富山市科学文化センター研究報告第14号，pp.79—86(1991)

植物標本データベース構造の一提案*

太田　道人
富山市科学文化センター

A Proposal of a Filing System of Herbarium Specimen Information

Michihito Ohta
Toyama Science Museum

A filing system of herbarium specimen information is proposed and discussed to build up effective computer retrievable database.

はじめに

コンピュータを使って標本データを登録し、ラベルや目録出力などを行う試みが各地で行われるようになってきた（黒崎，私信；狩山，1990；小川，私信など）。この背景には、パーソナルコンピュータの普及があると考えられるが、実際の業務においては、処理速度やディスク容量などのデータ処理の基本的な面で大きな制約を受けていることが多い。また、汎用コンピュータを使って処理を行った例もあるが、この場合は、随時要求される多種の細かなプログラムの改良に労力がかかりすぎるという問題が指摘されている（太田，1987；小川，私信）。

パーソナルコンピュータの性能は数万件のデータを処理するには、決して余裕あるものではないが、大規模な汎用コンピュータやエンジニアリングワークステーションと比べ、きわめて容易に導入でき、かつ柔軟な処理が可能であるというメリットがある。したがい、パソコンを利用して、より性能のよいアプリケーションやデータベース構造を摸索し、その性能を公表することは、同様の環境で標本データベースを有する機関にとって、システムの改善や検討の基準とすることができ、意義のあることである。

筆者は、富山市科学文化センターにおいてパーソナルコンピュータとリレーショナルデータベースソフトを使用し、リレーショナル構造による標本データベースを構築した結果、資料の受入から分布図出力用データ作成までの一連の処理において、比較的満足のいく性能を得ることに成功したので報告する。

国立科学博物館植物研究部金井弘夫博士には、地名索引購入時に便宜をはかっていただき、原稿に目を通していただいた。頌栄短期大学黒崎史平助教授には、標本データをデータベース化する際に留意すべき点を指摘していただいた。兵庫県立自然系博物館準備室橋本光政先生、同西井正和氏、倉敷市立自然史博物館狩山俊悟学芸員、徳島県立博物館小川誠学芸員には、データベースソフトウェアについての検討結果を示していただいた。これらの方々に深く感謝申し上げます。

この研究の一部に、文部省科学研究費補助金奨励研究（B）第63917012号が利用された。

使用したハードウェアとソフトウェア

ハードウェア

*富山市科学文化センター研究業績第102号

パーソナルコンピュータ
　NEC PC-H98model70
ハードディスク装置
　ICM HD7300ES（300MB）
ソフトウェア
リレーショナルデータベース
　ビーコンシステム R:BASE PRO ver.2.2
BASIC言語
　Microsoft Quick Basic ver.4.5
メモリー管理ソフト
　Megasoft Memory Pro 386 ver.1.5
オペレーションシステム
　Microsoft MS-DOS ver.3.30B

PC-H98mode170は、MS-DOSパソコンの中では、かなり処理速度が高速なものである。これに、メモリーを6メガバイト増設した。HD7300ES は内部に1メガバイトのキャッシュメモリーを持つ高速型のハードディスクである。この一領域128メガバイトをデータベース用に使用した。

R：BASE 選択の理由は、リレーショナル処理が充実していること、コマンドによりかなり柔軟なデータ処理が可能で、さらにこれが自動化できること、入力様式や出力様式の設計が容易であること、ソートが非常に高速であること、可変長項目が設定できそこからの検索が可能であること、各種データベースソフトとの互換性に富むことなどである。

Quick Basicは、R：BASE から出力されたデータを目録の体裁に整えるために利用した。

Memory Pro 386は、PC-H98model70のメモリーを有効に活用するためのソフトで、大きなデータベースからの検索を高速化するために利用した。

データベースが持つ機能

1 標本データの管理（登録、修正、検索、ソートなど）
2 標本ラベル出力
　(i)　標本受入時の一括出力
　(ii) 同定変更・学名変更時等の個別出力
3 分布図作図で必要な位置情報の自動付加
4 リスト出力（収蔵目録用・環境調査用）

1の処理はデータベースが備えるべき基本的な機能である。通常サポートしにくい備考データについても、最長4092バイトまでつけることができ、以下の全ての処理に出力される。

2のラベル出力は、同一産地データのラベル書き作業を軽減する目的で作成されたものであるが、ここではさらに機能を付加し、受け入れ時に標本の和名リストが与えられた場合には、学名付きの標本ラベル出力を行い、同時にデータ登録が完了するものとした。

3の処理は、これまで分布図自動作図に先立ち必要だった経緯度情報やメッシュコードなどの位置情報の標本データへの付加作業を大幅に軽減する目的で作られたものである。地名辞書には、石川・富山県地名索引（金井, 1987 b）を用いた。

4は、定期的に発生する収蔵目録のデータ編集作業や、地域ごとの生物相リストを証拠標本に基づいて短時間に作成するための機能である。最終的なレイアウトは、Quick Basic で作られたプログラムで行われる。

構　造

狩山（1990）は、パーソナルコンピュータを使った標本データベースにおいて、「標本管理マスター」の1レコードに、標本データをはじめ、学名や科名番号、備考等全てのデータを持たせる構造とした。この場合、複数の同一植物名に対し、その数に対応した学名と科名番号等が繰り返し記録されることになり、データベースをいたずらに大きくし、その大きさが検索速度の低下をもたらす。

テーブル名「標本データ」

1	**登録番号**	TEXT	7 文字	yes
2	**和名**	TEXT	16 文字	yes
3	県名	TEXT	6 文字	
4	**市町村名**	TEXT	10 文字	
5	**産地名**	TEXT	24 文字	
6	産地名補	NOTE		
7	標高1	INTEGER		
8	標高2	INTEGER		
9	採集日	TEXT	8 文字	
10	採集者名	TEXT	12 文字	
11	受入番号	TEXT	5 文字	

テーブル名「備考」

1	**登録番号**	TEXT	7 文字	yes
2	コメント	NOTE		

テーブル名「受入票」

1	**受入番号**	TEXT	5 文字	yes
2	受入日	TEXT	6 文字	
3	受入方法	TEXT	2 文字	
4	タイトル	NOTE		
5	点数	INTEGER		
6	採集地	NOTE		
7	収集日	TEXT	6 文字	
8	採集者名	TEXT	12 文字	

テーブル名「おぼえ」

1	**受入番号**	TEXT	5 文字	
2	メモ内容	NOTE		

テーブル名「学名辞書」

1	**科名**	TEXT	10 文字	yes
2	**和名**	TEXT	16 文字	yes
3	属名	TEXT	17 文字	yes
4	種名	TEXT	42 文字	
5	変品種名	TEXT	72 文字	
6	出典	NOTE		

テーブル名「科名ｺｰﾄﾞ」

1	配列順	INTEGER		
2	**科名**	TEXT	10 文字	yes
3	科名ﾗﾃﾝ	TEXT	17 文字	
4	分類ｺｰﾄﾞ	TEXT	1 文字	

テーブル名「地名辞書」

1	県名	TEXT	6 文字	yes
2	**市町村名**	TEXT	10 文字	yes
3	**産地名**	TEXT	24 文字	yes
4	地名読み	TEXT	20 文字	
5	2.5万地図	TEXT	8 文字	
6	東経	TEXT	5 文字	
7	北緯	TEXT	4 文字	
8	読み確認	TEXT	1 文字	
9	広域地名	TEXT	1 文字	

テーブル名：「富山県植物誌」

1	科名	TEXT	10 文字	yes
2	**和名**	TEXT	16 文字	yes
3	解説文	NOTE		
4	産地一覧	NOTE		

Fig. 1. 個々のテーブルの内容とテーブル間の関連を示す。テーブル内は、カラム名、データタイプ、長さ、索引指定の有無の順に記した。データタイプとはカラムに入力されるデータの形式。INTEGER は数字データ、TEXT は固定長文字データ、NOTE は自由長文字データを示す。囲み線や網かけ等、同一文字修飾を施したカラム名同志が、テーブル間の共通カラムである。

そこで筆者は、データベースの構造を、繰り返し現れるようなデータを別テーブルにして、必要なときに必要な項目だけを参照して表示するリレーショナル構造とした（Fig. 1）。例えば、「標本データ」テーブルと「学名辞書」テーブルとは、共通のカラム「和名」をたよりにたがいに参照することが可能であるということである。構造上は学名と標本データとが分割されているが、表示に際しては、あたかもつながっているかのように見えるのである。これにより、「標本データ」テーブルに学名や科名番号、備考等のデータを持たせる必要がなくなり、ディスク上のファイルの大きさを小さくし、ひいてはテーブル単位の検索速度を上げることが可能になる。データ件数が増加すればするほどこの差は大きなものになる。ただし、テーブルを関連づける処理をあらかじめ定義しておいたり、複雑な処理結果を得るためには、それなりの手順を踏む必要がある。

データベースに含まれる内容とデータ量

「　」内はテーブル名、“　”内はカラム名。

(1)「標本データ」　33846件

種子植物とシダ植物の標本データが記録されている、いわばメインテーブル。他のテーブルとの共通カラムを多く持っている。産地の区切り方は「地名辞書」に合わせ、富山県/立山町/芦峅寺/立山少年自然の家/のように区切って入力される。

カラム名の“和名”は必ずしも植物和名でなくともよく、同定が不完全なものについては、Poa sp.のように、和名のないものについても、学名を入るところまで記入する。

(2)「備考」　2563件

標本データの備考部分。カラム名“登録番号”を介して、“標本データ”と関連づけられている。同定記録、同定変更記録、採集地の具体的な記述、産地名表記に関する記述などを最大4092バイトまで記録する。

(3)「地名辞書」　5254件

富山県地名索引（金井，1987ｂ）のデータ。“市町村名”と“産地名”２つカラムを介して「標本データ」と関連づけられている。両方が一致した場合に限り、テーブル間の結合が行われる。この条件であれば、富山県地名索引で、同名異所の問題は生じない。また、同じ地点を示す地名が、地図が異なるために索引上に複数ある場合は、地図をチェックし、地名が最も広くかかると判断される地図の座標のみを採用した。ただし、線状をなす河川名などは、そのまま残した。元のデータに若干のデータを追加しエラーデータを修正した。

(4)「学名辞書」　5352件

“和名”を介して、「標本データ」と関連づけられている。

現在のところ、種子植物については、新日本植物誌（大井，1983）から抜粋したものを、シダ植物については、原色日本羊歯植物図鑑（田川，1959）のものを使用。これ以外の文献から採ったものは，“出典”に記入した。

科名と配列順に関するコードは別テーブル「科名コード」に分割してある。

「標本データ」で辞書にない和名が使われ、その名前によって分類配列を必要とする場合には、随時辞書登録を行っていく。

(5)「科名コード」　209件

“科名”を介して「学名辞書」と関連づけられている。

(6)「受入票」　1045件

“受入番号”を介して「標本データ」と関連づけられている。

資料受入れ時の情報。標本など受入単位ごとの詳しい情報や、数量、処理記録などを記録する。

(7)「富山県植物誌」　2505件

富山県植物誌（大田他，1983）のリスト部分のデータが入っている。

“和名”を介して「標本データ」や「学名辞書」と関連づけられているので、必要に応じ“解説文”や“産地一覧”を引用することが可能になっている。

(8)「おぼえ」　3件

データベース全体に関するメモ。

データベースのファイルの大きさ

画面レイアウト用ファイル　0.4MB
データファイル　7.2MB
索引情報ファイル　1.2MB
以上3ファイル。

性能（処理時間）

①単純検索（1）

(i) 索引にヒットする場合

（テーブル「標本データ」から和名「イノデ」を検索し該当データが表示されるまでの所用時間）

1件目表示　瞬時
20件表示　1秒以内

(ii) 索引にヒットしない場合

（「イノデ」を和名に含むものを検索）

1件目表示　17秒以内
20件表示　17.5秒以内

②単純検索（2）

（属名「Pteris」に属する標本データを表示する。あらかじめ「標本データ」と「学名辞書」とをビュー指定により関連づけておく。）

1件目表示　瞬時
20件表示　1秒以内

③学名付加

（「標本データ」から「学名辞書」を参照し、学名を付加した新たな標本データテーブルを得るまでの所用時間）

20件　0.5秒
100件　2.8秒
500件　11.5秒
1000件　23.5秒

④経緯度情報の付加

（「標本データ」から「地名辞書」を参照し、経緯度情報を付加した新たな標本データテーブルを得るまでの所用時間）

20件　5.5秒
100件　32秒
500件　2分3秒
1000件　4分37秒

⑤目録出力

（ある寄贈品目録（4500データ）を作る際、1．受入番号を指定して標本テーブルから該当データを検索し、2．学名付加処理、3．備考付加処理、4．多重ソート、5．テキストファイルへの出力、6．BASICによる整形までに要する、キー入力時間も含めた所用時間）

約40分

問題点と考察

（1）経緯度情報の自動付加について

標本データに経緯度情報が自動的に付加されることは、分布図の自動作図と相まって植物地理学的研究の進展に貢献することが期待される。ただし、次の制限がある。

一つはメッシュ特定の問題である。本処理では機械的に、市町村単位に地図に記された地名の位置のメッシュを特定する。しかし実際には、そのメッシュには、地形上・環境上、当該植物が産することはきわめて考えにくいことがある。手作業では、ここで経験的判断が入って近接する適正なメッシュに推定の上変更することが可能であるため、自動付加結果とでは若干の差が出ることがあり得る。ただし、分布図の利用目的が、当該植物の県全体における分布傾向を把握することであれば、この差はほとんど問題にならない。

また、河川や用水など線状をなす地名は、索引が複数記載されているため、標本データ

と地名索引とのリレーショナル処理によって、１標本データにつき複数の位置情報を生じてしまうことがある。したがって、このような産地記録を持った標本データについては、あらかじめ処理の対象からはずさなければならない。

経緯度情報の自動付加は地名辞書の存在が前提となるシステムである。現在の所、２万５千分の１地形図から作成された地名索引は、富山、石川、茨城、秋田県分しかデータベース化されておらず、早期に全都道府県分の発行が待たれるところである。

（２）本リレーショナル構造の問題点

今回報告したデータベースには、和名のない標本データの登録に際しては、16バイトの和名欄に学名を入るところまで入力するという不自然な対処を行わなければならないという欠点がある。和名カラムが学名辞書と共通項目になっているため必ず入力されなければならないからである。このため、目録出力などの際には、不自然な“和名”部分を印刷前に削除しなければならない。

また、標本ラベルにもとから経緯度情報が記録されている場合には、無条件にこのデータを位置情報として採用すべきであるが、今回の報告ではこの機能がサポートされていない。現在の所、このようなデータは非常に少ないが、時とともに地名が変化しても正確な位置の再現ができるよう、今後は経緯度情報も合わせた位置記録が不可欠となっていくと考えられる。今後、データベースに位置情報データ記録用のテーブルを新設し、これを経緯度情報自動付加処理に優先して使用するよう改善しなければならない。

（３）パソコンデータベースの限界について

データ件数の増加に伴う検索速度低下は、索引にヒットする場合においてはほとんど問題にならないが、索引にヒットしない場合には、深刻な問題となる。この速度低下は、テーブルのデータ件数にほぼ比例する。現在「標本データ」テーブルをひととおりアクセスするのに要する時間が17秒であることから、単純計算では、1分間で約12万件の標本データの検索が可能ということになる。ただし今回の条件では、増設メモリーの6メガバイトという大きさとデータベースファイルの大きさとがほぼ等しかったために、データベースのほとんどがメモリー上に展開されていたことになる。標本データが10万件になった時点でのファイルの大きさは10数メガバイトと推定され、この速度を維持するためには、メモリー増設が不可欠となってくる。

標本データに経緯度や学名を付加する場合には、データベース増大の影響はほとんどなく、対象とするデータ件数にのみ左右される。経緯度付加処理は複数のカラムを一致させるため、かなりの時間がかかるが、得られる結果の有用性を考慮すると実用範囲内と考えられる。標本ラベル出力については、レーザープリンタで印刷する時間が検索及びデータ処理にかかる時間よりもはるかに長いため、特に問題は生じない。

一方ディスク容量の面からの制約がある。MS-DOSが管理できるディスク容量は現時点では最大128メガバイトである。データベースは、時にそのファイルの大きさの２倍以上の空き領域を必要とするため、余裕と安全を見込んで40メガバイト程度をパソコンデータベースの限界としなければならない。本例のファイルの大きさの４倍程度、標本データ件数にして推定30万件程度が限界である。

（４）その他

データベース構造の決定をはじめ、処理速度、操作性、画面設計の自由度等はアプリケーションソフトの性能に大きく左右される。

今回は、筆者の判断によりR:BASEを選択したが、より目的に合致した高性能のアプリケーションソフトが存在すること、あるいは登場することは充分に考えられる。この問題については、データ処理や分布図作図など共通の目的を持った研究機関相互の情報交換が密に行われ、互いの貴重な実戦データが生かされることでかなり解決されていくべきものと考える。このような交流から、通信やメディア交換等による標本データの相互利用へと発展していくことが望ましい。ただ、ここで常に問題となるのは、標本の同定に関わることである。情報交換により、広範囲に標本の存在が判明することの意義は高いが、分類に関わるものに引用するためには、結局、実物を目で確認しなければどうしようもない。分類学においては、データ相互利用の限界がここまでであるということを示すと同時に、標本調査の重要性が改めて指摘されるところである。

参考文献

狩山俊悟，1989．岡山県におけるツツジ属の分布．倉敷市立自然史博物館研究報告4：1-15.

狩山俊悟，1990．パソコンを利用した標本データの登録と分布図の作図．倉敷市立自然史博物館研究報告5：23-32.

神奈川県植物誌調査会編，1988．神奈川県植物誌．神奈川県立博物館．1442pp.

金井弘夫，1986．長野県における普通植物の分布．国立科学博物館研究報告B．12(4)：155-165.

金井弘夫，1987a．石川・富山県地名索引．199pp.

金井弘夫，1987b．石川・富山県地名索引データベース．日債銀総合システム株式会社.

環境庁，1985．自然環境保全基礎調査「緑の国勢調査」.

大田弘・小路登一・長井真隆，1983．富山県植物誌，弘文堂．富山．430pp.

太田道人，1987．標本のデータベース化について，進野久五郎植物コレクション，富山市科学文化センター収蔵目録1：203.

太田道人，1991．地名索引情報を使った植物分布図の自動作図．富山市科学文化センター研究報告14．87-91.

太田道人，長井真隆・吉沢庄作植物コレクション．富山市科学文化センター収蔵目録4（印刷中）.

大井次三郎，1983．新日本植物誌顕花篇．至文堂．東京．1716pp.

田川基二，1959．原色日本羊歯植物図鑑．保育社．大阪.

参考資料

Herb. Toyama Science Museum

Polystichum • amboversum Kurata

No. [P007361]
アイツヤナシイノデ FNo.[B-90069]

Loc.富山県利賀村 田ノ島
川向
alt. 600m - 志村義雄同定1990.11。P0167
7と同一株。大島・太田1991
にアイツヤナシイノデとして引用。
Date 19790802 Coll.大島哲夫

備考データも出力された標本ラベル

標本データ登録画面

No.(例 S123456):[redacted]　　和名[ｻﾗｼﾞｸﾗ[redacted]]

産地データ	県　　名 [富山　] 市町村名 [大山町　　] 産地名１ [日尾　　] 産地名２ [～牧野　　]
	標高(alt.m)　[160]m～ [-0-]m
	受入番号 [B-83117]
	採集者名 [大島哲夫　　] 採集日　[19820724]

＊＊　学名入力画面　＊＊	
和名(ｶﾀｶﾅ)：[redacted]	科名(ｶﾀｶﾅ)：ｷｸ
属名(latin)：Cirsium	[COMPOSITAE　　]

種小名+命名者：babanum Koidz.

変品種名(var.以下)：var. otayae (Kitam.) Kitam.

特記出典：-0-

地名辞書入力

地名読み ﾍｲｿﾞｳﾀﾞﾆ	県名　市町村名　産地名 富山　立山町　平蔵谷
2.5万 [十字峡　] ;東経13737 ;北緯3636	読み確認(0･1) [0] ; 広域地名 [0]

[資　料　受　入　票　（入力／修正）]	
受入番号(例:88001) [[redacted]]	受入年月日(例:880115) [870430]
受入方法(寄／採／交) [採]	
件　名 [ｲﾁﾘﾝｿｳ ｱｵﾊﾞｽｹﾞ　　] 点数[33] [　　]	
採集年月日(複数の場合は代表値で.例:880115) [870430]	採集者名[太田道人　]
採集地(複数の場合は適宜区切って) 富山県大沢野町神通峡薄波 alt.250m ・ 細入村割山 alt.250m ・ 富山市杉谷 alt.40m	