

## 北海道植物データ処理システムの開発について (その2)

日野 彰

(前報\* より続く)

### ④植物分類データファイルのフォーマットの検討

植物名については和名をカタカナでそのままデータ化する方法も考えられるが、現実的にはカタカナ(あるいはひらがな)での入力是非常に繁雑であり、それに要するバイト数も大きいため非現実的である。植物目録の作成等の利用を勘案して、植物名に対しては植物分類学上の分類体系に基づくコード化が妥当と判断し、これを植物番号と呼ぶこととした。

植物分類データファイルの第一の役割は、植物番号と和名あるいは学名との対応をあきらかにすることであり、植物分類データファイルによって、数字でデータ化されている植物名を和名や学名に置き換えられるようにする必要がある。

植物分類データファイルの第二の役割は、例えばオオヨモギとエゾヨモギあるいはオオブキとアキタブキのように同一の分類単位に対して複数の和名(あるいは学名)が使用されている場合に、同じ分類単位であることを予め整理するためのデータファイルとしての役割である。群落組成表や植物目録の作成などの際に、同一の植物を正しく判定する必要があるため、ひとつの分類単位に対して一組の代表する和名(正和名)と学名(正学名)を設定し、それ以外の和名・学名(異和名・異学名)については、植物分類データファイルを使用して正和名・正学名に変換できるシステムが妥当であると判断した。この変換プロセスを植物名の標準化と呼ぶこととする。

以上のほか、データ処理上の利便性を考慮して、分類上の位置に関するデータ、分布に関するデータ、出現頻度に関するデータ、形態に関するデータ、貴重性に関するデータを植物分類データファイルのアイテムとして採用した。

以上の検討の結果、11種類のアイテムを登録すべき情報とし、各アイテムについてデータ形式を検討・決定し、カテゴリーの数を勘案して必要バイト数を決めた。選定・決定した各アイテムの内容は次のとおりである。

#### a. 植物番号

植物名を識別するためのコードである。フィールドデータファイルにおける出現植物名と同じく、データ形式は数値(実数)型、バイト数は6とする。

#### b. 正名・異名の区分

当該の植物番号に対応する和名および学名が正名であるか異名であるかを示すコードである。右表の区分に従いコード化する。データ形式は数値(実数)型、バイト数は1とする。

コード	区 分
1	採用する正和名・正学名
2	採用しない正和名・正学名
5	正和名・異学名
6	異和名・正学名
7	異和名・異学名

#### c. 正名の植物番号

正名・異名の区分のコードが2、5、6、7の場合に、植物名の標準化の際に変換先となる正和名・正学名の植物番号である。データ形式は数値(実数)型、バイト数は6とする。

#### d. 自生分布範囲あるいは帰化・栽培・植栽等の区分

当該の植物に関する既知の分布情報をコード化したものである。次頁右上表の区分に従いコード化する

\* 北海道植物データ処理システムの開発について(その1)、善多尼訶9:22-28(1993.4)

る。データ形式は数値（実数）型、バイト数は1とする。

e. 科番号

当該の植物が所属する科を識別するためのコードである。データ形式は数値（実数）型、バイト数は3とする。

f. 属番号

当該の植物が所属する属を識別するためのコードである。データ形式は数値（実数）型、バイト数は3とする。

g. 出現頻度の区分

当該の植物のフィールドデータファイルにおける出現頻度を大まかにランクづけしたコードである。フィールドデータをコーディングする時に使用する「和名索引」の編集の際に利用する情報である。右中表の区分に従いコード化する。データ形式は数値（実数）型、バイト数は1とする。

h. 形態の区分

当該の植物の形態を識別するためのコードである。右下表の区分に従いコード化する。データ形式は数値（実数）型、バイト数は1とする。

i. 和名

当該の植物の和名を示す。データ形式は文字型、バイト数は20とする。

j. 学名

当該の植物の和名を示す。データ形式は文字型、バイト数は60とする。

k. 貴重性

当該の植物の貴重性について、既存の資料からまとめた情報である。当面は、環境庁(1986)における貴重植物と、我が国における保護上重要な植物種及び群落に関する研究委員会(1989)にとりあげられている絶滅種等に関して下表の区分に従いコード化する。データ形式は数値（実数）型、バイト数は3とする。

コード	区 分
1	国内～国外に分布
2	日本特産
3	北海道特産
5	北海道移入種
6	帰化植物
7	栽培植物
8	植栽種
9	園芸種

コード	区 分
1	非常に多い
3	普通
5	やや少い
6	まれ
7	ごくまれ

コード	区 分
1	地衣・コケ
2	草 本
3	つる状草本
5	サ サ
6	落葉木本
7	つる状落葉木本
8	常緑木本
9	つる状常緑木本

コード	区 分
1	「自然環境保全基礎調査報告書」における貴重植物
4 0	「我が国における保護上重要な植物種の現状」における絶滅種
5 0	「我が国における保護上重要な植物種の現状」における絶滅危惧種
6 0	「我が国における保護上重要な植物種の現状」における危急種
7 0	「我が国における保護上重要な植物種の現状」における現状不明種
8 0 0	(北海道における分布がやや疑わしい植物)
9 0 0	(北海道における分布がかなり疑わしい植物)

以上のアイテムの内容をもとに、図3に示す植物分類データファイルのフォーマット（入力用）を設計した。

■分類・和名データファイル

植物	科	属	種	和	名	頁
植物	科	属	種	和	名	頁
植物	科	属	種	和	名	頁

■学名データファイル

植物	科	学	名
植物	科	学	名
植物	科	学	名

図3 植物分類データファイルのフォーマット (入力用)

⑤文献データファイルのフォーマットの検討

文献データファイルに登録すべきアイテムとしては調査年および文献番号のほか文献名、資料名、著者名 (調査者名)、発行年月日、発行者名がある。これらのうち識別コードに含まれる調査年代と文献番号を除くアイテムはいずれも文字型のデータ形式での登録が妥当と判断されることから、文字型のデータ全体を一括して文献内容という1アイテムにまとめることとした。したがって文献データファイルのアイテム数は3となる。各アイテムの内容は次のとおりである。

a. 調査年

調査年は文献・資料を識別するために必要な識別コードの一部である。フィールドデータファイルにおける調査年と同じ基準により、データ形式は数値 (実数) 型、バイト数は4とする。

b. 文献番号

文献番号もまた文献・資料を識別するために必要な識別コードの一部である。フィールドデータファイルにおける文献番号と同じ基準により、データ形式は数値 (実数) 型、バイト数は3とする。

c. 文献内容

文献内容についてはすべてコメント欄として取扱う。ただし一応の目安として文献名記入欄、図書名・発行所名および発行年月日の記入欄、著者名の記入欄の3つの記入欄を想定し、それぞれに65バイトづつを用意した。データ形式は文字型、バイト数は195となる。

以上のアイテムの内容をもとに、図4に示す文献データファイルのフォーマット (入力用) を設計した。

年	REF	文献内容①
年	REF	文献内容②
年	REF	文献内容③

図4 文献データファイルのフォーマット (入力用)

⑥処理条件設定データファイルのフォーマットの検討

処理条件設定データファイルに登録すべきアイテムの内容は、どのようなデータ処理を行おうとするのかを決めた後に検討すべきものであることから、データ処理システムの作成の項で検討される。ここでは検討の結果登録すべきと判断されたアイテムの名称のみを以下に掲げる。

	略記号
a. 計算処理項目の指定	JST
b. フィールドデータの並びの指定	JSU
c. 植物名の標準化の方法の指定	JSV
d. 識別コードによる検索条件の指定	MYR, MRF, MST
e. 出力する帳票の指定	KKZ, K KU, K KY, K KS, K KL
f. 類似度計算の対象層の指定	KT1, KT2, KSS, KH1, KH2, KHL, KMM
g. 標題の指定	SHY, SJH, SLH
h. 群落の区分種の指定	GK, JG
i. 標題部の出力内容の指定	LYR, LRF, LOC, LPA
j. 分布図作成植物の指定	NPL
k. 植物目録作成対象地図番号の指定	MPN
l. 植物目録の出力形式の指定	MPM
m. 計算処理対象年代の指定	MYMIN, MYMAX

### (3) コード化の基準の作成

データをデータファイルとして入力するためには、まず資料となるデータを一定の基準によっていくつかの形式に分類し、形式ごとに入力方法の基準を明確にしておくとともに、あらかじめ各アイテム中のコードの意味するところを整理して各形式のデータに対応したコード化の基準を作成しておく必要がある。

コード化すべきアイテムのうち、文献番号、調査番号、調査方法、植物名の表示方法の区分、発表の形態の区分、所属階層のそれぞれについてはすでに行ったフィールドデータファイルのフォーマットの検討の過程で実質的にコード化が終わっている。したがってここでは調査地点位置と植物名に対するコード化の基準について検討する。

#### ① 調査地点位置のコード化の基準

調査地点位置を示すコード（以下地図番号という）はすでに述べたように、当該する調査スタンドの調査地点の所属する国土地理院発行1:25,000地形図を示す4桁の数字であり、地形図1枚の範囲は概ね10km×10kmである。E C P L A N Tにおいて1:25,000の1地形図をメッシュの単位とした理由は以下のとおりである。

メッシュを利用した植物分布のマッピングについては金井(1972)によって詳しく研究されている。金井(1972)はメッシュシステムの数値的命名法として、1:50,000地形図の16分の1までを表現できるようにされた Locality Index を提案している。その後金井(1983)は、Locality Index の大きさが県単位の分布図に用いるには粗い場合があるとして、経緯度を分の単位で表示する座標系を用いて茨城県と富山県におけるいくつかの植物の分布図の作成を試みている。Locality Index を用いて長野県における植物分布図の作成を行った金井(1988)の例によると、県全体のメッシュ数は574個となっている。これらの例をみると、特定の地域における植物の分布状況を把握するには一つの図面に1,000個前後のメッシュがあれば一応足りると判断され、北海道においてその数に対応するメッシュの大きさはちょうど1:25,000地形図となる。また、E C P L A N Tでは植物群落組成調査票を基準としてデータファイルが設計されている。図6に示されるとおり植物群落組成調査票では一般に1:50,000地形図を上下左右に分けて調査地点の位置を表現するようになっており、これがちょうど1:25,000地形図の図幅に対応する大きさのメッシュとなっている。過去の多くの資料をできるだけ同じ基準でデータ化しようと考えた場合、植物群落組成調査票の位置表示をそのまま利用できることは1:25,000地形図をメッシュとすることの大きな利点



②植物名のコード化の基準

植物名のコード化は本データベース作りの中で最も労力を要した作業のひとつであり、結果的に北海道に生育が報告されている高等植物のリストを作成することとなった。

北海道内に自生するといわれる高等植物（シダ植物および顕花植物）の種類数は約 1,800～2,000といわれており、数の上ではコードとして4桁の数字があれば足りることとなるが、実際にはひとつの植物種に対して数多くの異名（異和名および異学名）があり、これらを原典にできるだけ忠実にデータ化するためにはさらに2桁の数字が必要とされた。さらに、植物分類学上の分類体系自身時代や研究者によってかなり異なるという問題や新産種への対応という問題があり、それらに対してできるだけフレキシブルに対応できるコード体系を作り上げるため、植物分類データファイルの中に正異名の判定や所属する科や属の判定などに必要なデータを含めることとした。

植物番号についての詳しい解説は「北海道高等植物目録」に記載されているが、その概要は以下のとおりである。

- ・植物番号は4桁の種番号と2桁の枝番号からなる。
- ・種番号は原則として種、亜種および変種を区別するものである。
- ・枝番号は異名・品種・植栽種・栽培種等を区別するものである。
- ・種番号2000～2999はシダ植物、3000～3999は裸子植物、4000～5999は単子葉類、6000～7999は離弁花類、8000～9998は合弁花類にあてられている。

(4) コーディング

フィールドデータのコーディングは図2（前報参照）に示したECPLANTフィールドデータ用コーディングシートに記入して行う。コーディングをスムーズに行うため、「ECPLANTコーディングマニュアル」と「ECPLANT和名索引」を作成した。

植物群落組成調査票とそのコーディングの例を図6に示す。

(5) データ入力

コーディングされた情報をファイルとして入力・保存し、フィールドデータファイルとした。

図6のコーディングシートの入力例を図7に示す。

(次報へ続く)

表4 市町村コードおよび支庁・離島・山系コード（部分）

コード	市 町 村	コード	支 庁	離 島	山 系
8 0 0 1	松前町	8 0 0 0	渡島		
8 0 0 2	福島町	8 0 4 1		大島	
8 0 0 3	知内町	8 0 4 2		小島	
8 0 0 4	木古内町	8 0 4 6			千軒山系
8 0 0 5	上磯町	8 0 5 0	檜山		
8 0 0 6	大野町	8 0 9 1		奥尻島	
8 0 0 7	七飯町	8 0 9 6			遊楽部山系
8 0 0 8	函館市	8 0 9 7			狩場山系
8 0 0 9	戸井町	8 1 0 0	胆振		
8 0 1 0	恵山町	8 1 4 6			樽前山系
8 0 1 0	(尻岸内町)	8 1 5 0	日高		
8 0 1 1	綴法華村	8 1 9 6			日高山系
8 0 1 2	南茅部町	8 2 0 0	後志		
8 0 1 3	鹿部町	8 2 4 6			余市山系
8 0 1 4	砂原町	8 2 5 0	石狩		
8 0 1 5	森町	8 3 0 0	空知		
8 0 1 6	八雲町	8 3 4 6			夕張山系



```

***** ***** TOP OF DATA *****
000001 202 191 718カミトフ`ウ,アショロ N45E15      3504145シラカンハ`-ミヤコサ`サCOM.
000002 12.0 70 8.0 30 4.0 40      1.20100
000003 19 100HINOMA,SHIDA  QUAJPN      NP
000004 T144614119      6971      117611  T222788317  227352  S117352  11756620
000005      22696504  11840360  11618803H2444702  442180  116675  117490
000006      6612      6719      118427      9510      8536      5517      6568
000007      +2756620  23505415  618603      9107      9999
000008 202 291 718カミトフ`ウ,アショロ N50E20      4004144シラカンハ`-ミヤコサ`サCOM.
000009 14.0 70 8.0 20 4.0 30      1.20 98
000010 19 100HINOMA,SHIDA
000011 T1446141  T211756620  226186  S226186  11756620  116174  H2444702
000012      222220      226675  127276  237250  119510  9107      6186
000013      7046      7490      572596  9999
000014 202 391 718カミトフ`ウ,アショロS50W 3      4004144ミス`ナラ(L-R)-ミヤコサ`サCOM.
000015      6.0 20 4.0 60      1.20 98
000016 19 25HINOMA,SHIDA
000017 T2226174  S337352      226016  117516  117870  116719  H2554702
000018      222220      116674      7197  227516  119414  6016  226719
000019      6472      119107  224262      7215      9999
000020 202 491 718カミトフ`ウ,アショロS35E 5      3604144ミス`ナラ(L-R)-チシマサ`サ・ミヤコサ`サCOM.
000021      7.0 20 4.0 30      1.30100

```

図7 フィールドデータの入力例