

【研究ノート】

# 時系列資料の検索システム RSTについて\*

椿 康 和

## I. は し が き

近年のコンピュータ技術の発達を反映して、本学の総合情報処理センターにおいてもその提供する各種サービスが質的にも量的にも充実してきており、利用者の環境もより使いやすい方向に大幅に改善されている。これを背景として社会科学の分野においても、研究目的に適合した資料の検索・加工・分析を、プログラミングの知識なしに容易に行うことのできるシステムの整備が求められている。

本稿で紹介する時系列資料の検索システム (A Retrieval System for Time-Series Data, 略称 RST) は、このような要請の中でとりわけ時系列資料のファイル検索のために開発された、TSS 環境下での対話型検索システムである<sup>1)</sup>。その特徴は、

- a) 検索対象となる時系列資料のファイル (以下 RST ファイルと呼ぶ) と検索プログラムが、社会科学の分野、とりわけ経済分析に使用する時系列資料の特殊性を考慮していること<sup>2)</sup>、

---

\* 本システムの作成にあたり、アルゴリズム等において広島大学経済学部横山和典教授のご指導を受け、同前川功一助教授にもかみかみ助言をいただいたことに対して謝意を表します。現在、本システムは本学情報センター設置の HITAC-M180, VOS3/TSS の下で稼働中である。なお、使用言語は VOS3/最適化 FORTRAN 77 である。

- 1) 時系列資料等の加工・分析を主な目的としたシステムとしては、本号に紹介されている TQASS などがある。
- 2) RST ファイルの作成については横山和典教授または筆者にご相談されたい。

- b) 操作が簡単であること,
  - c) さまざまな形態の検索要求に柔軟に対応できること,
  - d) 他のプログラムでもただちに使用できる形式で検索結果をデータ＝セットに出力すること,
- などである。

## Ⅱ. RST の 機 能

もともと社会科学の分野で使用する時系列は、作成されてきた目的・歴史の相違や度々行われる概念変更などのために、通常、収録される期種（測定間隔）および期間などの属性が系列ごとに異なる。他方、それらの系列群を検索・加工する場合にはできるかぎり同一条件のもとでまとめて使用することが多い。RST ではこの資料の収録と検索との間に存在する環境の相違を配慮して、系列群に対する検索条件の一括指定方式を採用した。これにより、収録系列固有の期種と検索条件における期種が異なる場合には、適切な編集を行った上で条件に合わせて出力して、期間外部分の切り捨てと欠落部分に対する 0 値の補充を自動的に行っている。

RST の操作については、使用頻度の少ない利用者を考慮して、システム起動後の操作はすべて端末に表示される入力指示にしたがって行う形式を採用している。このため、利用者側に要求される準備作業は検索される系列の選択だけである。

また、検索結果の出力方法とその内容に関して利用者側の大幅な裁量を認めている。端末以外の検索結果の出力先としては、センタに設置されているライン＝プリンタ（大量出力用）とデータ＝セット（他のプログラムへのデータ受け渡し用）が指定できる。さらに、端末に出力する情報の種類についても状況に応じた選択が可能となっている。これらにより、利用者はその目的に適合した効率の良い検索作業を行うことができる。

検索結果の利用度を高めるため、それが出力されるデータ＝セットの形式にも柔軟性をもたせてある。検索結果を保存するデータ＝セットは、プログラム＝パッケージ TQASS および QASS でそのまま扱うことができ

る。また、このデータ=セットのレコード形式と書式を、他の利用者プログラムなどでただちに使用できる汎用性のあるものに変換する機能も備えている。

### Ⅲ. RST の 構 成

RST は、

- a) 時系列資料ファイルの検索、
- b) 系列コードの変更、
- c) データ=セットの変換、

の機能を異にする三つのプログラムにより構成され、それぞれ利用者の業

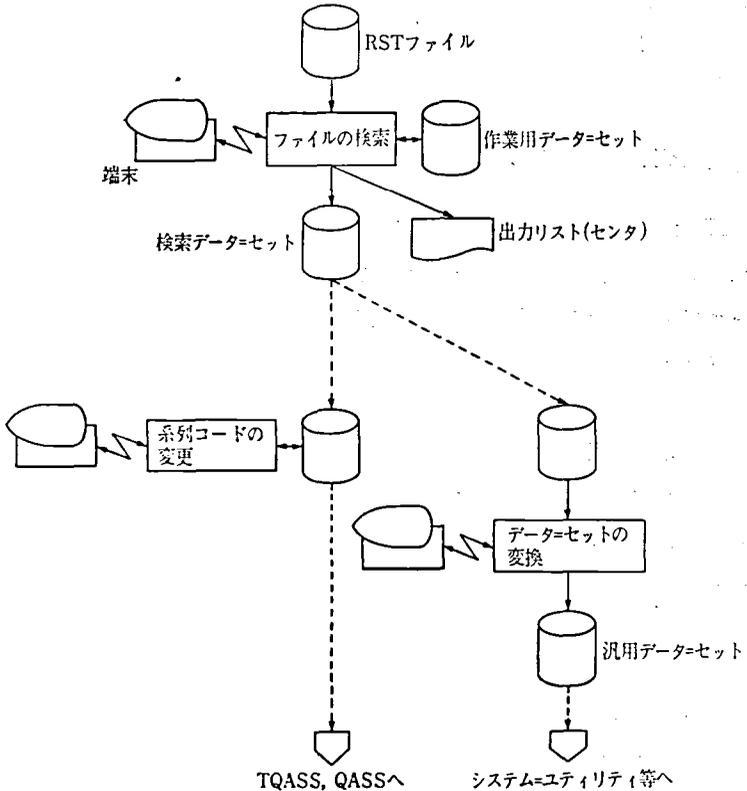


図1 RST の構成 (プロセス=フローチャート)

務選択に応じて実行される。業務により必要な資源が異なるためこのような機能別のプログラム構成を採用したが、それによりプログラムの肥大化が防止され、処理効率の向上とメンテナンス負荷の軽減という効果も期待できる。プログラム相互間の関連性と実行に必要な各種資源を図1に示す。

TSS 環境下でシステムを動作させる場合、実行に必要な資源の割り当てとプログラムの呼び出しを各種の TSS コマンドで行う必要がある。しかし操作性を向上させるためには、環境設定に必要なこれらの作業を簡略化する工夫が求められる。RST ではこれらの作業を、TSS コマンドを一括して登録したコマンド=プロシジャによって行うことにより、利用者の負担軽減と操作ミスの防止を図っている。この結果、RST の利用時に必要な予備知識は TSS 端末の操作に関する初歩的なものに限られることになった。

#### Ⅳ. RST ファイル<sup>3)</sup>

RST ファイルには高い検索効率を達成するために直接編成形式が採用されている。検索コードは収録された資料の検索単位である系列ごとに設定されており、その体系は利用者の便宜を考慮して系列の内容にもとづいた分類体系となっている。このため、全体として必ずしも連続的ではない検索コードを直接編成データ=セットのアクセス=キー（相対レコード番号）にそのまま用いた場合には、検索効率は上がるもののデータ=セット中に空き領域が生ずることになり、格納効率の悪化を招く。他方、データ=セットの記憶媒体である直接アクセス装置は共同利用資源となっているため、個々のデータ=セットはなるべくコンパクトであることが望まれる。RST では、利用者本位の検索コード体系と高い格納効率を直接編成データ=セットで実現するために、利用者プログラムによって検索コードからアクセス=キーへの変換を行う方式を採用しており、RST ファイル

---

3) 現在、RST ファイルには、日本経済新聞社が提供している日経総合経済ファイルを圧縮・編集して収録している。

## RST ファイル

インデックス部	第1レベルインデックスレコード (第2レベルに対するインデックス)
	第2レベルインデックスレコード (第3レベルに対するインデックス)
	第3レベルインデックスレコード (検索コードと相対レコード番号の組)
	データレコード (系列ごとの情報)
データ部	

(注) 第1レベルインデックスは検索コードより求める。

図2 RST ファイルの構造

の構造もそれに適合したものとなっている (図2)。

RST ファイルは、系列ごとの情報が収録されているレコードから成るデータ部と、検索コードからそのレコードを求めるための情報を収めたレコードから成るインデックス部とに分けられる。検索プログラムは検索コードにもとづいてインデックス部を3段階にわたってサーチし、該当系列が実際に収録されているレコードの相対レコード番号を求め、そのレコードを読み出す。

## V. RST の 操 作

### §5.1 RST の起動と業務選択

端末とセンターが接続されると端末に READY が表示される。最初に EXEC コマンド

#### EXEC RST

を入力し、RST 実行用のコマンド=プロシジャを起動する。

図1に示すように業務ごとに使用する資源が異なるため、プログラムの呼び出しに先立って適切なデータ=セットを割り当てる必要がある。このため、RST の起動により業務選択メニュー (図3) が表示され

```

READY
EXEC RST
*** START RST *** 81-11-16 12:55:38
** SELECT YOUR JOB **
*-----*
* 1. RETRIEVAL *
* 2. RETRIEVAL AND CREATION OF DATA-SET *
* 3. CONVERT VARIABLE CODE IN YOUR DATA-SET *
* 4. CONVERT THE FORMAT OF YOUR DATA-SET *
*-----*

*SELECTION =
    
```

図3 RST の起動と業務選択

表1 各業務のデータ=セットに関する入力指示

業 務	入 力 指 示
1	なし
2	*** TYPE IN NAME OF THE RETRIEVAL DATA-SET *** *NAME=
3	*** TYPE IN NAME OF RST/TQASS DATA-SET *** *NAME=
4	*** TYPE IN NAME OF RST/TQASS DATA-SET *** *NAME= *** TYPE IN NAME OF NEW DATA-SET *** *NAME=

る。利用者はその中から一つを選択し、その該当番号を入力する。選択に応じて使用するデータ=セットに関する入力指示がそれぞれ表示される(表1)。

## §5.2 検索

### (1) 検索条件の設定

RST による時系列資料の検索・編集は以下の6種類の条件にもとづいて行われる。

#### a) 同時検索系列の最大数

- b) 期種
- c) 始期
- d) 終期
- e) 出力先
- f) 端末出力情報種別

利用者はこれらの条件を入力指示にしたがって設定し、最後に確認する。確認の際に条件の変更が必要であれば、再度設定する。

a) 同時検索系列の最大数

b) ~ f) の諸条件を同一にして同時に検索・編集する系列の最大数を入力する。実際に指名する系列の個数はこの値以内であればよい。その上限値は400である。

b) 期種

月次・四半期次・暦年半年次・年度半年次・暦年次・年度次の6種類の指定が可能である。時系列資料ファイル内に収録された系列の年間観測度数 (SF) と、ここで指定した期種の年間観測度数 (RF) が異なる場合は以下の処理が行われる。

(i)  $SF < RF$  の時

その系列は検索されない。

(ii)  $SF > RF$  の時

系列の特性に応じて集計する。

c) 始期, d) 終期

西暦年の下2桁と、期種により定まる期次の組み合わせで指定する。系列の長さの上限値は400観測点である。

e) 出力先

短時間に多くの系列を検索したい場合や、検索結果である系列の属性と

観測値を整理された一覧表の形で保存したい場合には、Lを指定して、結果をセンタのライン=プリンタに出力させる。

f) 端末出力情報種別

端末に出力される情報には、

- (i) 検索コードと系列コード、
- (ii) 系列属性表、
- (iii) 系列観測値表、

の3種があり、状況に応じてその種別を指定する。たとえば、すでにライン=プリンタへの一覧表の出力を指定済の場合は、端末に出力する情報を最小限に抑えることにより検索時間が短縮できる。また、すでに属性表を出力済の系列群に対する再検索の場合には、観測値表のみを出力させることにより同様の効果が期待できる。

a) ~ f) の検索条件の入力完了時に、それらの確認を目的とする一括表示が行われる。条件設定に変更がなければ、OK を入力し次の作業に移る。変更があれば、NO を入力し条件の再設定に戻る。

検索条件の設定に関する入力指示を図4に示す。

図4 検索条件の入力指示 (以下の各図の下線部分は入力を示す)

TYPE IN CONDITIONS FOR RETRIEVAL

\*\*\* NUMBER OF VARIABLES \*\*\* : MIN=1, MAX=400

\*NUMBER OF VARIABLES=10

\*\*\* PERIODICITY \*\*\*

SELECT ONE CODE FROM THE NEXT TABLE

+	-----	+	
*	MONTHLY	--> MM	*
*	QUARTERLY	--> QQ	*
*	CALENDER HALF YEARLY	--> CH	*
*	FINANCIAL HALF YEARLY	--> FH	*
*	CALENDER YERLY	--> CY	*
*	FINANCIAL YEARLY	--> FY	*
+	-----	+	

\*PERIODICITY=QQ

\*\*\* FIRST PERIOD \*\*\*

TYPE IN ACCORDING TO THE FOLLOWING FORMAT

```

+-----+
* XX-YY ; XX IS THE LOWER 2 FIGURES OF A. D. *
*      YY IS THE TERM-NO OF INTEREST      *
*      MONTHLY      ; 01 TO 12      *
*      QUARTERLY    ; 01 TO 04      *
*      HALF YEARLY  ; 01 OR 02      *
*      YEARLY        ; 01 OR 02      *
*      YEARLY        ; 01 ONLY      *
*      1981 4TH QUARTER --> 81-04    *
+-----+

```

\*PERIOD=75-01

\*\*\* LAST PERIOD \*\*\*

IN THE SAME FORMAT AS FIRST PERIOD

\*PERIOD=83-04

\*\*\* LIST PRINT OPTION (CENTER) \*\*\*

IF YOU NEED THE LIST, TYPE IN L

\*LIST OPTION=N

\*\*\* SELECT INFORMATION OF TERMINAL OUTPUT \*\*\*

```

+-----+
* 1. ALL *
* 2. ATTRIBUTES LIST *
* 3. DATA LIST *
* 4. RT-CODE AND VARIABLE CODE *
+-----+

```

\*INFORMATION OF TERMINAL OUTPUT= 1

\*\*\* CONDITIONS FOR RETRIEVAL \*\*\*

TYPE IN OK OR NO

```

+-----+
* NUMBER OF VARIABLES 10 *
* PERIODICITY          QUARTERLY *
* FIRST PERIOD         75 1 *
* LAST PERIOD          83 4 *
* LIST OPTION          NONE *
* INFORMATION          ALL *
+-----+

```

\*ANSWER=OK

## (2) 検索コードの入力

RST では系列の検索コードを最大 6 桁の整数で表わしており、RT コードと称する。利用者はこれをあらかじめコード=ブックより選択した上で検索作業を行う。

RT コードは条件設定時に指定した最大系列数を越えて入力しても無視される。また、RT コードの個数とその最大系列数に達しない場合には、最後の RT コードの後にスラッシュ (/) を付加して、RST にその入力完了を通知する必要がある。

RT コードは空白又はコンマ (,) で区切って 1 行に複数個入力できる。全ての RT コードが 1 行に収まりきらない場合には、それらを複数行に分けて入力する。ただし 1 個の RT コードを複数行にまたがって入力することは許されない (図 5)。

図 5 検索コードの入力

```
*** TYPE IN 10 RT-CODE SEPARATED BY ONE SPACE OR COMMA  
          AND TYPE IN / AFTER THE LAST RT-CODE ***  
          *** FOR EXAMPLE ; 10009, 10002, 10003, 10004/  
*RT-CODE=66044 66021 66065 66002/
```

## (3) 検索結果の出力

端末に出力される検索結果は系列属性表と系列観測値表とに分けられる。両者の形式を図 6・図 7 に示す。

個々の観測値の識別を容易にするために、系列に対しては先頭から V001, V002, ... という 4 桁の系列コードが、観測点に対しては期種に対応

図 6 系列属性表

```
*** ATTRIBUTES LIST *** FOR TERMINAL  
*--* INFORMATION OF VARIABLES *--*  
  
NO= 1 RT-CODE=66044  
TOTAL POPULATION OF JAPAN (BEGINNING OF MONTH)  
UNIT ; 10000 PERSONS  
PERIODICITY ; MM  
CONVERT CODE ; L  
VARIABLE CODE ; V001
```

図7 系列観測値表

	V001	V002	V003	V004
	R066044	R066021	R066065	R066002
Q751	11116.000	8405.332	5162.332	107.333
Q752	11147.000	8431.332	5385.000	94.667
Q753	11181.000	8455.332	5399.332	94.333
Q754	11213.000	8480.332	5344.000	103.000
Q761	11237.000	8506.332	5248.332	125.667
Q762	11268.000	8529.664	5425.332	108.333
Q763	11298.000	8550.332	5450.332	101.000
Q764	11327.000	8572.332	5389.332	96.333

表2 観測点コードの付加規則

期種 \ 桁	1	2~3	4
月次	M	西暦年の下2桁	1~9, O, N, D
四半期次	Q		1~4
半年次	H		1~2
年次	A		空白

図8 検索データセットへの出力形式

a). 出力並び

WRITE (11) NV, NS, N, (V(I), I=1, NV), (S(I), I=1, NS), (RT(I), I=1, NV)  
 WRITE (11) (X(I), I=1, N)

b). 変数の型と長さ

変数名	型	長さ (byte)	内 容
NV	整数	4	系列数
NS	"	4	観測点数
N	"	4	NV×NS
V	実数	4	変数コード
S	"	4	観測点コード
RT	"	8	RTコード
X	"	4	観測値

c). 観測値の並び

第1系列 第1観測点~第NS観測点	第2系列	...	第NV系列
----------------------	------	-----	-------

した3～4桁の観測点コードが、それぞれ新たに付加される(表2)<sup>4)</sup>。また、時系列資料の分析によく使用される、1より逡増する傾向変数(系列コード; TIME)と固定値1をとる単位変数(同; UNIT)が、検索プログラムによって生成され、系列観測値表と検索データ=セットの最後に追加される。

検索データ=セットの形式と内容を図8に示す<sup>5)</sup>。

#### (4) 検索終了の確認

検索結果の端末出力に続いて検索業務の終了を確認するための入力指示が表示される(図9)。条件を変更して再検索を行う場合は、C、検索業務を終了させる場合は、Q、と入力する<sup>6)</sup>。

図9 検索終了の確認

```
*** RST RETRIEVAL END ***
      YOU CAN SELECT Q OR C
      +-----+
      * END THIS STEP                -> Q *
      * RETURN TO BEGINNING OF THIS STEP-> C *
      * (YOU CANNOT ADD VARIABLE TO DATA-SET) *
      +-----+
*SELECTION=Q
*** END RST ***          81-11-16 13:52:34
READY
```

### §5.3 系列コードの変更

検索データ=セットに収められた系列を TQASS などを使用する場合、その識別には検索プログラムが自動的に付加した4桁の系列コードを用いてもよい。しかし、このようなコード体系よりも利用者自身の命名による

- 4) これらのコードは、検索結果を TQASS, QASS で利用する場合には不可欠なものである。
- 5) 検索データ=セットは、ブロック長 208 byte の可変長スバンド非ブロック形式をとっている。
- 6) 検索結果をデータ=セットに出力しているときには、出力済の結果を消去したい場合を除いて、必ずQと入力する。本システムでは、検索条件が同一であっても、すでに作成済の検索データ=セットに対する系列・観測点の追加は許されない。

図10 系列コードの変更

\*\*\* START VARIABLE CODES MODIFICATION PROCESS \*\*\*  
\*\*\* SELECT THE PROCEDURE OF MODIFICATION IN THE FOLLOWING MENUE \*\*\*

- ```

+-----+
* 1. CONTINUOUS (FOR WHOLE MODIFICATION) *
* 2. DISCRETE (FOR PARTIAL MODIFICATION) *
* 3. DISPLAY ALL VARIABLE CODES *
* 4. FINISH MODIFICATION *
+-----+

```

\*SELECTION= 3

```

***** VARIABLE CODES *****
NR P X000 V090 X003 X004 X005 X006 X007 X008
EWI EMN HWWI HWMN TIME UNIT

```

\*SELECTION= 1

\*\*\* CONTINUOUS MODIFICATION START \*\*\*

\*\*\* TYPE IN THE FIRST AND LAST CODE IN DATA-SET SEPARATED BY SPACE OR COMMA \*\*\*

\*\* FOR EXAMPLE --> V001, V010 \*\*

\*FIRST AND LAST CODE=X000, X008

\*\*\* TYPE IN 8 NEW VARIABLE CODES SEPARATED BY SPACE OR COMMA AND TYPE IN / AFTER THE LAST CODE \*\*\*

\*VARIABLE CODE=RH, QADS, DRS, QW, TIHC, TIHJ, WPDN, WPDG/

\*SELECTION= 2

\*\*\* DISCRETE MODIFICATION START \*\*\*

\*\*\* TYPE IN OLD AND NEW CODES SEPARATED BY SPACE OR COMMA \*\*\*

\*\* FOR EXAMPLE --> V010, GNPR \*\*

\*OLD AND NEW CODE=HWWI, HAWI

\*\*\* HWWI IS REPLACEDBY HAWI \*\*\*

\*\*\* IF YOU WANT TO FINISH THIS THIS PROCEDURE, TYPE IN Q OR TYPE IN C \*\*\*

\*ANSWER=C

系列の内容に即したコードの方が都合の良いことが多いと考えられる。そのため RST では、検索データ=セットを対象とした系列コード変更機能

を提供している。

(1) 入力方法

利用者は検索時に出力される系列属性表などにもとづき、変更する系列コードを決定する。新コードの入力方法には、

- a) 連続的入力、
- b) 個別的入力、

の2種類があり、目的に応じて使い分ける。連続的入力の変更対象の系列コード群の先頭と最後尾のコードを指定して、それに対応する新コードを連続的に与える方法であり、個別的入力は特定の系列コードを指定して、それに対応する新コードを与える方法である。また必要な場合には、全ての系列コードを表示させることもできる。

(2) 使用例

系列コードの表示、新コードの連続的入力および新コードの個別的入力に関する入力指示などを図10に示す。

#### § 5.4 データ=セットの変換

RST は検索データ=セットのレコード形式とレコード=レイアウトを、他の利用者プログラムなどでただちに使用できる汎用性のあるものに変換する機能を提供している<sup>7)</sup>。

(1) 変換方式

変換される検索データ=セットには系列コードや観測点コードも含まれているが、作成される汎用データ=セットには系列の観測値だけが含まれる。汎用データ=セットにおいては、観測値は小数部をもつ10進数の形式で表現されるため、検索データ=セットにおける内部表現の変換にFORTRAN のF変換を用いる。

利用者は変換に際して、

---

7) TQASS および QASS で作成されたデータ=セットも変換対象とすることができる。汎用データ=セットはレコード長 80 byte, ブロック長 800 byte の固定長ブロック形式をとっている。

- a) F変換の書式仕様,
- b) 観測値の配列形式,

の2種類のパラメータを指定する。

- a) F変換の書式仕様

汎用データ=セットの1レコード (=80 byte) に含まれる観測値の個数, 各観測値の小数点以下を含めた最大桁数, および, その小数点以下の桁数を空白又はコンマ ( , ) で区切って順に指定する。

- b) 観測値の配列形式

複数個の系列に対して複数個の観測点があるとき, それらの観測値を汎用データ=セットの各レコードに収めるには2通りの形式がある。一つはそれらが系列ごとに観測点順に並んでいる形式であり, Vと指定する。いま一つは観測点ごとに系列順に並んでいる形式であり, Sと指定する。

## (2) 使用例

F変換の書式仕様と観測値の配列形式に関する入力指示等を図11に示す。

図11 データ=セットの変換

```

*** START CONVERSION OF DATA-SET ***
* INPUT FORMATTING PARAMETERS FOR ONE RECORD
                                     (F-CONVERSION) *
* -----*
* N1 : NUMBER OF DATA IN ONE RECORD      *
* N2 : TOTAL NUMBER OF FIGURES IN ONE DATA *
* N3 : NUMBER OF FIGURES IN ITS DECIMAL PART *
*      EXAMPLE 10F8.3 --> 10, 8, 3         *
* -----*
*N1, N2, N3=5, 16, 4

*SELECT TYPE OF OUTPUT FORM *
* -----*
* V : VARIABLE TYPE      *
* S : SAMPLE TYPE       *
* -----*
*TYPE OF OUTPUT=V
    
```

## Ⅵ. お わ り に

RST は特に時系列資料ファイルを対象とした小規模の検索システムであり、数量経済分析における原資料の作成・提供を主な狙いとしている。本システムがその分野における研究・教育の一助となれば幸いである。

なお、今後の機能強化と操作性向上のために利用者の方々の積極的なご助言を希望するものである。

(1981.11.30)