKS, 4901	BROCHURES
KB101	----TRAVEL GOODS, BRIEF- 4202. 11	CASES OF LEATHER	KC020	----NEWSPAPERS, JOURNALS 4902	AND PERIODICALS
KB102	----TRAVEL GOODS OF 4202. 12	PLASTIC OR TEXTILE	KC030	----PLANS AND DRAWINGS 4906	
KB200	----HANDBAGS 4202. 21.....	4202. 29	KC040	----ADVERTISING MATERIAL 4911. 10	COMMERCIAL CATALOG
KB201	----HANDBAGS OF LEATHER 4202. 21		KC050	----PRITED PICTURES, 4911. 91	DESIGNS & PHOTOGRAPH
KB202	----HANDBAGS OF PLASTIC 4202. 22	OR TEXTILE			
KB300	----WALLETS, PURSES AND 4202. 31.....	SIMILAR CONTAINERS 4202. 39			
KB301	----WALLETS AND PURSES 4202. 31-100 4202. 32-100				
KB400	----SPORTS BAGS AND 4202. 91.....	OTHER CASES 4202. 99			
KB401	----SPORTS BAGS & OTHER 4202. 91	CASES OF LEATHER			
KB402	----SPORTS BAGS ETC OF 4202. 92	PLASTIC/TEXTILE			

*** BDC SYSTEM MASTER (JAPAN) ***

IMPORT

FROM:KD000

TO:KD206

KD000	---FOOTWEAR 64			KD106	*-----FOOTWEAR FOR TENNIS, BASKET OF LEATHER 6403. 91-021 6403. 99-011
KD100	----SPORTS FOOTWEAR 6401. 10-010 6401. 92-010 6402. 11.....6402. 19 6403. 11.....6403. 19 6403. 51-021 6403. 59-021.....6403. 59-029 6403. 91-011 6403. 91-021 6403. 99-011 6404. 11			KD107	-----FOOTWEAR FOR TENNIS, BASKET OF TEXTILE 6404. 11
KD101	-----SKI-BOOTS 6401. 10-010 6401. 92-010 6402. 11 6403. 11			KD108	-----CANVAS SHOES FOR TENNIS, BASKET 6404. 11-010
KD102	----SPORTS FOOTWEAR, OTHER 6402. 19 6403. 19 6403. 51-021 6403. 59-021.....6403. 59-029 6403. 91-011 6403. 91-021 6403. 99-011 6404. 11			KD200	----FOOTWEAR, EX. SPORTS FOOTWEAR 6401. 10-090.....6401. 91 6401. 92-090.....6401. 99 6402. 20.....6402. 99 6403. 20.....6403. 51-012 6403. 51-022.....6403. 59-019 6403. 59-031.....6403. 59-099 6403. 91-012.....6403. 91-019 6403. 91-022.....6403. 91-029 6403. 99-012.....6403. 99-029 6404. 19.....6405
KD103	----SPORTS FOOTWEAR, OTHR, OF RUBBER, PLAST. 6402. 19			KD201	----FOOTWEAR OF RUBBER, PLASTICS 6401. 10-090.....6401. 91 6401. 92-090.....6401. 99 6402. 20.....6402. 99
KD104	----SPORTS FOOTWEAR, OTHR, OF LEATHER 6403. 19			KD202	----BOOTS OF RUBBER, PLASTICS 6401. 92-090
KD105	----FOOTWEAR FOR TENNIS, BASKET, GYMNASTICS 6403. 51-021 6403. 59-021.....6403. 59-029 6403. 91-011 6403. 91-021 6403. 99-011 6404. 11			KD203	----SHOES OF RUBBER, PLASTICS 6401. 99-010 6402. 99-010
KD106	----FOOTWEAR FOR TENNIS, BASKET OF LEATHER 6403. 51-021 6403. 59-021.....6403. 59-029 6403. 91-011			KD204	----SANDALS OF RUBBER, PLASTICS 6402. 99-020
				KD205	----FOOTWEAR OF LEATHER 6403. 20.....6403. 51-012 6403. 51-022.....6403. 59-019 6403. 59-031.....6403. 59-099 6403. 91-012.....6403. 91-019 6403. 91-022.....6403. 91-029 6403. 99-012.....6403. 99-029 6405. 10
				KD206	----FOOTWEAR OF LEATHER, STRAPPED (SANDALS) 6403. 20

*** BDC SYSTEM MASTER (JAPAN) ***

IMPORT

FROM:KD207

TO:KE203

KD207 -----SHOES OF LEATHER
6403. 59-031 6403. 59-099

KD208 -----FOOTWEAR OF TEXTILE
6404. 19 6404. 20
6405. 20

KD209 -----CANVAS SHOES
6404. 19-220
6404. 20-211 6404. 20-212
6404. 20-310

KD210 -----* JIKATABI*
6404. 19-210

KD300 -----PARTS OF FOOTWEAR. IN-SOLES, HEEL CUSHION
6406

KE000 ---RECORDS, TAPES & OTHR SIMILARLY MEDIA
8523 8524

KE100 ----UNRECORDED MEDIA FOR SOUND REC. & OTHER
8523

KE101 ----MAGNETIC TAPES, UNRECORDED
8523. 11 8523. 13

KE102 ----MAGNETIC TAPES, FOR SOUND REC, UNREC.
8523. 11-010
8523. 12-010
8523. 13-010

KE103 ----MAGNETIC TAPES, FOR VIDEO REC, UNREC
8523. 13-021 8523. 13-029

KE104 ----MAGNETIC TAPES, OTHER, UNRECORDED
8523. 11-090
8523. 12-090
8523. 13-090

KE105 ----MAGNETIC DISCS, UNRECORDED
8523. 20

KE200 ----RECORDED MEDIA FOR SOUND REC
8524

KE201 ----MAGNETIC TAPES, RECORDED (INCL. VIDEO)
8524. 21 8524. 23

KE202 ----GRAMOPHONE RECORDS, RECORDED
8524. 10

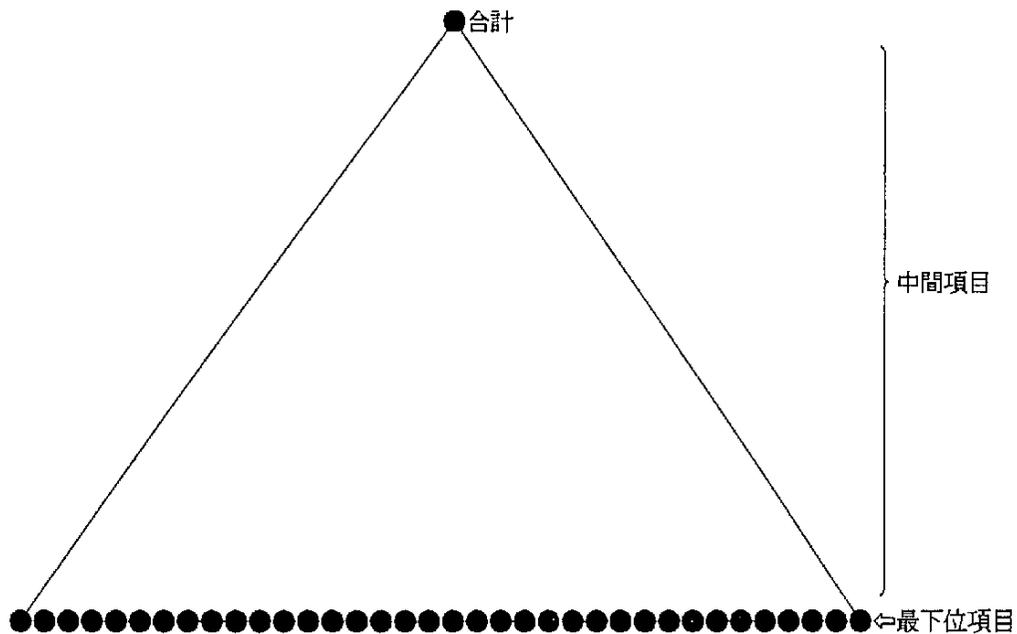
KE203 ----DIGITAL AUDIO DISCS
8524. 90-010

付-25

文責：日本貿易振興会 野村玲子

付録3 分類体系の問題

1 分類体系のとらえ方



- 中間項目のとりかたは、幾つもあり得る。

発表されるデータは、その内、発表者自身の判断で合理的と思われるものを選択して発表している。

- ユーザーは発表されているもの以外の見方をしたいと思う事がある。

この時、普通は手作業で計算して求めている。



収録されているデータから組上げる事が可能な時には、DBはその結果を答えるべきである。

2 分類のその他の問題

① 分類間の関係の問題

異なる「分類体系」間の一般論的な違いについては提供可能か、異なる分類体系に含まれる項目間の関係を記述する事は殆ど不可能。

② 異動の問題

分類の組み換えの例については資料参照。

この問題を一般的に解決するのは困難。

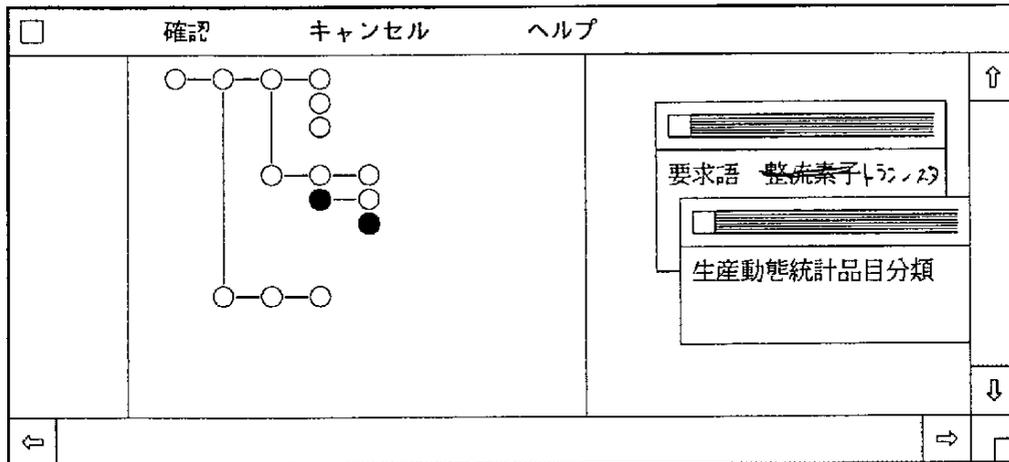
当面、分類間・異時点間の問題は棚上げとし、1分類・1時点内の問題に限定してはどうか？

3 どうやって答えるか？

「改善情報」 組み上げられるかどうかを判断できる情報かノシステム内に無ければ ユーザーに決定させるしか無い。

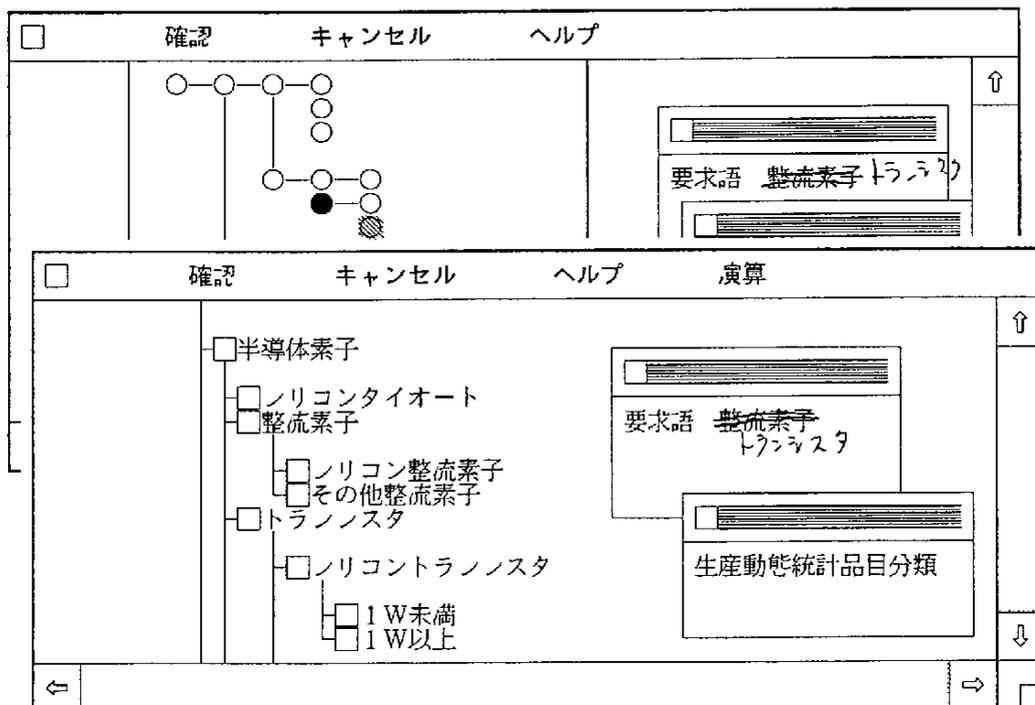
次の様なインターフェースの例はとつか？

- ① 検索要求語の投入 例えは「トランスタ」
- ② 該当項目位置の表示 例えは下記ウィンドウの例



×全体の階層イメージ図が表示され、要求語かヒットした項目かフリンクする

- ③ 詳細の表示と決定。



×前画面で、ヒットした項目をクリックすると該当項目の近辺を表示。

この画面で更に、下位項目による組上げを編集できる様にする。

○考え方の概要

- ① 「次善情報」を決定できるのはユーザー自身である。
- ② 分類項目を意識した場合の「次善情報」の範囲はシステムが十全に提供できる程単純な関係では無い。
- ③ そうはいつでも対象とする範囲は1つの分類体系の中なので そのまま分類体系を見せてしまう事の方が効果がある。

●補足 要求語から項目へのマッピングの方法

ユーザーが選択する事を考慮して、マッピングは厳密で無くても可ては無いか？

以下の方法はどうか。

- ① 各分類項目毎に説明文を収録して置く。(これについては上記画面で参照できる様にする。)
- ② 要求語はこの「文書」を検索する事として考える。
(キーワードを付与する。 説明文からキーワードを抽出する等々。)
- ③ 類義語辞書は、要求語のノソーラスとして位置つける。

以 上

文責：㈱アイ・エヌ情報センター 篠原昇

付録4 Frame Work of COMPASS

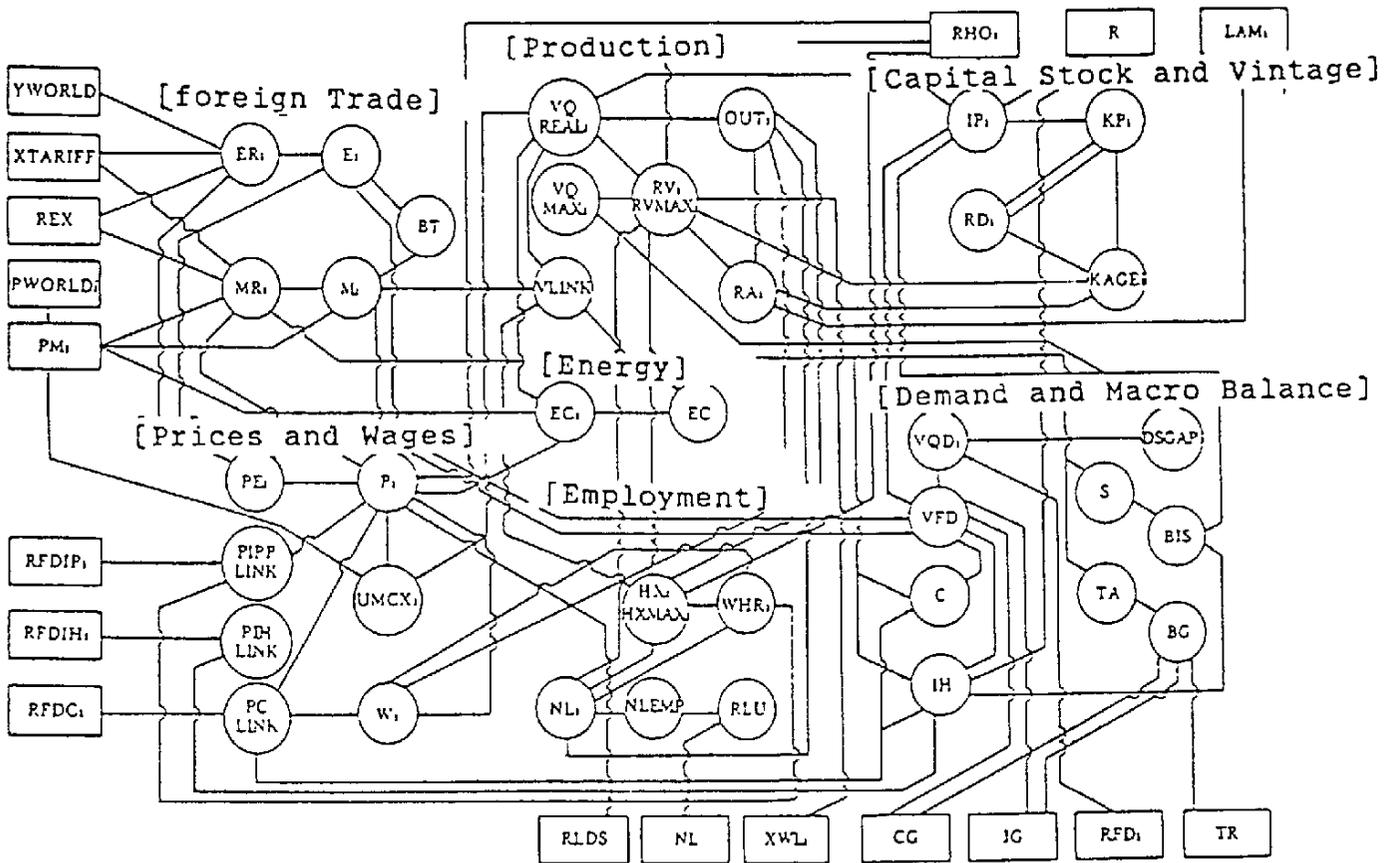
Figure 1. Simplified flow diagram of COMPASS

[Foreign Investment]
IFi
QFi

[Environment and Resources]
IZi
KZi
Zij
MRAW
MFUEL, etc.

[High-téch]
RKCOMi
RKNCi
RKROBOi
RKVENDi

[Research and Development]
R&Di
QPATi



[Household]
Ya
Sa, SSa

[Demography and Time Budget]
N, NM, NF
RNLMa
RNLFa
HMa, HFa

[Quality of Life]
NNW or MEW
KH, KCD, KG
HL
Z

Table 2-2 Sectoral Breakdown of Main Variables

Variables	IP _i	KP _i	KAGE _i	RIDI _i	NL _i	XL _i	HX _i	IQ _i	IQ _i	IK _i	IQRIDI _i	IQLAM _i	IQKAG _i
	Invest- ment	Capital Stock	Vintage of Capital	Capaci- ty Utili- zation	Labor	Labor's Share	Work- ing- Hours	Contri- bution of NL	Contri- bution of HX	Contri- bution of KP	Contri- bution of RDI	Contri- bution of lambda	Contri- bution of KAGE
Code Sectors	EN(eq)	EN(def)	EN(def)	EN(eq)	EN(eq)	EN(eq)	EN(eq)	EN(def)	EN(def)	EN(def)	EN(def)	EN(def)	EN(def)
Chapters	3	4	5	6	7	8	9	9	9	9	9	9	9
1) 01& Agriculture, etc.	x	x	x	---	x	x	---	x	x	x	x	x	x
2) 10& Mining	x	x	x	---	x	x	x	x	x	x	x	x	x
3) 18 Food processing	x	x	x	---	x	x	x	x	x	x	x	x	x
4) 20& Textiles	x	x	x	x	x	x	x	x	x	x	x	x	x
22) 22 Lumber and wood	---	---	---	---	---	---	[x]	---	---	---	---	---	---
23) 23 Furniture and Fixture	---	---	---	---	---	---	[x]	---	---	---	---	---	---
5) 24 Pulp and paper	x	x	x	x	x	x	x	x	x	x	x	x	x
25) 25 Printing and publish.	---	---	---	---	---	---	[x]	---	---	---	---	---	---
29) 29 Leather and products	---	---	---	---	---	---	[x]	---	---	---	---	---	---
28) 28 Rubber products	---	---	---	[x]	---	---	[x]	---	---	---	---	---	---
6) 26 Chemicals	x	x	x	x	x	x	x	x	x	x	x	x	x
7) 27 Petroleum and coal	x	x	x	x	x	x	x	x	x	x	x	x	x
8) 30 Nonmetal minerals	x	x	x	x	x	x	x	x	x	x	x	x	x
9) 31B Iron and Steel	x	x	x	x	x	x	x	x	x	x	x	x	x
10) 32 Nonferrous metals	x	x	x	x	x	x	x	x	x	x	x	x	x
11) 33 Fabricated metals	x	x	x	x	x	x	x	x	x	x	x	x	x
12) 34 General machinery	x	x	x	x	x	x	x	x	x	x	x	x	x
13) 35 Electrical machinery	x	x	x	x	x	x	x	x	x	x	x	x	x
14) 36M Motor Vehicles	x	x	x	x	x	x	x	x	x	x	x	x	x
15) 36S Other transport	x	x	x	x	x	x	x	x	x	x	x	x	x
16) 37 Precision instrument	x	x	x	x	x	x	x	x	x	x	x	x	x
17) 38& Manufacturing n.e.c.	x	x	x	---	x	x	x	x	x	x	x	x	x
18) 13 Construction	x	x	x	---	x	x	x	x	x	x	x	x	x
19) 70 El. gas, water	x	x	x	---	x	x	x	x	x	x	x	x	x
20) 40 Wholesale and retail	x	x	x	---	x	x	x	x	x	x	x	x	x
21) 50& Fin., ins., real estate	x	x	x	---	x	x	x	x	x	x	x	x	x
22) 60I Transportation	x	x	x	---	x	x	x	x	x	x	x	x	x
23) 60C Communication	x	x	x	---	x	x	x	x	x	x	x	x	x
24) 97 Public admin.	---	---	---	---	x	x	---	---	---	---	---	---	---
25) 75P Public services	---	---	---	---	x	x	---	---	---	---	---	---	---
75M Private services	x	x	x	---	x	x	x	x	x	x	x	x	x
00 All industries	x	x	x	x	x	x	x	x	x	x	x	x	x
0M Manufacturing	x	x	x	x	x	x	x	x	x	x	x	x	x

Note: [] indicates that the variables are available and discussed in the text but are not used in the COMPASS.

付録5 システムテスト概要

実験システムをもとに、4人の委員が試用を行い、主に操作性に関して評価を行った。

①知識獲得の為のシステム仕様について

・操作のし易さ

ファンクションキーの操作が分かりにくい。
メニューを見るまで、どれだけ知識が入っているか判らない。
(最初からメニューが開いた方がよいか?)
入力ミスをした場合修正がしにくい。(エディタによる修正)
ユーザによっては、コードを見て確認をしたい場合がある。
(標準分類が信頼できるかどうかという問題を含む)

・知識構造の妥当性について

類義語辞書にコードが含まれるのは危険(ユーザが誤操作をする)なので、類義語と、コードは、編集時には、分離されていることが望ましい。

②収集された知識の性質について

個人辞書の分析は未だ行っていないが、類義語のレベルでは、多量の類義語が収集されることが予想され、また、同じ類義語が異なる基本名詞に対応する場合がある。
これらの処理基準が必要。

- ・個人的な差異の大きい知識と小さい知識の分類
- ・利用目的と必要な知識の関係に関する分析

③知識ベースの運用管理方法について

具体的な個人辞書を獲得していない為未評価

- ・拡充仕様書に示した運用イメージの妥当性
- ・知識獲得の際のチェック基準として必要なものはなにか

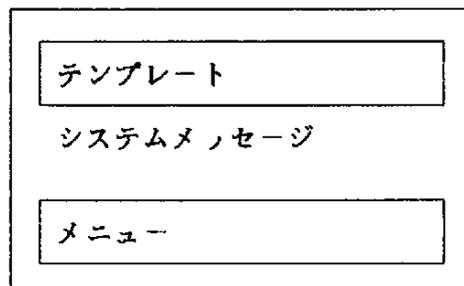
④知識ベースの効用について

- ・利用可能なユーザの範囲は、知識を使わない場合と比べ、広がったか。
操作性があまりよくないため未評価
- ・未だ不足している知識があるか。
検索失敗が、データの欠測か、0値かの違いをデータベース毎に整理した知識が必要。

⑤その他システム全体の評価

・操作性

システムメッセージは、中央に表示したほうがよいとの意見があった。



(いままでのままでよいとの意見もあり、判断不能)

ファンクションキーのガイドは、最下段が分かりやすい。(全員)

- ・機能

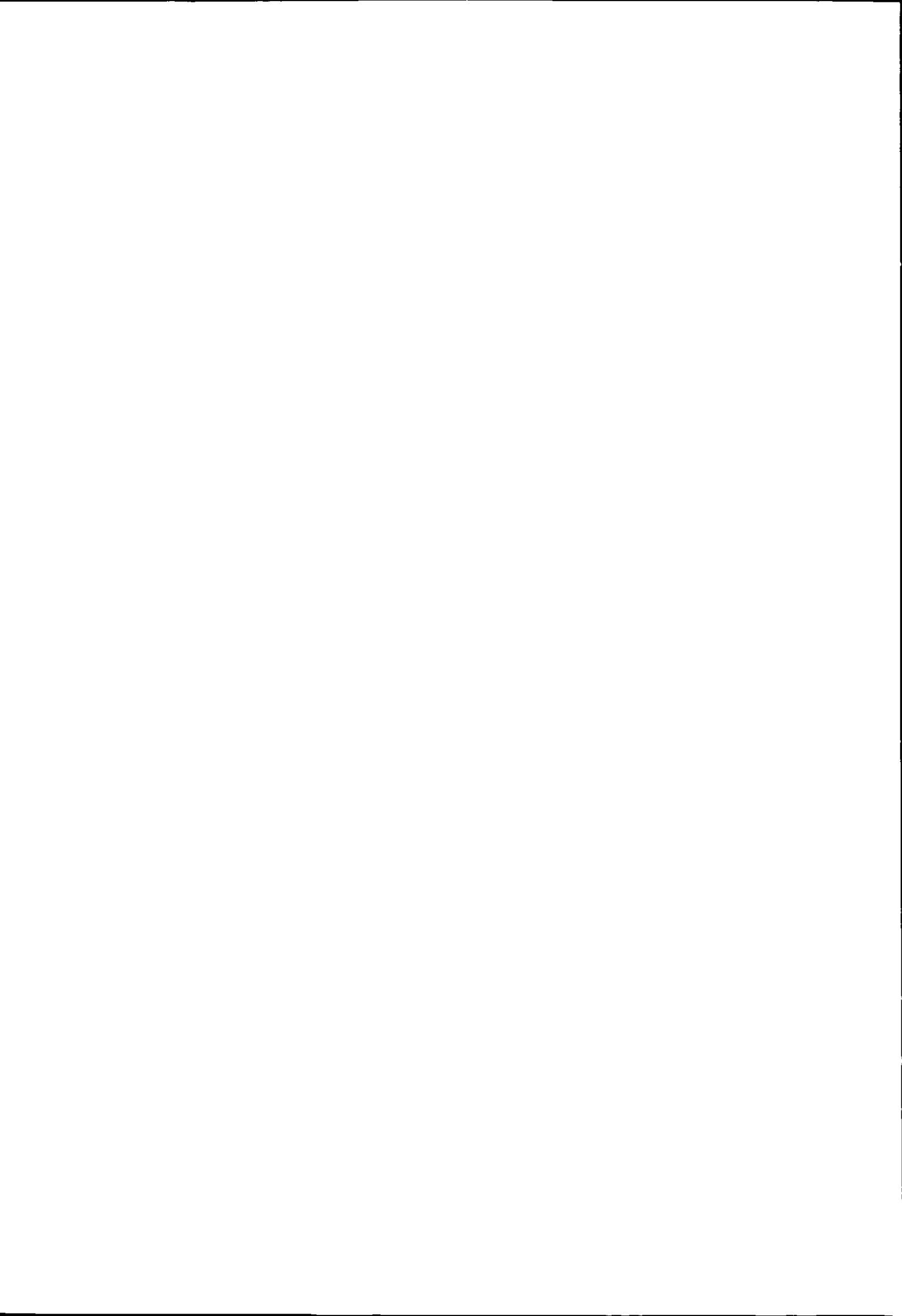
カイドが不親切で、マニュアルレスになっていない。
キャンセルの機能が不足している。

- ・性能

検索要求を発生する度に、ホストにログインしているのは非効率。

参 考 文 献

- (1) (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究報告書
－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
- (2) (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究報告書
－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.
- (3) 石川徹也：文意解析処理に基づく主題索引語作成支援システム，情報処理学会論文誌、
Vol.32 No.2 pp220-228, 1991.
- (4) 通商産業省：通商白書，1990.
- (5) 総務庁：日本標準商品分類，1990.
- (6) 総務庁統計局統計基準部：統計調査総覧，1990.
- (6) 総務庁：昭和60年産業連関表－計数編(2)－，1989
- (7) アスキー Introduction to OS/2 . 1990.
- (8) DPC . データベース・サービスに関するユーザ意識調査，1988.



—— 禁 無 断 転 載 ——

平成 3 年 3 月 発行

発行所 財団法人 日本情報処理開発協会
東京都港区芝公園 3 丁目 5 番 8 号
機械振興会館内
TEL 03(3432)9372(代表)

印刷所 株式会社 方 泉 社
TEL 03(3253)2771(代表)

02-R 002



