

会話型データ処理—その9—

GEOCAPS用のデータファイル作成

吉井 守正 (鉱床部)
Morimasa YOSHI

はじめに

前回にプログラム GEOCAPS(地球化学データ解析プログラムシステム)のあらましをご紹介した。今回からそのシステムの利用法を中心に述べよう。旧システム(岩石化学データ処理システム)については1980年から1982年までにすでに本誌の会話型データ処理シリーズ1~6で記してある。したがってそれらの記事との重複はなるべく避けながらこのシステムの特徴や旧システム以後の改良点について述べることにしよう。

1. 準備

岩石・鉱物・水質そのほかの化学分析値やそれに付随する物理量などのデータまたはこれらに類似した性質をもつ二次元配列の数値データはこのシステムで処理可能である。まずつぎのような準備作業を行う。

a) 試料番号の整理

試料番号は9文字以内の文字列(数字およびローマ字等)となるように調整する。なるべく同番ができないようにする。システムではデータを試料番号順にソートできるので同種のデータは最初の文字を同じにしたり全体の文字数を一致させておくとあとから便利である。

なお文字数は物理的には10字まで書くことができる。最初に9文字までに制限するわけは入力データを何らかの理由で一時廃棄しようとするときに試料番号の前に-(マイナス)記号を追加して処理から除外させる機能を利用できる可能性を残すためである。

b) コードの設計

データに付けた分類コードによって必要なデータを検索して処理をする方式は筆者らによる旧来からのシステムでの一大特色である。この機能を十二分に活用するためにデータの性質を全体的に良く見極めた上で周到なコードの設計をしてもらいたい。コードの考え方と付け方のコツについてはすでに本誌315号(吉井1980)で述べたのでここでは要点だけ記す。

コードは全体が10文字の文字列である。これはつぎの4種のサブコードに細分されている。

- 第1サブコード (1文字)
- 第2サブコード (3文字)
- 第3サブコード (3文字)
- 第4サブコード (3文字)

データに分類コードを付けるときに各サブコードに対して使用者はまったく任意に分類の定義をしてよい。最近の実例では

- 第1サブコード: 地質区または地域
- 第2サブコード: 地質時代・地層または岩体
- 第3サブコード: 岩質
- 第4サブコード: 地区 文献コード その他

といった形式のものが多い。

コードの設計で念頭に置くべき点はデータをコード別に検索する際にはコードの範囲を指定して行うという方式に適合させることである。ここで言う範囲とは文字列のもつASCII文字コードの番号範囲のことである。

ASCII文字コードによる文字の“値”の大小関係はつぎのとおりである。

・<0<1<…<9<A<B<…<Z<a<b<…<z<~

すなわち数字 ローマ字の大文字および小文字の順に値が大きくなる。このシステムでは・(ピリオド)を最小値 ~ (ティルダ)を最大値と定義している。

だからたとえば地質時代のように一定の順序に配列させる必要のあるときはサブコードの少なくとも最初の文字は上記の序列に沿ったものでなければならない。それもできれば1, 4, 7, ……とかA, D, G, ……といった具合に間を少し飛ばして付けるようにしておく方がよい。将来その中間に新しい区分を追加することができる。

第2サブコードから第4サブコードまでは3文字が使えるのでデータの細分を要しないときはそれら各サブコードの第1文字目で主要な分類および序列が定

第1表 鉱物のデータに付けたコードの実例

第1サブコード(1文字): 出所	
R	文献
Y	筆者自身の分析
第2サブコード(3文字): 鉱物名	
BTM	バスタム石
PXM	パイロクスマンガン石
RDN	バラキ石
TPR	テフロカンラン石
第3サブコード(3文字): 化学分析法	
△△	湿式
E△△	EPMA
△印に適当な細分を定義する。	
第4サブコード(3文字): 文献コードその他備考	
M64	文献コード表を別に作って対照する。
Y78	

まるようにする。そして残る2文字はその分類内容の略号として判読できるようにすると便利である。

最良の実例とは言えないが筆者が現在使っているコードの例を第1表に示す。

C) 成分名

ここでいう成分名とは 化学成分だけでなく 一般の項目名も含めた名称を意味する。したがって 帯磁率・屈折率・温度・pH といった項目であってもよい。目下のところ 成分名の文字数は 9字以内、項目数は 35以内である。

とくに Total または Sum の項を加えておくとデータ入力の際で 各試料の最初の項から それらの項の直前までの項の数値の合計を算出し Total(Sum)の項に記入する機能がある(吉井, 1981)。微量成分や物理量は Total (Sum) の項のあとに付ける。

なお 成分式の計算結果を 使用者の指定する項目に属する数値配列へ記入することができる。そこで 最大試料数が比較的少なく 成分数に余裕がある場合は 計算結果を書き込むための専用項目を登録しておくことでデータ処理結果をもとにした 次の段階の処理もできるようになる。成分名に関する注意事項を次に示す。

ア) 同名の項目を(大文字・小文字の別にかかわらず)作ってはならない。とくにノルム用プログラムでは内部的に ノルム鉱物名および D.I. の項目が

追加されるので これらと重複しないよう注意を要する。

- イ) 計算式で用いる記号すなわち + - * / () ^ = , を含む名称は避ける。
- ウ) 数字で始まる名称は避ける。 “0” という名称を作ってはならない。
- エ) log, ln で始まる名称は 大文字・小文字の別にかかわらず避ける。
- オ) [] を含む名称を作ってはならない。

以上について 少し補足する。データ処理の際には 使用者は 処理すべき成分名を 単独または算式の形で指定する。入力された成分名の検索が行われるので同名が2箇所以上あると 2番目以降の成分の検索ができなくなる。与えられた算式から成分名および定数と演算記号とを分離させる都合でイ, ウ, エの制約を生じる。ただし その場合 [H₂O+] とか [2V] といった具合に ブラケットで囲むことにより 上記の制約に抵触する成分名の指定ができる。したがって このブラケット自身を成分名に使うことはできない。

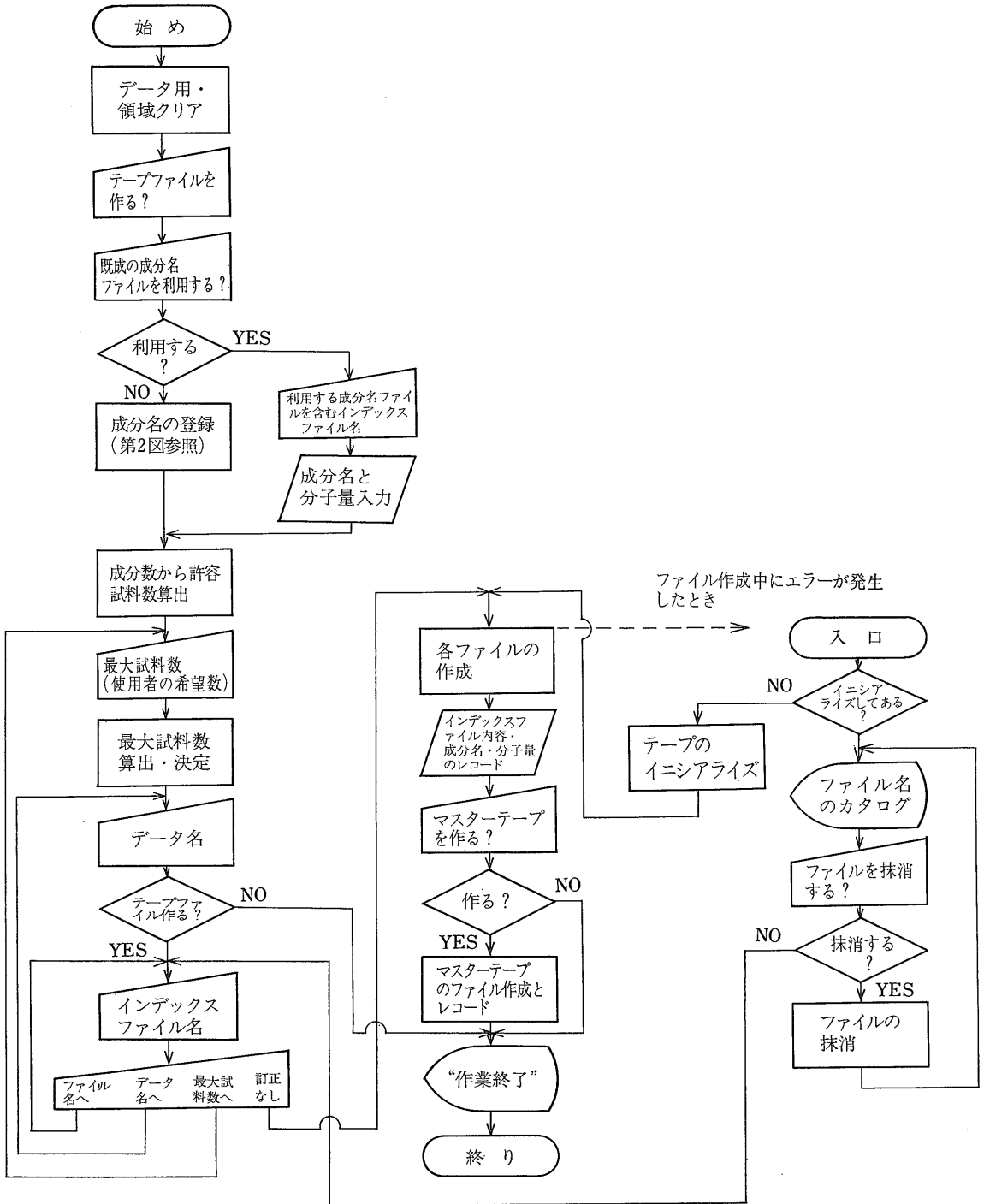
d) 準備段階での全般的な注意

ひと口にデータを整理するといっても その規模や出所によって 状況は千差万別である。とくに注意すべきは データの量が多いか または多くなることが見込まれる場合である。なかでもデータが自分のものでなく 異った複数の文献から引用する場合は問題になる。すなわち 出典によって化学成分の組合せや順序が異なる。このようなときは 面倒でも 原稿をコピーして切り張りし 順序を統一するなどの工夫が要る。その上 将来データが増したとき 分類や成分を増さねばならないかも知れない。それらを先読みしておく必要もある。

しばしば起こるのが同一データの多重登録である。とくに文献からの引用のときに起り勝ちで 原著の分類をう呑みにして コードも違えてしまったときは 見破りにくい。このシステムでは そのような場合でも 多重に登録されたデータを発見し 削除する機能が付いている。しかしこのような手戻りのないように 十分気を配るに越したことはない。

2. データファイルの作成

データの準備ができたなら 計算機と向い合う。以前から何回も述べているように このシステムは目下のところ 横河ヒューレット・パッカー社製 YHP-9845T



第1図 データファイル作成の行程

終了後プログラムは一時停止 (PAUSE) する。プログラムを進めるキーを押すと データの入力・訂正等に必要のセグメントを連結して その行程へ進入する。

第2表 ノルム計算用の“標準組成”

成分名	区間	説明
SiO ₂ TiO ₂ Al ₂ O ₃ Cr ₂ O ₃ * Fe ₂ O ₃ FeO MnO MgO NiO* CaO Na ₂ O K ₂ O P ₂ O ₅	A	1. ノルム計算可能なデータは 区間Aでの成分の組合せと序列が 左記のとおりでなければならない。 2. *印を付けた成分は 共に存在するか除かれるかの いずれかとする。 3. 区間Bの成分は変更できる。またこの区間に任意の成分を追加してもよい。 4. 区間Dには任意の成分を追加してよい。 5. 上記1-2の条件を満足するデータタイプは NR (*印の成分を除く) NC (*印の成分を含む) となり ノルム計算が可能となる。
H ₂ O+ H ₂ O- Others	B	
Total	C	6. ノルム計算用プログラムでこのタイプのデータ処理をする際には 左記成分のほかに ノルム成分の指定ができる。
	D	7. 成分の合計値を100%に再計算する場合は 区間Aの成分の合計値について行われる。

用である。プログラムテープをかけ 自動立ち上がりボタン(AUTOST)を押し込み 電源ON。プログラムメニューの1番(データファイルツクリ)を選択し 日付などを入力すると プログラムテープが走り 必要なセグメントの連結 という段取りとなる。

ファイル作成行程の流れ図を第1図に示す。

a) 成分名の登録

データファイルの作成をする行程は 最初に必ず通らねばならない。不慣れた人にとっては面倒な部分でもあろう。とくに成分名を登録する行程は 使用者が自分のデータを持ち込むのだから 手間がかかり 細心の注意を要する。このシステムでは 成分名はとくにプログラム自体には書かれておらず まったく白紙だから 使用者の責任で 成分名の登録をする。その際に 成分名そのものについての制約があることは 前項で述べたが そのほかに つぎのようなルールがある。

- ア) 化学成分名は 大文字・小文字の区別を厳密に守って入力すること。これにより あとで説明する分子量の自動算出機能が利用できる。
- イ) ノルム計算をするデータの化学成分は 一定の序列をしていること。

すなわち：SiO₂, TiO₂, Al₂O₃, (Cr₂O₃), Fe₂O₃,

FeO, MnO, MgO, (NiO), CaO, Na₂O, K₂O, P₂O₅の各成分が この順序で配列している必要がある。ただし括弧内の成分は同時に含むか または除外するかのどちらかにする。合計値を100%に再計算する必要のある場合には P₂O₅のあとに適切な位置に Total(Sum)の項を設けること。

以上のうちで とくにノルム計算をするデータでは 成分の種類と序列が規定されているので 一般用およびノルム計算用プログラムでは 珪酸塩用の“標準組成”を準備している。使用者は これを選択すると ノルム計算のできる成分の組合せを間違いなく登録することができる。これは 第2表に示すとおり 上に述べた成分のほか P₂O₅のあとに H₂O+, H₂O-, Others および Totalの項が付けられている。

この標準組成を選択した場合でも 上記イに違反しない限り P₂O₅のあとに任意の成分を追加できる。微量成分等も Totalの項の後に付け加えてよい。成分の追加は 登録成分名の訂正・追加をする行程を利用して行く。

水質分析用プログラムでも“標準組成”を用意している。これについては省略する。

b) 分子量の算出

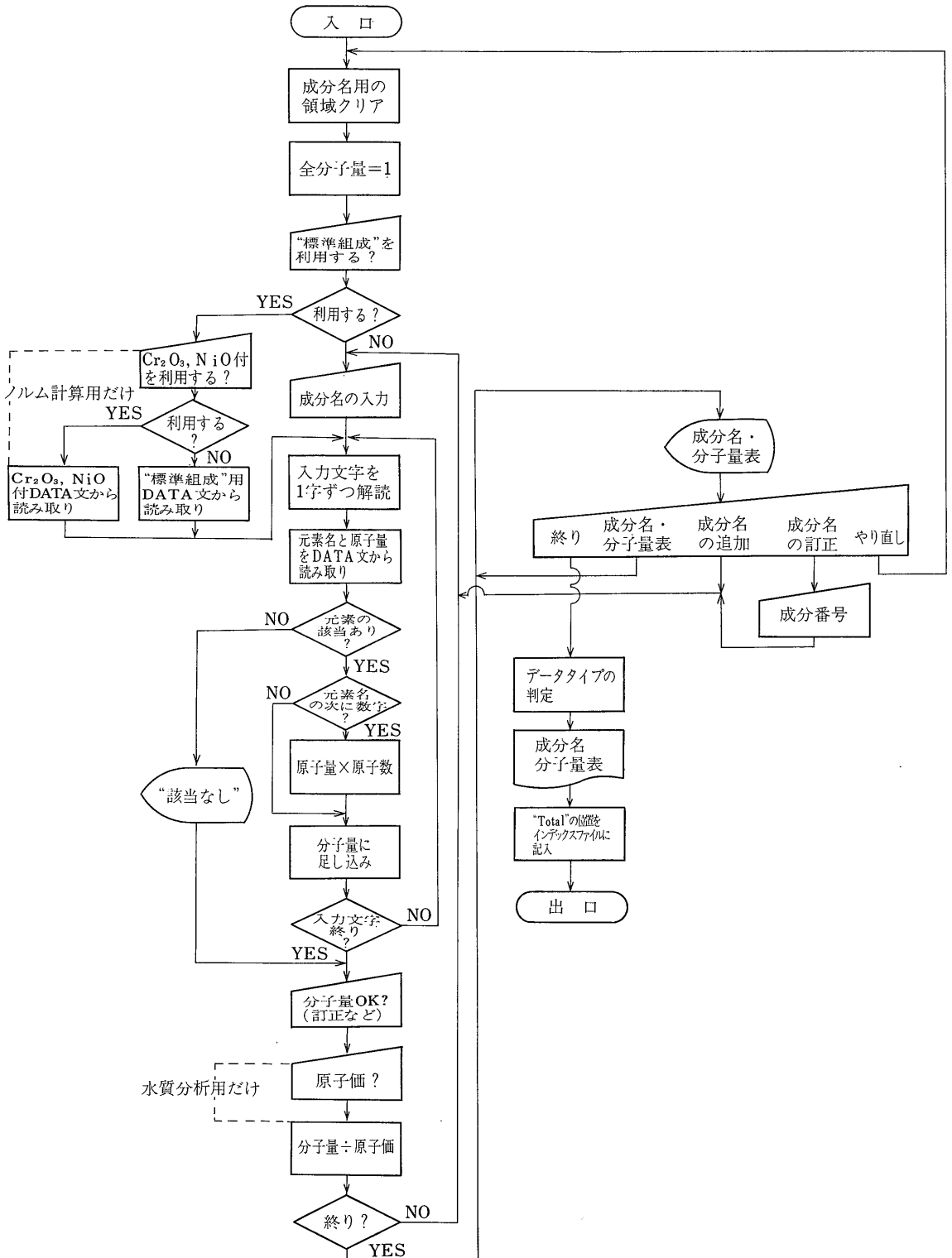
化学成分名が入力されると その文字を解読して 入力成分の分子量が算出される仕組みになっている。これについて述べてみよう。

使用者がキーボードから入力した文字は DATA文中にある元素名との照合が行われる。まず特殊な項目すなわち T, Fe₂O₃, T, FeO, Others, Total, Sumとの照合が行われる。該当があれば これらに併記されている分子量も読み取る。Others, Total, Sumの“分子量”は1とされる。

これらに該当がない場合は 入力成分名の第1字を判別しこれがローマ字の大文字であるときには ただちに DATA文中の1文字から成る元素名を検索する。該当があれば その元素に併記されている原子量を読み取る。該当がなく 第2字目がローマ字の小文字ならば DATA文中の2文字から成る元素名を検索する。

以上で元素名に該当があれば 次の1文字を判読する。もしこれが数字のときには すでに読み取った原子量にこの数値を乗じる。以下 上に述べた手続きを繰り返す。このようにすると たとえば Al₂O₃のような場合は

$$Al \text{の原子量} \times 2 + O \text{の原子量} \times 3$$



第2図 成分名登録の行程

第1図の中にある“成分名の登録”行程（サブルーチン）の詳細を示す。使用者が入力した文字列を解説し、元素名と原子量を読み取ったDATA文と照合しながら、化学成分の分子量を算出する機能が特徴的である。

という計算がなされて 分子量が自動的に求められる。

元素名が DATA 文中に見出されない場合 または第 1 字目が 数字かローマ字の小文字であるときは 化学成分名に該当がない旨の表示が出され 使用者に分子量の入力を仰ぐ。このとき 化学成分名以外の項目のときには “分子量” を 1 とするように規約する。 化学分析値をモル数に換算するには この分子量で割るのだから分子量の値が 0 であってはならない。 プログラムとしては 使用者がたとえ分子量に 0 を入力しても それを 1 に変えるように仕組んである。

このようにして 得られた分子量を成分名とともに成分名ファイル中に記憶する。

水質分析用のプログラムでは さらに 各成分の原子価を使用者に入力してもらう。 SiO_2 のように電荷のバランスが取れて 原子価が 0 となる場合や 化学成分以外の項目の “原子価” は 1 と規約する。 この場合も 0 が入力された場合には 強制的に 1 に変換する仕組にしてある。 各成分の分子量÷原子価の値が成分名ファイル中に取められる。

成分名登録の手間を省くために すでに作られている成分名ファイルの内容を利用することもできる。 この手続をするには 使用者は 既成の成分名ファイルの利用する行程を選び 成分名ファイルの付属しているインデックスファイル名を入力する。 これによって成分名ファイルの内容がテープから計算機に入力される。

一般には 研究テーマが同じとき 取扱うデータの成分の数や組合せは一定する場合が多いので 上に述べた方法は 手続きの簡素化という点でも便利である。

成分名の登録作業が終了し 使用者が終了の指示をすると プログラムでは ただちに成分名の序列を検査し データタイプを決定する。 目下のところつぎの 4 種がある。

データタイプ： NR ノルム計算可能のデータ
NC 同 Cr_2O_3 , NiO 成分付き
CG 一般用データ
CW 水質分析用データ

各タイプの特徴を簡単に述べておこう。

NR および NC

- ア 成分の組合せおよび序列が第 2 表の区間 A の条件を満たす。
- イ 合計値を 100% に再計算するのは同表区間 A の範囲の成分に限る。
- ウ データ処理の際に ノルム成分の指定ができる。
- エ ノルム計算サブルーチンで ノルム値の算出可。

(ただし ノルム成分の指定があった場合またはデータ印刷用プログラムの場合に限る)。

CG

- ア Total (Sum) の項がある場合 その直前の項目までについて合計値を 100% に再計算できる。
- イ ノルム成分の指定をすると “成分名該当なし” を表示。
- ウ ノルム計算サブルーチンへ入れず 入口から出口に短絡。

CW

- ア 水質分析用プログラムでファイルが作成されたとき付けられる。
- イ “分子量” 用ファイルに 分子量÷原子価の値を記憶。
- ウ その他の機能は CG のイ, ウに同じ。

データタイプの決定のあと Total (Sum) の項目の有無 およびそれがあつた場合 その位置 (成分番号) が検索・記憶される。 これらの事項・データタイプ・成分数などは インデックスファイルに記入される。

以上で 成分名登録 (サブルーチン) から抜けて つぎの行程へ進む。

成分名登録の行程の流れ図を第 2 図に示す。

e) 最大試料数の決定

成分名の登録が終ると 成分数が決まる。 この成分数をもとにして 最大試料数の算出が行われる。 この方法として YHP-9845T の BASIC 中には REDIM 文というのがあり 最初に宣言された配列規模 (=行数×列数) の範囲内で 配列の行数および列数を任意に変更できる (吉井, 1981)。 目下 化学分析値を入れる配列の宣言時の規模は 1500行15列である。 そこで成分名登録の結果 成分数が確定することにより 許容試料数がつぎの算式で求められる。

$$\text{許容試料数} = (22500 \div \text{成分数}) \text{の商}$$

ただし 許容試料数は 1500 以下とする。 そして 使用者が入力を希望する試料数と 許容試料数のうちの少ない方の値を最大試料数と呼ぶことにして この数だけの試料を収容できるように配列の修正が行われる。

d) 各ファイルの作成

使用者が 比較的少数の試料についての “使い捨て” 的なデータ処理を選択する場合には テープファイルの作成は省略される。 しかし一般には このシステムの

第3表 データ用ファイルの構成

インデックスファイル インデックス・成分名・データの各ファイル名 ファイル作成・データレコード各年月日 データ名・データタイプ 最大試料数・試料現在数 成分数・合計計算に係る成分数
成分名ファイル 成分名・分子量(水質分析用では分子量÷原子価) 最大35成分
コードファイル("ECD") コード群の名称 コードの組 最大36群
データファイル 試料番号・コード・各成分の値 最大1500試料(15成分のとき)

趣旨に従って つぎのような各種ファイルの作成 (CREATE) が行われる。データ用ファイルの構成を第3表に示す。

- ア インデックスファイル
- イ 成分名ファイル
- ウ コードファイル
- エ データファイル

使用者は インデックスファイルに命名する。これにより 成分名ファイルとデータファイルには自動的に名称が付けられる。このシステムでは インデックスファイルにたとえば“DATA”の名を与えると 成分名ファイルには“DATA a” データファイルには“DATA 1”が付けられる。コードファイルは つねに“ECD”および“ECD 1”という名が付けられる。なお コードファイルは データ処理の際にコード群を記憶させておくもので 詳しくは後の回で述べよう。

インデックスファイル名が与えられるとともに 同ファイルの作成と その内容のレコードがされる。続いて 成分名ファイルの作成と成分名・分子量等のレコードが行われる。そのあとコード用ファイルとデータファイルが作られる。

これら一連の動作のあと 使用者の選択により マスターテープの作成が行われる。これで作業は終了し プログラムは一旦停止 (PAUSE) する。プログラムを進めるキー (CONT) を押すと データ入力の行程などを含むプログラムセグメントを連結し その行程へ入る。

3. ファイル作成中のエラーと その対応

今回述べているのは GEOCAPS の入門行程ともいう

べき部分で このシステムにまだ慣れていない人が いわば おっかなびっくり操作を行うという場合も少なくない。このようなときに 操作上その他の都合でエラーが発生すると 使用者は戸惑うし 計算機そのものへの拒否反応を起こすこともあり得る。しかし テープにファイルを作る行程では 経験的にも エラーの発生率が比較的高い。

この原因の主要なものは 誤って同名のファイルを作ってしまうか または 試料数が多過ぎてテープに入りきらない場合である。とくに後者は 同一のテープに多数のファイルを作るときに起りやすい。新品のテープに対して INITIALIZE の作業をしなければ使えないのに その事を知らないで エラーを起こす人もいる。

このような状況にプログラムとしては一定の対応ができるようであればならない。

YHP-9845T では エラー発生時の処理ができるように ON ERROR GOTO というコマンドがある。これを利用して エラー発生とともに その処理をするプログラム行程へ移る仕組にしている。第2図にその行程を示してある。

a) INITIALIZE されていないテープ使用の場合

テープが新品で INITIALIZE されていないにもかかわらず そのテープにファイルを作ろうとした場合は まず使用者にその状況を表示し 確認を取る。使用者が INITIALIZE の許可を出せば これを実行し そのあとただちにインデックスファイル以下の作成に移る。

b) 上記以外の場合

ファイル名のカタログを表示し 使用者の注意を喚起する。そして 作りかけたファイルを消去 (PURGE) するかどうかの指示を仰ぐ。指定されたすべてのファイルを消去した上で 行程はインデックスファイル名の命名へ戻る。ファイルを作りかけている途中でエラーが発生した場合は その時点までに作られたファイルのすべてを取り消さないと 再びエラーが出る。

ただし 成分名ファイルとコードファイルについては 同一テープ内に同名のファイルが作られそうになった場合は 自動的にその行程を飛ばして 次へ移る仕組になっている。

引用文献

吉井守正 (1980) 文字列を使ったコードによるデータの選択。地質ニュース no.315, p.13-17.
 —— (1981) REDIM 文を使った配列操作。地質ニュース no.318, p.34-37.