学際研究用データベース・システム SMLESIS

II メッシュデータ検索 (GRIDREQ) 及び ネットシソーラス処理 (NESSY)

島津 康男*・浦部 達夫*・杉山 公造***
平松 敏祐**・上松 峰夫*

SHIMAZU, Yasuo,* URABE, Tatsuo,* SUGIYAMA, Kozo,***
HIRAMATSU, Toshisuke** and UEMATSU, Mineo*: SMLESIS
——Database System for Interdisciplinary Researches.

1. SMLESIS とは

われわれは、1966年以来SMLES研究グループを作って研究行動を行ってきた。SMLESはseamless earth science (ぬいめなしの地球化学)の略で、その名の通り地球に進行する現象をトータルとしてみようというものである。その主な研究手法がシステム・シミュレーションなので、SMLESはsimulation earth scienceの略でもある。

やがて、SMLESグループは、地球というトータルシステムを、陸海空の無生物自然界((地球)物理システム)・生物界(生態システム)・人間社会(生産・消費の技術、経済プロセスから社会構造、文化までを含み、広義の社会システム)の集合としてみることになった。そして「環境学(environmentology)」とわれわれが名づけた領域へ進むことになり、データは多種多様となって、必然的にこの学際研究を支援するSMLESIS (SMLES information system)が必要になってきた。

SMLESIS自身は、環境学に止まらず、自然、

社会、文化の各科学にまたがる一般の学際研究を支援するデータベース管理システムである。その意図の詳しいことは〔文献1〕,〔文献2〕にゆずり、ここでは概要のみを記しておく。

(1) バックグラウンドは多様でも、共通のイメージ(問題意識または哲学)をもつ10人のオーダーの研究者グループのためのものである。いわば研究グループが一つの大きな本を共有するためのもので、研究者ファイルの拡大版といってもよい。そこで二次情報をうるためではなく、0、一次情報を使う仕事を支援するものである。

(2) 文献・数値データ・非数値データ・文書(環境法令及びその適用対象別inverted file)・地図(メッシュデータ)・研究者(アドレス・学会・バックグラウンド)などを受け入れる。どこへ行けばデータがえられるかの情報ルートの情報も貴重である。データ内容は多様だが、量は数10Mb止りでそんなに多くない。

(3) 多様な検索のために、ソフトなシソーラスを用意する。このシソーラスは、目的とする問題の概念構造をモデル化したものと理解される。シソーラスをとりかえることによって、多様な利用

* 名古屋大学理学部; Faculty of Science, Nagoya University

** 名古屋大学大型計算機センター; Computation Center Nagoya University

*** 富士通国際情報社会科学研究所; International Institute for Advanced Study of Social Information Science, Fujitsu

が可能である。多様なバックグラウンドの学生ごとに教育カリキュラムをシソーラスの形で提供することもできる。

(4) (3)に共役な性質として、DATA構造はなるべく固定しない。特にポインターなどはつけない。またDEMANDの内容が一定しないので、マルチファイル方式は妥当でない。

(5) 文献には主観的なコメントをつけ、検索対象にする。学際的とは、一つの方分野での成果を全く別の観点から評価することでもあるから、著者自身の要約と同じく、読者のコメントが重要である。同じイメージをもつグループ内のことだから、1人のメンバーのコメントは他の人にも役だつ。

要するに、学際研究とは、多分野の研究者が集まることではなく、1人1人が多分野をこなす「個人学際」が本来の姿であると、われわれは信じている。

2. 文献検索システム

文献検索システムの詳細については〔文献1〕を参照されたい。なお〔文献1〕はバッチ処理用のものであるが、現在は追加・削除・修正・検索ともオンライン方式になっている。FACOM 230/75に50ボー端末をつけて使用するが、プログラム自身はFORTRANで書いてある。これはプログラムの公開（他機種への変換）を容易にするためである。主なコマンドは次の通りである。

- >NUMBER 対象とするデータのID_Nを指定する。Nは次の行にFREE FORMATでうつ。
- >OLD マスターファイル上のデータを、コア上の作業領域へ転送
- >SAVE >OLDの逆転送
- >KILL 作業領域上のデータを消去（フラッグを立てるだけ）

```
>NUMBER
>1942
>LIST
  1942  IS NOT FOUND
>OLD
>LIST
  1942
WEISS*H.V.,M.KOIDE,E.D.GOLDBERG (1971) MERCURY IN A GREENLAND ICE SHEET;
EVIDENCE OF RECENT INPUT BY MAN/SCIENCE 174,692-694/TOXITY;BIOGEOCHEMICALCYCL
E@
>CORRECT
>"TY;"TY,"
>CORRECT
??
>CORRECT
>"CHEMICALCYCLE"CHEMICAL CYCLE"
>DEBUG

  1942
WEISS,H.V., M.KOIDE AND E.D.GOLDBERG (1971) MERCURY IN A GREENLAND ICE SHEET;
EVIDENCE OF RECENT INPUT BY MAN
SCIENCE 174,692-694
( TOXITY,BIOGEOCHEMICAL CYCLE )
>SAVE
>BYE
```

```
* END OF FORTRAN *
. プログラム メイ: NONAME , カンリョウ コード: 000.
#BYE
. CPS カンリョウ コード : 000
*** CPS RIYOU KIROKU ***
```

>REVIVE 誤ってKILLを実行したデータの復活
(フラッグの復活)

>CORRECT A Aの文字列をB B Bにおきかえる。文字数は双方とも自由。削除の時は>'AA'。'AA'がデータ中にないと'STRING IS NOT FOUND'と印刷される。

>NEW 新規データの入力。「@」をうつまでFREE FORMAT

>LIST データ内容の印刷

>DEBUG データ内容を所定の型式で印刷

>DELETE マスターファイルからの削除(フラッグを立てるだけ)

>KEEP >DELETEコマンドの結果とり消し

>LOOK 作業領域上のデータの一覧表印刷

>CONDENSE KILLでできた空き領域を圧縮(これを実行したあと>REVIVE,>KEEPは無効)

>BYE 終了して一般TSSコマンドに移る。

実例を図1に示す。

3. メッシュ・データ検索システム (GRIDREQ)

メッシュ・データとは、地図上に方形の網目(メッシュ)をかけ、これによって地図情報をデジタル化したものである。地球科学はもちろん環境学においても、メッシュ・データのもつ意義は大きい。すなわち、面積が規格化されているので、ことなつた地域・時点のデータの比較が容易であり、自然現象と土地利用状況など異分野で集められたデータとの組み合わせも容易である。統合、分解、標本抽出などの操作や、位置や距離に関連する情報の操作は、コンピューターを前提とする

ことによって、楽である。ただし、地図とするのが適当なタイプの情報かどうかは、十分注意しなければならない。また、地図情報を作成するには、原データからの加工に多くの労力を要するので、既に作成されたものを加工して利用することが多く、データ評価の点で問題を生ずる場合がある。

SMLESISでは、国土地理院図の経緯度をもとにした「標準地域メッシュ体系」によるデータを蓄積し、これらの検索・オーバーレイ・四則演算・図化を行う検索システム(Grid*Request: GRIDREQ)をもっている。今のところGRIDREQはバッチ専用である。

(A) データ・ファイル

約1Km×1Kmの総理府標準メッシュにもとづいて、

自然環境(地形・地質・土壌・水文・植生・天然記念物・気候・公害状況など)
土地利用(現況・変化・法的規制・計画など)
社会環境(人口・地価・交通量・生産・消費・水利用・文化財など)

の経年変化を含む229項目を愛知県(5,200メッシュ, 5.5MB)・東京都(2,100メッシュ, 1MB)についてファイルした。そう大量でもないので、DATA構造もRelational modelのように高級なものとはせず、

メッシュ コード	流域 コード	市町村 コード	境界 コード	←データ→
-------------	-----------	------------	-----------	-------

のように単純なものにしてある。ここで

メッシュ・コード 1Km×1Kmメッシュの位置を18で示す。

流域コード そのメッシュが含まれる河川のコード

市町村コード そのメッシュが含まれる市町村のコード

境界コード 県境、海岸などの境界がメッシュの中を通過していることを示すフラッグ。これでリンクを書く。

*メッシュは和製英語で、本来はgrid squareとよぶべきである。

であり、データには

(1) 1つのメッシュコードに属性値が1つ対応するもの — 高度、人口など

(2) 1つのメッシュコードに属性名が1つ対応するもの — 優先植物名、規制の有無 (yes, No) など

(3) 上2つの混合 — 種Aが20%, 種Bが50%, 種Cが30%のような植生図, 地質図など。の3つのタイプがあるが, ここでは(3)は(1)に還元し, 属性Aの数値を20%というようにしてある。これによってデータの項目がふえるが, 全体の量がそう大きくない場合は, DATA構造を単純にするというメリットの方が大きい。また, 各データのタイプ, 内容その他のコメントは, ヘッダーとして属性ファイルの形とし, データ・ファイルへはこれを通してアクセスする。

(B) 検 索

例えば

10 (SHIZENDO. GE. 6). AND. 1 (GE-NZON SHOKUSEI. EQ. TABU) という REQUEST は, 環境庁の定義による自然度が6又はそれ以上で, かつ現存植生がタブであるようなメッシュを抽出し, 各々の条件に評価の重み10, 1を与えて総合評価点を求めよ, とのいわゆるオーバーレイを示している。一般に検索命令は,

抽出条件: .GT., .GE., .LE., .LT.,
.EQ., .NE.

組合条件: .AND., .OR., +, -, *, /
からなっている。AND, OR, + は20種類まで結合可能, -, *, / は2種類までに制限される。なお, 流域, 市町村の検索は任意指定は入力パラメーターによって行われる。

結果の統計処理, 図化結果はラインプリンターに出力される。この種の機能はSPSSにあるが, 入力型式がハードで使いにくいので, あえて自作したのである。従って多変量解析・林理論などのサブプログラムも自作である。データの蓄積, このシステムを使用しての環境の総合評価の問題も

進行中である。

メッシュ・データは元来表型式データ (table data) の一種であって, 前述のようなデータ構造は

turtle →

愛知県	データ コード	人 口 (S40)	人 口 (S45)	地 質	

domain
↓

のようなDATA構造に対応している。従って利用目的に応じて, 次に示すようないくつかの型の

	S50	人口	交通	S0x
S45	人口	交通	S0x	
S40	コード	人口	交通	S0x

時系列データとして利用したい

一次 コード			
一次 コード			
一次 コード	一次 コード		

特定地域を検索したい

マルチ・ファイル構造がありうる。また同じ表型式データでも, 産業連関表のように domain

と **turple** とが独立でないマトリクス型式のデータを受け入れることが、環境学としては必要になってくる。

環境問題のデータベースといえば、大気汚染の監視など公害規制への利用が流行している。この場合、空間パターンと時系列パターンとが同等に重要であり、データベースとしては時系列更新がネックになる。われわれは環境学データベースを環境管理に利用することも行っているが、そこでは、

(A) 地域計画の支援

(B) 環境アセスメントなど特定地区の環境評価の支援

(C) 環境監視の支援

の3つに分類し、GRIDREQは(A)、(B)に適用されるものと考えている。一般に(C)は別システムを必要とするものとする。〔文献3〕

4. ネット・シソーラス処理システム

(NESSY)

SMLESISでは、概念構造を表現する論理シソーラスが重要な意義をもった。もっとも単純な論理構造はトリー (tree) で表わされるが、図2のようにループを含む構造はトリーでは表現

図 2

できない。トリー表現を包含し、より一般的な概念構造を検索に適用するには、ネット (網) 構造で記述したネット・シソーラスを扱う必要がある。そこで会話的にネット・シソーラスを処理する **net thesaurus processing system (NESSY)** を開発した。NESSYはネット構造

の生成・修正にも有用である。

局所的な論理 (因果) 関係がわかったとして、その二項関係の群が全体としてどんな論理システムになっているかがわからない場合は多い。

そこで二項関係を

$A > B$ 原因 A が結果 B を生ずる。

$A < B$ 結果 A は原因 B によって生ずる

$A < > B$ A , B は双方向関係

$A = B$ A , B は同義

の4種類とし、入力データを次のように与える。

(FUDO < CLIMATE > POLLUTION)

(POLLUTION > ECONOMIC POLICY)

(POLLUTION > ENVIRONMENTAL POLICY)

(NATURAL CONSERVATION < VEGETATION)

(LAND USE > NATURAL CONSERVATION)

(ENVIRONMENTAL POLICY < NATURAL CONSERVATION)

(FUDO < VEGETATION)

(VEGETATION > FOOD < POPULATION > LAND USE)

(ENVIRONMENTAL POLICY < > LAND USE)

()

ここで空白の () はデータの終りを意味し、1行に複数の関係を書いてもよい。以上のデータを入力すると、内部ではすべてを二項関係に分解し、さらに文字列をHash関数によって1~1024の整数値に変換して、**relation table** に入れる。< > は < と > に再分解し、< は > に反転して、**table** をソートする。その結果、> と = とだけにする。ここで、グラフ理論における隣接行列とか、特別のリストを作って構造全体を表現することはしない。

上記のデータ群は外部的には図2を表現しているわけである。**brain storming** で次々に出てくる局所的な概念を、自動的にまとめあげる操作だと思えばよい。なお、図2は国土計画に関するシソーラスの簡略形をあらわしている。

次にネット・シソーラスの検索について述べる。次のような6種類のコマンドがある。

/F (KW=) KWと同義のものを探せ

/F (KW>) KWより下位のものを探せ

スと違ってスピードを重要せず、きめ細かさを重視した。

さらに、SMLESISの開発自身を学際研究の方式の見直しとみることもあって、われわれ自身による開発を旨とし、いわゆるプロに任せることはしない。これまで、研究内容自身に“自前主義”をとることが、SMLES研究グループの根本方針だったのである。performance をぎせいにしても、いわゆる学際研究の方式を分析し、これをきめ細かく表現することが第一である。

現在までに開発の終わった、

文献検索システム

メッシュ・データ検索・解析システム

ネットシソーラス処理システム

は、いずれも公開提供可能であり、FACOM 230/75 + 50 ボー端末ならそのまま利用できる。さしあたり、300 ボー端末及びディスプレイへのアップを行う。データそのものも、

環境学文献（4000件、毎月30件程度追加されている）

研究者（500名）

字引きシソーラス（環境学用英和・和英、同義語、語尾変化）

論理シソーラス（環境学、国土計画、地震災害対策、環境六法インバーテッドファイル）

メッシュデータ（愛知県、東京都）

は提供可能である。

われわれは、必ずしもシステム全体を一つにまとめるのが有利とは考えていない。むしろ、定量・図解シソーラスの機能アップとともに、いわゆる研究オリジナルデータとの関連を考えている。例えば、化学分析の研究データは、ある分析手法をもった一人の研究者があらちのサンプルを分析するといった蓄積タイプをもっており、

WHO/HOW/WHAT/WHEN/WHERE

Dr.A-X法-PCB	1976.9	└─α村—1 PPB └─β村—2 PPB
	1977.3	
Dr.B-Y法-PCB	1976.10	α村—3 PPB

のDATA構造をもつものと表現される。ところが環境学では

PCBの地域分布 (WHAT/WHEN/WEFRE)

α村のPCB変化 (WHERE/WHAT/WHEN)

のように全くことなったDATA構造が要求される。これは

DISCIPLINE → MISSION

ORIENTED DATA ORIENTED DATA

との構造差といってよく、研究者、研究法に伴うデータ評価の問題が入ってくる。この点でGEODASとの結合をはかりたい。

なおデータベースの開発といえば、できあがった磁気テープの導入を考える人が多い。しかもそれは専ら二次情報をうるためのものである。研究領域が確立しかつ同業者も多い分野にあっては、この種のデータベースが必要なことは確かである。しかし、これですべての研究がカバーできるものでないことも確かであろう。大きくわければ discipline oriented research には二次情報が有効であり、mission oriented research には有効でないということになるうか。

discipline oriented と mission oriented とがタテヨコの関係にあるとの考えがある。とすれば、discipline oriented のデータベースが、各領域について完成してはじめて、mission oriented のデータベースとして利用できることになるが、果してそうであろうか。第一、各disciplineごとの内部でのデータ規格化ができていない現状で、それらを横につなぐのは無理というものである。

われわれは、必ずしも

学際研究=mission oriented research とは考えず、むしろ1つの新しいdiscipline を作ることを考える。だからこそ、多くの specialist の集合としてでなく、1人1人が多discipline をこなすことを目標にしているのである。従ってDATA自身も自前で作らねばならない。環境学のDATAは環境問題のDATA

と同じではないのである。

文 献

1. 浦部達夫・島津康男・茂泉俊夫(1976), 学際
研究用データ・ベース・システム SMLESIS
文献検索システム・マニュアル(パ
ッチ・システム): 1-33.
2. 島津康男(1977), MISSION ORIENTED
DATA SYSTEM- 環境問題を例として
情報システムの形成過程と学術情報の組織
化, 52年度夏期研究会報告: 21-30.
3. 島津康男(1977), 環境管理と情報システム
環境情報科学6, (1): 56-59.
4. HIRAMATSU, Toshiyuki (1976),
Thesaurus oriented docu-
ments retrieval system
(THEODORES) for laboratory
database. Mem. Res. & Deve-
lopment Division, Computation
Center, Nagoya University
(2): 24-48.
5. 平松敏祐(1977), THEODORES-Tree
thesaurusからnet thesaurusへの
拡張, 情報処理学会18回 全国大会 予稿:
293-294.