

# ネットワーク志向の 経済時系列統計資料検索システム

椿 康 和

## 1 はじめに

最近の、ダウンサイジング、ネットワーク化やオープンシステム化の潮流は、情報提供=検索システムの世界にも大きな変革をもたらしている。学術データベースの目的の1つは知識の共有とその流通の促進であるが、大型汎用機とその上で稼動する DBMS を用いた従来型のシステムでは、広域ネットワークへの対応が困難で、サービスの提供範囲に自ずと限界がある。それに加えて、構築・運用にかかるコスト上の問題もあり、かなり大規模な組織においても維持そのものが困難になりつつある。

これに代わるものとして、現在、社会全体としての「情報の共有と再利用」を促進する中心的役割を果たしつつあるのが、インターネットを基盤とする情報提供システムである。その理由は、インターネットに接続されたホスト上に分散して維持されている情報を対象とする検索手段として、従来の telnet や FTP (File Transfer Protocol) 等の方法に加えて、WAIS (Wide Area Information Servers), Gopher, WWW (World Wide Web) などの、広域情報システムが急速に普及してきたためである。前者の方法では、情報の入手にあたって、その所在場所を確認し、それぞれの保存形態に応じた操作手順によって個別に入手するという操作を伴う。これに対し、後者においては、記述されたリンクをたどることによって関連する情報に導かれるという、非常に簡素化された操作で情報が入手できる。また、これらのシステムは、ワークステーションやパソコン上での運

用が容易で、不特定多数の利用者を対象とした情報の発信や公開を、極めて低いコストで行うことを可能にしている。

その結果、インターネットの広く普及した米国では、その上で流通する情報の種類は極めて広範囲に及んでおり、株価や各種の経済指標など、経済・経営分野においても多種多様な統計資料を提供するサーバが様々な政府機関や大学で運用されている<sup>1)</sup>。

これらの情報の提供にあたっては、telnet や FTP, Gopher といった手段が広く用いられているが、上述の広域情報システムの中でも、最も効果的な情報提供サービスを実現することができるかと評価されているのが WWW である。本稿では、この WWW を用いて構築した、経済時系列統計資料の検索システムについて述べる。

## 2 WWW による情報提供＝検索システム

### 2.1 World Wide Web

WWW は、インターネットで結ばれた各地のサイトに散在する、様々な情報（テキスト、音声、画像等）を、関連する情報に対する記述されたリンクをたどるという方法で容易にアクセスするために、CERN (Conseil European pour la Recherche Nucléaire:ヨーロッパ素粒子物理学研究所) によって開発された、インターネット上の広域情報システムである。

WWW では、この機能を、URL (Universal Resource Locator) とハイパー・テキスト (Hyper Text) とによって実現している。URL は、インターネット上に存在する情報資源の所在地を、情報を提供するサーバ名とファイル名とによって記述し、HTTP (Hyper Text Transfer Protocol) や FTP など、情報の送信に用いられる通信プロトコルや情報の利用形態を定義するものである。ハイパー・テキストは、通常の文書では、脚注や

---

1) マクロ経済資料を例にとると、U. S. Department of Commerce の Economic Bulletin Board (EBB) や STAT-USA, University of Maryland の EconData などがある Goffe[4]。わが国では、一部の組織において株式指標等の情報提供の試みが始まったばかりである。

参考文献リストなどによって間接的にしか提供できない、他の情報資源に対するアクセスの方法を、一定の形式で文書中に埋め込むことによって、直接的にアクセスできるようにする仕組みである。WWW では、この機能の実現にあたり、情報を記述する手段として、HTML (Hyper Text Markup Language) を用いている。HTML は、文書の相互流通の促進のための標準化手法として ISO で制定された、SGML (Standard Generalized Markup Language) を基礎とした、文書の構造を記述する言語である。

WWW による情報の提供は、クライアント/サーバ形式で行われる。すなわち、情報を提供する WWW サーバは、一般に WWW ブラウザと呼ばれている、情報利用者側のクライアント・プログラムからのリクエストに応じて、指定された情報を送信する。クライアント側では、受信した情報を解読し、文字・画像・音声等の特性に応じて、ハイパー・テキスト形式で利用者に提示を行っている<sup>2)</sup>。

## 2.2 WWW による情報提供＝検索システムの特徴

WWW の機能は、サーバからの一方向的な情報提供や、他のサーバへのリンク機能だけではない。利用者からサーバへのデータの入力機能の他、CGI (Common Gateway Interface) 機能を利用して、クライアントからのリクエストに応じて、さまざまなプログラムをサーバ上で実行させることが可能である。

これらの機能を利用すると、対話的な情報提供＝検索システムを比較的容易に実現することができる。とりわけ、Awk 等で作成されたテキスト処理プログラムと組み合わせて、文献目録や人名録等のテキスト形式の情報を対象とするシステムが、多くのサイトで構築されている<sup>3)</sup>。

---

2) この WWW ブラウザとしては、ワークステーションからパソコンまで幅広いプラットフォーム上で動作する Mosaic や、多彩な表示機能をもつ Netscape Navigator などがある。

3) <http://www.ll.chiba-u.ac.jp>, <http://www.hiroshima-u.ac.jp> など。

WWW を用いた情報提供=検索システムは、汎用機と DBMS によって構成されてきた従来型の情報検索システムと比較して、以下の特徴を有している。

(1) 低コストと開放性

情報提供=検索システムを従来型のシステムの上で稼働させるためには、DBMS や TSS の運用のために、相当な規模の設備を必要とする。また、ネットワークについても、個々のホストを中心とする垂直型の形態をとることから、その利用者はそれぞれのネットワーク内部に限定されがちである。

これに対し、WWW ではワークステーションやパソコンクラスのコンピュータを情報提供サーバとして利用できるため、システムの構築や運用にかかるコストは極めて低く押さえることが可能である。また、ネットワーク部分については、組織内の LAN と組織間のインターネットという普遍的な情報流通の枠組みを利用することで、情報の利用範囲が大きく広がっている。さらに、WWW サーバ上の情報は、原則として利用者の資格に関する制限なしに、インターネット全体に広く公開されている。

このような、極めて開放的な情報提供システムは、システムの維持に必要なコストを利用者から回収することが困難であることや、デジタル化によって、画像等についても複製や加工が容易になったことに起因する、情報の二次的利用における著作権保護の問題なども生じさせているが、文献等の情報源に関する案内情報や、政府作成の諸統計などのように、社会全体で広く共有すべき公共財的性質をもつ情報の提供には最も適した手段であるといえる。

(2) 共通的な利用者インターフェース

従来型の情報提供システムでは、検索コマンド等の利用者インターフェースは、個々の DBMS や検索プログラムごとに異なったものとならざるを得ない。これに対し、WWW では利用者インターフェース部分が HTML 形式によって統一的に記述されているため、利用者側は、WWW ブラウザという検索ツールを用いることで、どの情報源についても共通の

インターフェースで検索することが可能となっている。

その反面、機能面ではブラウザや HTML の仕様に直接的に依存しており、表現上の制約が存在する。このため、テキストや画像に比べ、多量の数値情報など、従来あまりその対象とされてこなかったタイプの情報の場合、膨大な量の HTML 記述を余儀なくされたり、情報の内容に即したユーザ・ビューのサポート等が別途に必要となる場合もある。

### (3) 分散データベースの実現

WWW は、インターネット上に散在する情報資源を、効率的に利用することを目的として開発されたものであり、インターネット上の WWW サーバ群は、全体として分散データベースを構成している。これらのサーバ間では、URL によって所有する情報へのポインタを記述することで、情報資源を共有する仕組みであるため、インターネット上の他の情報源との相互の関連づけが容易である。また、データベースの構築と維持が各サーバ側で行われることから、データの発生源により近い場所でデータベースの更新を頻繁に行って最新の情報を提供することや、情報の利用に関するコントロールを情報の生産者が直接に行うことも可能になる。

### (4) 情報のデジタル化と相互利用の促進

専門性の高い情報の場合、共有化のメリットは、組織外部も含めれば大きいものの、組織内部でみる限り小さくなる。従来型のシステムでは、デジタル化情報の共有範囲は組織内部に限定されがちである。その結果、このような情報については、システム運用上のコスト・パフォーマンスが低くなり、これが、専門性の高い情報の、デジタル化・共有化を阻げる一因となってきている。

WWW は、小規模なシステムでも情報の提供が可能なことや、組織外部での大きな潜在的需要を想定できることから、これらの情報についても、デジタル化と組織内外での流通を飛躍的に促進させる効果をもっている。

### 3 経済時系列統計データ提供システム

#### 3.1 時系列データベースの構造

##### (1) 単純時系列データベース

規則的な時間間隔ごとに観察された時系列データを対象としたデータベースでは、測定間隔や収録期間、時間的な集約方法などの時系列属性が系列ごとに異なることから、これらと値のリストを組にして、個々の時系列を単位に収録する形式が採用されることが多い。この形式は単純時系列データベースをよばれ、収録時の管理や時系列データ間の演算の容易さ、さらには、期間途中における分類の変更への対処の容易さなどの利点もっている。

その反面、この形式では、時系列データが、地域など、時系列属性とは異なる属性によっても構造化されている場合、収録にあたって、その構造に関する情報が系列名に含まれるなど、表面的には失われてしまうことから、明示的な構造に依拠した条件検索が不可能となる。また、利用に際しては、分析に先立って、時系列間で異なる属性をできる限り均一に揃えて提供するための編集・加工が不可欠であったり、規模が大きくなると、収録された多数の時系列の中から必要な時系列を探すのが困難になるなどの問題も生じてくる。

本システムでは、時系列をグループ化し、それを単位とした時系列の選択や検索条件の設定等を行うインターフェースを提供することで、これらの問題点を解決しようとしている<sup>4)</sup>。

##### (2) 時系列のグループ化

時系列をグループ化する基準には、収録された統計表の定義や、地域や産業分類などの時系列属性以外の一般の分類属性が考えられる。経済時系列の場合、一般に、関連する複数の時系列が1つの統計表に集約され、統計書の形式で公表されることが多いため、統計表を単位としたグループ化

---

4) 佐藤 [1] では、統計表の構造情報を記述する「統計項目解釈情報ファイル」の利用が提案されている。

は容易である。また、これらの時系列は、期間種別や収録期間が揃っている傾向があり、利用にあたっては、同時に利用される可能性が高いことから、グループを単位とした検索方法は適しているといえよう。また、一般の分類属性については、属性そのものをグループ名とし、属性値を系列名としたグループの定義もできるであろう。

(3) 情報ファイル

例えば、時系列を統計表を基準としてグループ化した場合、グループ名（統計表名）からだけでは、必要な時系列を探すことは困難であり、また、同種の時系列が複数の統計表に収録されている場合も多く、時系列名からその含まれるグループ（統計表）を特定化する手段も欠かせない。

このため、このグループ化にあたっては、1つの統計表に含まれる各時系列に共通の属性情報等を記述した「統計表情報ファイル」と、各時系列の所属（統計表名など）と時系列固有の属性等を記述した「系列情報ファイル」を編成し、統計表と時系列名の両方の視点から、統計表の検索を行うこととした。それぞれのファイルには、表3.1に示す項目が含まれている<sup>5)</sup>。

(4) 統計表定義ファイル

本システムでは、情報提供の枠組みとして WWW を用いている。このため、上記の情報ファイルとは別に、各統計表ごとに、検索対象の時系列の選択と検索条件の指定にあたって、利用者インターフェースを定義する、

表3.1

情報ファイル	項 目 名
統計表情報ファイル	分野, 統計書名, 統計表名, 時系列属性 (期間種別, データ種類), その他の共通属性 (単位等), 統計表定義ファイル名
系列情報ファイル	系列名, 系列コード, 統計表定義ファイル名

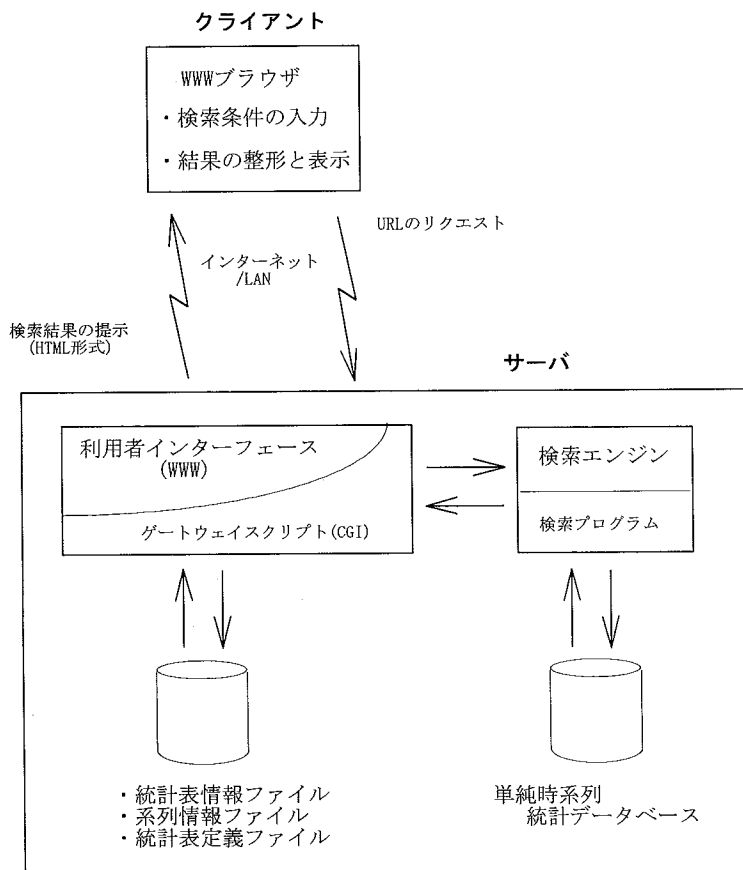
5) 系列情報ファイルには、系列の検索方法を多様化するために、ここに挙げたものの他、例えば系列分類等の項目も含めることも可能であり、それに伴うシステムの変更も容易である。

HTML 形式の「統計表定義ファイル」を用意している。このファイルには、さらに、統計表や各系列に関する記述の情報も一部含まれている<sup>6)</sup>。

### 3.2 システムの概要

本システムは、(1)検索対象となる時系列を、統計表や統計項目の定義にもとづいて対話的に特定化し、検索条件を設定する利用者インターフェー

図3.1 システムの概要



6) HTML 形式の統計表定義ファイルは、別のプログラムを用いて情報ファイルから自動的に生成されている。



スト、(2)設定された条件にもとづいて、単純時系列データベースから対応する系列を検索する検索エンジン、から構成されている(図3.1)<sup>7)</sup>。

(1) 利用者インターフェース

2. 2で述べたように、WWW を用いることで、ネットワーク環境における一般的で簡易な利用者インターフェースを容易に提供することが可能となる。本システムでは、WWW における CGI 機能を利用して、これを実現している。

サーバ上の CGI スクリプトは、WWW ブラウザからのリクエストによって起動される。起動されたスクリプトは、必要に応じて他のプログラムを呼び出し、それらの処理結果やスクリプト自身の生成した情報を、クライアントの WWW ブラウザに送る。ブラウザ側は、受け取った情報を整形して表示する。

本システムにおける CGI スクリプトは Perl で記述されており、以下の処理を行っている<sup>8)</sup>。

a) 統計表と統計項目の検索

指定されたキーワードによって、統計表情報ファイルと系列情報ファイルを検索し、該当する時系列を含む統計表の一覧を HTML 形式で提示する。

b) 検索条件の入力

選択された統計表の統計表定義ファイル (HTML 形式) を提示して、検索条件と検索対象の時系列を特定化させる。

c) 検索プログラムの起動

入力された検索条件をチェックし、単純時系列データベース検索プログラムを起動する。

d) 検索結果の表示

---

7) 本システムの構築では、ワークステーションとして FUJITSU S-4/10 を、WWW として CERN httpd 3.0 を用いている。

8) Perl はテキスト処理やシステム処理に適合した、汎用的な手続き型のインタプリタ言語であり、C 言語に比べ、プログラムのステップ数の大幅な節減が可能である。

検索プログラムから検索結果を受け取り、HTML 形式で提示する。

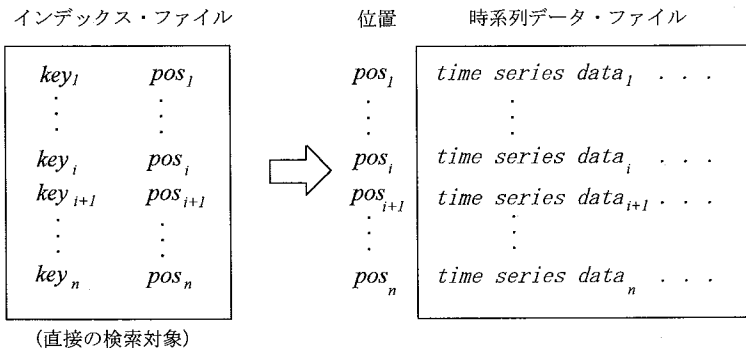
(2) 検索エンジン

CGI スクリプトから起動される検索エンジンは、受け取った検索パラメータにもとづいて、単純時系列統計データベースから必要な系列を取り出し、期間集約などの適切な編集を行って CGI スクリプトに渡している。この検索プログラムとデータベースの原形は、これまで汎用機上で運用されてきたものであるが、本システムの構築にあたり、大幅な改良を加えてワークステーション上に移植した<sup>9)</sup>。

特に、検索の性能を上げるためにデータベースのファイル構造を変更した。データベースは、検索対象となるインデックス・ファイルと、データ本体を収録する時系列データ・ファイルとから成る。図3.2に示すように、検索プログラムは、検索キーから時系列データへのポインタを格納したインデックス・ファイルをサーチし、時系列データ・ファイル中の該当時系列が格納されたアドレスを求め、時系列データを取り出している<sup>10)</sup>。

3.3 実行例

図3.2 インデックス・ファイルと時系列データ・ファイル



9) 椿 [2].

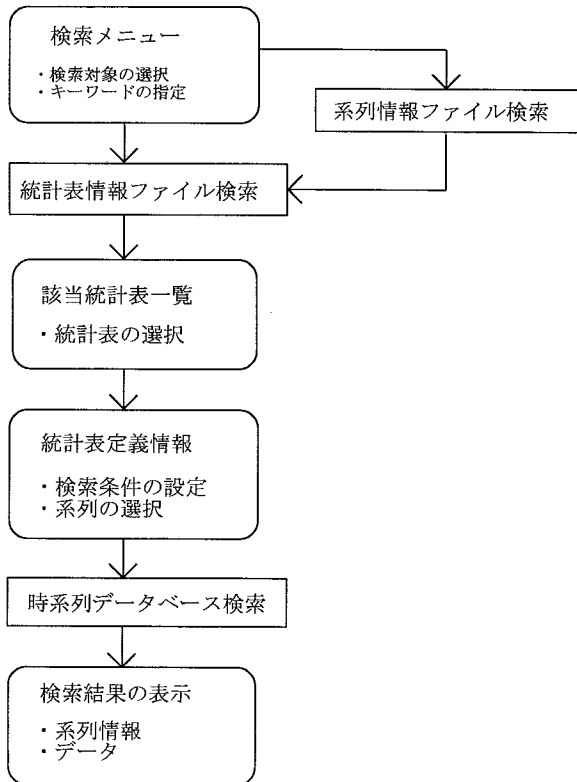
10) インデックス・ファイルは、系列ごとに異なる収録データサイズの情報にもとづいて、時系列データ・ファイル作成後、別のプログラムによって生成される。

本システムによる検索の流れを図3.3に、実行例を図3.4～3.7に示す<sup>11)</sup>。

まず、検索メニュー（図3.4）において、対象分野と期間種別、検索対象（統計表情報ファイル又は系列情報ファイル）を選択し、さらに、検索対象に関するキーワードを指定する。指定されたキーワードは、統計表名や時系列名との部分一致検索に用いられる。

系列を対象とする検索では、系列情報ファイルをサーチして該当の系列を検出し、さらに、それを含む統計表を統計表情報ファイルから取り出して、該当するものを一覧表示する（図3.5）。

図3.3 検索の流れ

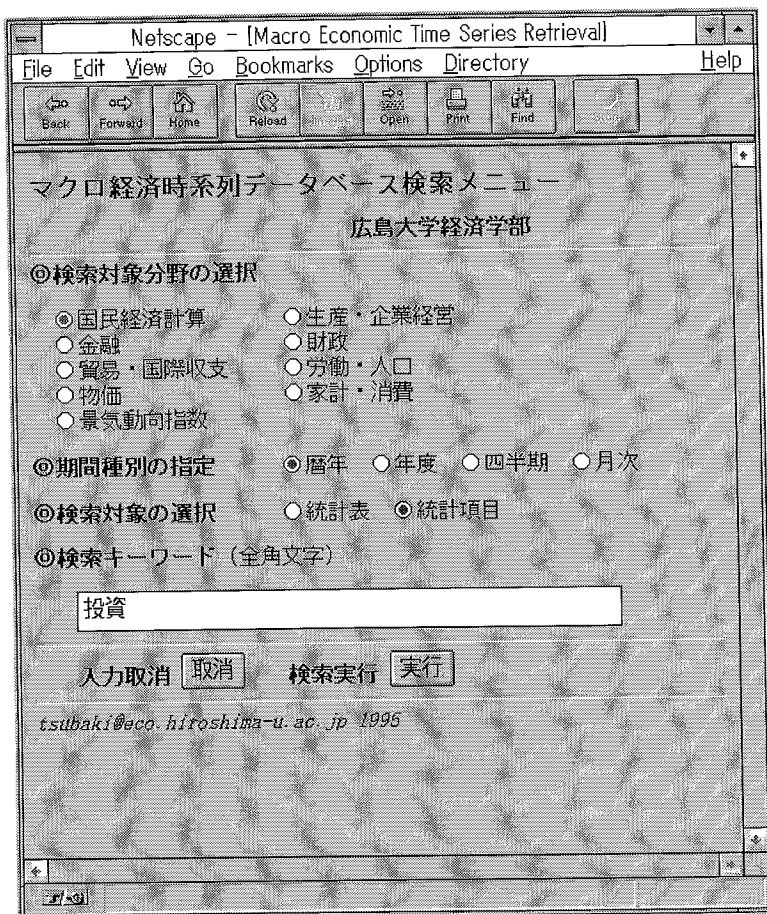


11) ここに示した実行例は、Netscape Navigator Ver1.1n によるものである。

表示された一覧表の中から特定の統計表を選択すると、URL で指定された統計表定義ファイルが呼び出され、該当の統計表に含まれる系列が一覧表示される (図3.6)。この画面では、グループ化された時系列に関する記述的情報を提供するとともに、検索期間の指定と対象となる個々の時系列の選択を行う。

検索条件を特定化して検索の実行を指示すると、検索エンジンにその情

図3.4 検索メニュー

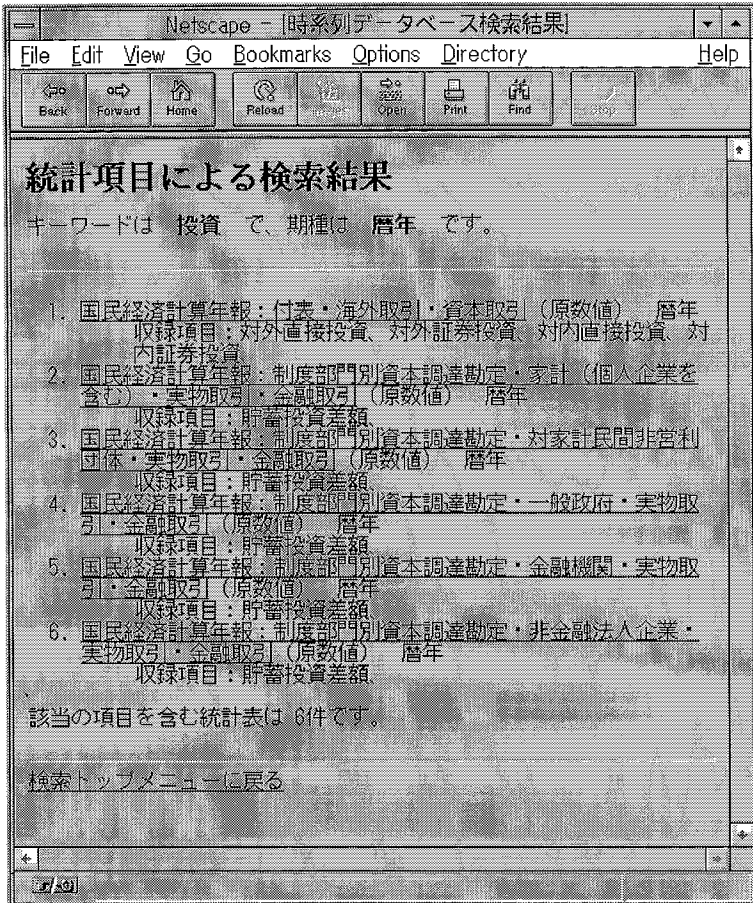


報が渡され、検索プログラムが起動される。検索結果の表示は図3.7のようになる。

#### 4 今後の課題

今回のシステムの構築を通じて、WWW をベースとする情報提供システムが、経済時系列資料を対象とした場合にも相当の有効性をもつことが

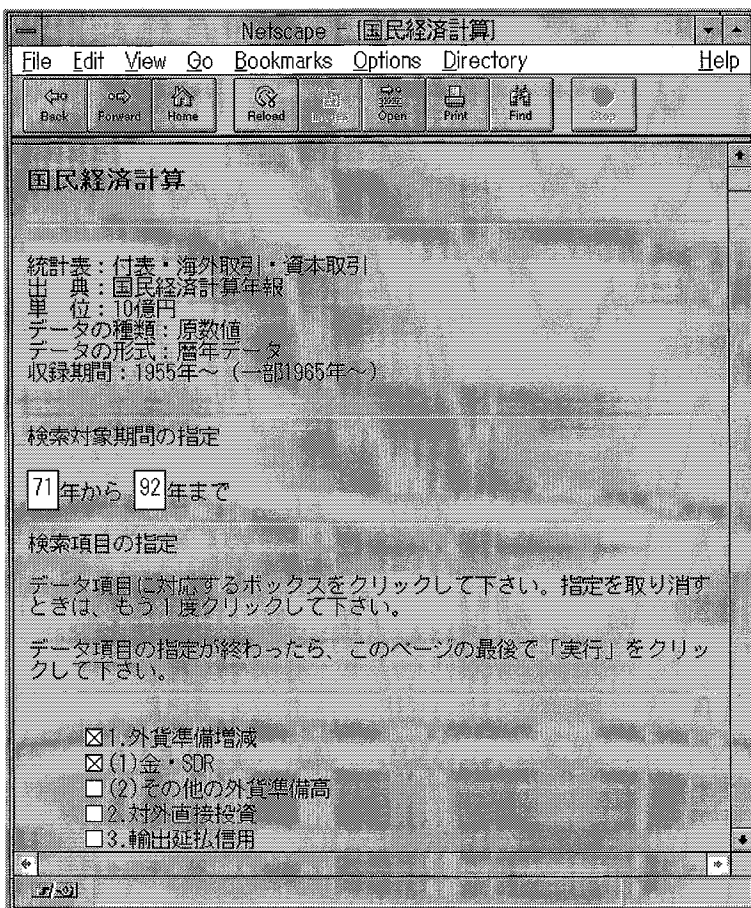
図3.5 情報ファイルの検索結果



検証できた。

しかしながら、本システムは試験的で小規模なものであるため、広域情報検索機能やマルチメディア対応といった、WWW の優れた特性を十分に活用するまでには至っていない。例えば、前者についてみれば、統計資料を有効利用する上で、関連する統計文献に関する情報や各時系列の特性情報など、本システムの外部の他の情報源へのハイパー・リンクも必要と

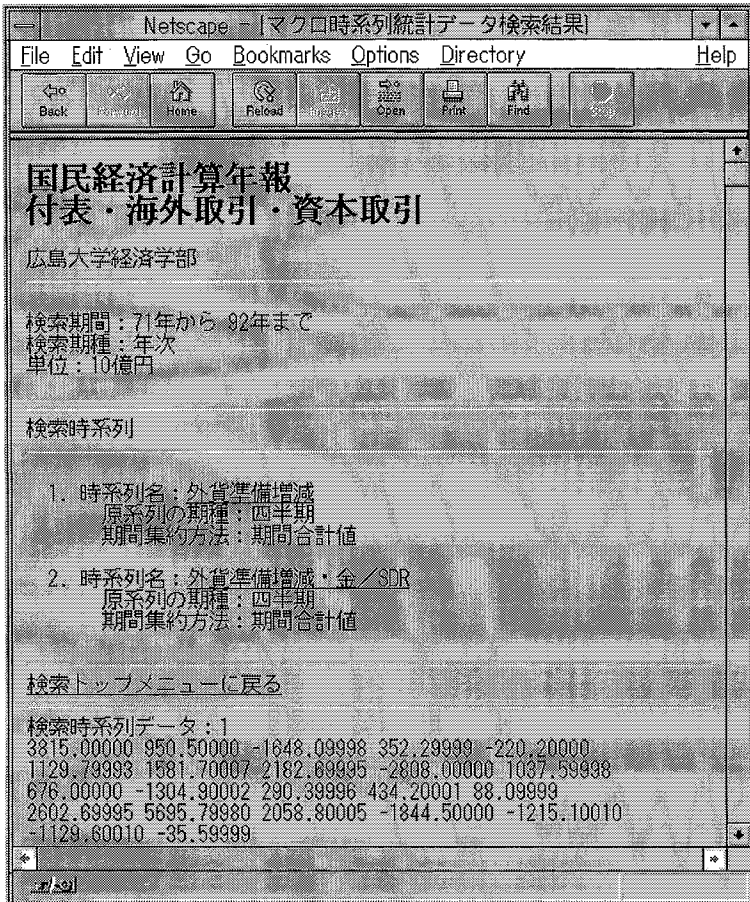
図3.6 統計表の定義情報の表示と検索条件の入力



されるであろうし、後者の場合には、リクエストに応じて時系列のグラフを画像として提供することなども考えられる。また、膨大な資料の中から分析に必要とされる時系列に導くための過程をきめ細かくサポートするための、統計データ辞書の充実も不可欠である。

他方、現行の HTML の仕様—HTML 2.0—は、必ずしも数値データの提供に適しているとは言い難い側面もある。ダウンロード後の再利用など

図3.7 時系列データベースの検索結果



を考慮すれば、anonymous FTP や Gopher など、他の広域情報提供機能をシステムに組み込むことを検討する必要もあろう (96. 1. 6)。

なお、本研究は、平成7年度文部省科学研究費補助金一般研究 (B) (課題番号06451103) によるものであることを付記し、ここに謝意を表する。

### 参 考 文 献

- [1] 佐藤英人, 『統計データベースの設計と開発—データモデルと知識ベースの応用』, オーム社, 1988
- [2] 椿康和, 「時系列資料の検索システム RST について」, 広島大学経済論叢, 第5巻第3号, 1982
- [3] WIDE Project 編, 『インターネット参加の手引き1995年度版』, bit 別冊, 1995
- [4] Goffe, B., *Resources for Economists on the Internet*, Vol.1, No. 11, Oct., 1995
- [5] Lemay, L., *Teach Yourself Web Publishing with HTML in a Week*, Sams Publishing, 1995 (武倉他訳, 「HTML 入門—WWW ページの作成と公開」, プレンティスホール出版, 1995)
- [6] Wall, L. and R. L. Schwartz, *Programming perl*, O'Reilly and Associates, Inc., 1990 (近藤訳, 「Perl プログラミング」, ソフトバンク, 1993)