富山県生物学会誌 第27号, pp. 76～83 (1987)

コンピュータを使った手軽な標本データ整理
――ＭＳ－ＤＯＳを使って――

富山市科学文化センター　太　田　道　人

生物標本は分類学的情報のみならず，生物地理学的並びに生態学的情報も提供してくれる。生物地理学的解析の基本資料となる植物誌や分布図の作成作業は，従来，標本その他に基づいたデータを全て手作業で行っていたが，近年コンピュータが普及し始め，大量の標本データを短時間で処理し，これらの作業を合理的に行うことが可能となってきた。

コンピュータを利用して，標本データをデータベース化し，植物誌や分布図作成を先進的に行っているものには金井（1982），清水（1984）などがある。富山市科学文化センターでも富山市電子計算課の協力で，同様に大型コンピュータを用い，ソフト開発を行って収蔵資料目録の作成を行った（太田，1987）。これらのような大型のデータベースは，いずれも大きな設備費用とプログラム開発に時間がかかるが，そのメリットには，データの検索・分類・修正・追加・再出力などの処理が高速で容易に行えることに加え，データの一括集中管理によって，将来共同利用を可能にすることもあげられる。

標本データベース作成の目的として，1.検索とソート（並べかえ）を行うこと，2.体裁を整えて植物誌などの目録を作成すること，3.分布図を作成すること，があげられる。2と3の処理を行う既製のソフトは，本処理の特殊性や作成様式が研究者によって異なる面が多いため，ほとんど市販されていない。このため納得のいくものを作成するには，どうしても独自のプログラムを作成しなければならない。これに対し，1の処理はいずれもコンピュータの機能として基本的なものであり，特別なプログラミングの必要がない。検索とソートが可能になるだけでも，今まで多大な労力を必要とした，採集地別，採集者別等のリスト作成や任意項目での昇順や降順のソートがきわめて容易に行えるようになるため，フロラやファウナの把握，採集計画立案などに大きく貢献する。本稿では，標本データ整理において1の作業をプログラムを作らずに行うことを所期の目的と考え，その具体的方法を紹介させていただく。

レコード数が数千件程度の小規模なデータ処理には，最近普及し始めたパソコンを用い，市販の簡易言語ソフトを購入して利用する場合とＢＡＳＩＣやＣＯＢＯＬなどのやや複雑な言語によるプログラムを作成して利用する場合とが一般的である。しかし簡易言語は，標本データのように，入力した後は，さしあたり検索とソートしか必要としないという利用目的に比べ，機能が多く高価であり，一方のＢＡＳＩＣ等は，プログラム作成によってかなり複雑な処理が可能であるが，その多機能さゆえにマスターするのが大変で，プログラミングには時間がかかるなどの問題がある。

そこで筆者は，標本データ整理のうち検索とソートをマイクロソフト社のＭＳ－ＤＯＳ ver 2.11

という，16ビットパソコンのオペレーティングシステム（一部機種に標準装備）として使用されているものを使って行っている。この方法は，脇（1986），他によっても紹介されており，操作が簡単で処理速度も比較的速い上，ソフトに費用がかからないという利点がある。

標本データベースの概要

筆者が作成している標本データベースは標本番号，学名，和名，採集地，採集年月日，採集者，標本受入番号の7項目で，現在約2500件である（図1）。学名や和名など，文字の長さが不統一なものについては字数制限を行い，年月日を6桁とするなど，1レコードには必要最小限の文字数しか入力していない。これによって一枚のディスクに，より多くのデータを書き込むことができ，さらに後述するソートにそなえソートを行う可能性のある項目の先頭を揃えることができる。

ＭＳ－ＤＯＳで作成されたデータベースは，市販の簡易言語ソフトにも取り込ませることが可能であり，将来ソフトを購入した場合でも，容易に移行できるという特徴がある。

図1の学名は入力された7文字の和名を基に，別に作成された学名辞書を参照して自動的に挿入されたものであるが，この合成機能は，残念ながらＭＳ－ＤＯＳには装備されていないため，ＢＡＳＩＣのプログラムを作成して行った。この概要については，今回は学名辞書の項で簡単にふれる程度にとどめ，詳しくは別の機会に譲らせていただく。

図1　シダ植物標本データベースから，大島哲夫氏採集標本を検索し，和名順にソートしたものの出力の一部。ファイル名OSM（本文参照）。

```
3330,ASPIDI Polyst  ｱｲｱｽｶｲﾉ,富山県井波町東城寺,821010,大島哲夫,B86-017
1662,ASPIDI Polyst  ｱｲｱｽｶｲﾉ,富山県魚津市湯上,770923,大島哲夫,B85-096
1666,ASPIDI Polyst  ｱｲｱｽｶｲﾉ,富山県高岡市須田,770805,大島哲夫,B85-096
2765,ASPIDI Polyst  ｱｲｱｽｶｲﾉ,富山県滑川市本江,780605,大島哲夫,B86-077
2766,ASPIDI Polystl ｱｲｱｽｶｲﾉ,富山県滑川市本江,780605,大島哲夫,B86-077
2694,ASPIDI Polybl  ｱｲｱｽｶｲﾉ,富山県魚津市黒谷,780716,大島哲夫,B86-077
2688,ASPIDI Polybl  ｱｲｱｽｶｲﾉ,富山県魚津市鹿熊,780903,大島哲夫,B86-077
1627,ASPIDI Dryocea  ｱｲﾉｺｸﾏｸﾞ,富山県滑川市本江,780605,大島哲夫,B85-096
2911,HYMENO Crokino  ｱｵﾎﾗｺﾞｹ,富山県平村祖山,801010,大島哲夫,B86-077
3148,LYCOPO Lycompla  ｱｽﾋｶｽﾞﾗ,長野県上伊那郡人笠山,560603,大島哲夫,B86-068
1557,ASPIDI Cyrenul  ｲｯﾎﾟﾝｸﾗ,富山県朝日町滝淵上流,850824,大島哲夫,B85-096
1548,PTERID Dirsut  ｲﾇｼﾀﾞ  ,富山県八尾町桐谷,850730,大島哲夫,B85-096
1432,ASPIDI Yniponi  ｲﾇﾜﾗﾋﾞ  ,富山県平村上梨300m,830915,大島哲夫,B84-083
1393,ASPIDI Yniponi  ｲﾇﾜﾗﾋﾞ  ,富山県立山町岩窒,820915,大島哲夫,B84-083
1520,ASPIDI polybl  ｲﾉﾃﾞ  ,富山県細入村猪谷,741025,大島哲夫,B85-096
1542,ASPIDI polybl  ｲﾉﾃﾞ  ,富山県八尾町桐谷,850730,大島哲夫,B85-096
0353,ASPIDI polybl  ｲﾉﾃﾞ  ,富山県氷見市殿田20m,820923,大島哲夫,B83-117
1527,ASPIDI tagawa  ｲﾉﾃﾞﾓﾄﾞ,富山県八尾町桐谷,850730,大島哲夫,B85-096
1540,ASPIDI tagawa  ｲﾉﾃﾞﾓﾄﾞ,富山県八尾町桐谷,850730,大島哲夫,B85-096
0418,PTERIS multif  ｲﾉﾓﾄｿｳ  ,富山県魚津市升田40m,820905,大島哲夫,B83-117
1327,ASPIDI suboch  ｲﾌﾞｷｼﾀﾞ,富山県立山町白岩90m,750821,大島哲夫,B84-083
1531,PTERIS interm  ｲﾜｶﾞﾈｾﾞ,富山県八尾町桐谷,850730,大島哲夫,B85-096
0328,ASPLEN varian  ｲﾜﾄﾗﾉｵ  ,富山県魚津市東又谷,800820,大島哲夫,B83-117
1574,POLYPOL puncta  ｲﾜﾋﾒｸﾗﾋ,富山県小矢部市小森谷,840805,大島哲夫,B85-096
0355,ADIANT atrata  ｲﾜﾍｺﾞ  ,富山県井波町東乗寺,821010,大島哲夫,B83-117
1326,DIPLAZ cavale  ｲﾜﾔｼﾀﾞ  ,富山県立山町白岩90m,750821,大島哲夫,B84-083
```

データ入力

データ入力は，コンピュータを扱って仕事を進める際に，最も重要で神経を使い，時間がかかる作業である。しかし，データベース完成後の大きなメリットを得るためには，どうしても行わなければならないものである。

データの入力には，ワープロで作成した文書をＢＡＳＩＣと互換性のある形式に変換して，そのままデータベースとする方法と，ＭＳ－ＤＯＳのEDLIN（エドリン，ラインエディタ）という編集プログラムを用いる方法とがある。前者の方法で行うためには，ワープロにファイル変換機能が付いていなければならない。

◇ｉ　EDLINによる場合

入力方法は比較的簡単で，基本的には図1のように，1行づつそのままの順にキーボードをたたいていけばよい。ＥＤＬＩＮの起動は，ＭＳ－ＤＯＳ起動後Ａ：（またはＡ＞）のプロンプトが表示されてから行う。EDLIN TESTと入力すると「新しいファイルです＊」と表示され，TESTというファイルがこれからのデータ入力に備え用意されたことを示す。ここではファイル名にＴＥＳＴを用いるが，使用する人のニーズにより８文字以内であれば，別に何であってもかまわない。＊のマークはEDLINが命令受付状態にあることを示すもので，これが表示されている時にＣ（行コピー），Ｅ（終了とファイル保存），Ｉ（入力，行挿入），Ｌ（リスト表示），Ｑ（編集中断），Ｒ（文字置き換え），Ｓ（検索），Ｔ（ファイル合成），その他の命令をキーボードから入力する。

データを入力するにはまずＩと入力する。データの行番号を示す1：が表示されるので，あとはデータを左から順に，例えば２３３０，アイアスカイノ，富山県・・・・・（学名省略）という具合に打っていく。カンマはデータの区切りとなるので，一定の場所に入れておく方が，後にこのデータを別の簡易言語やＢＡＳＩＣで利用する時に都合がよい。1行目の入力が終ったらリターンを押す。1行目のデータが記憶され，２：と表示される。ひき続き2行目のデータを入力していくが，この時，1行目に入力したデータと共通する文字がある場合は，その共通位置でＰＦ１キーを押すことにより1文字づつ2行目にコピーすることができる。例えば2行目で１６６２，と入力した後でＰＦ１を7回押すと一文字づつア．イ．ア．ス．カ．イ．ノとコピーされてくる。このコピー機能は，ＰＦ１を押し続けている間続けられる。途中で入力の誤りに気付いた場合は，一応その行のデータを最後まで入力してリターンを押した後，ＣＴＲＬキーとＣキーを一度に押して入力を中断する。＊が表示されたら，誤りのある行の番号を入力する。指定の行のデータが表示され，その下段に訂正用の行が用意されるので，まずＰＦ１を押し続け誤りの直前まで上の行をコピーする。そして，正しいデータを入力しＰＦ３を押すとカーソルの位置以降に上段のデータがすべてコピーさ

れ，最後にリターンを押すと訂正が終了する。入力再開には，行番号の最後を意味する♯に続けてＩを入力する。

今まで入力したデータを見る場合は，＊が表示されている時にＬをいれる。入力方法は1，15Lというふうに，表示開始行と終了行とを順に指定する。終了行を省略すると開始行から２４行表示する。２４行はちようど1画面分である。1行に入れるデータはあらかじめできるだけ７２字以内に収まるよう調整しておくと1レコードが2行にまたがらず画面が見やすくなる。

入力に関して必要な命令は以上である。さらに，使用すると便利なEDLINの機能にＳ（Search）とＲ（Replace）とがある。Ｓは編集中のファイルから文字を探す命令で，指定した文字がファイルのどの行に存在するかを検索するのに用いる。例えば，1，500，Sアイアスカと入力すると，EDLINは1行目から500行目の範囲にアイアスカという文字列を含む最初の行を一瞬のうちに表示する。500を省略すればデータの最後までが検索の対象となる。該当するデータをすべて表示させるには，Ｓの直前に？を入力する。こうするとＥＤＬＩＮは，該当するデータが見つかるたびに「いいですか＜Ｙ／Ｎ＞」と聞いてくるので，Ｎを入力し続ければ該当データが見つからなくなるまで検索が続けられる。Ｓによる検索は，条件を1項目しか指定できないが，編集中に実行できるのでかなり有用な命令である。複雑な検索には後述のＦＩＮＤを用いる。

Ｒは編集中のファイルに存在する文字列を，新たに指定する文字列に置き換える命令である。例えば，1，300，RXXΛＺ太田道人の命令は1行目から３００行目に含まれるＸＸという文字を太田道人に置換せよということである。ΛＺはＣＴＲＬキーとＺキーとを同時に押して入力する。この機能は，データを入力していく際，原稿の複数の行に同一の文字列が含まれている場合に便利で，同一部分をあらかじめ簡単な記号で代用しておき，後で正しい文字に変換するという使い方ができる。

EDLINを終了するには，Ｅを入力する。これにより，入力したデータがディスクに（本例では）TESTというファイル名で記録される。これ以後の編集や処理は，このファイルを単位として行うことになる。

◇ii　ワープロソフトによる場合

手持ちのワープロソフトに文書変換機能が付いている場合は，データ入力がワープロ画面で行えるため，データベース作成作業が極めて容易になる。

ファイルの変換は，ワープロ編集画面の1行が1レコードとして行われていくので，あらかじめ書式を設定しデータが2行に渡らないようにしておく。入力にあたっては，数字，カタカナ，カンマは半角（１バイト文字）で打っておくことが望ましい。これはＭＳ－ＤＯＳやＢＡＳＩＣにおいては2バイト文字の入力はやや面倒である上にレコード長も長くなり，さらに，プログラムによる数値計算は1バイト文字でないと行えないからである。また，カンマはデータベースにおいてデータの各項目の区切

りを意味する記号で，１バイト文字でないと意味をなさない。したがって地名や人名など漢字の方が見やすいデータ以外は半角で入力しておく必要がある。この方法で作成されたファイルの訂正，変更は◇ｉのＥＤＬＩＮによって行うことができる。

検　　索

標本データベース作成の一番のメリットは速い検索が可能なことであろう。「ワラビの標本では富山県のどこで採集されたものがあるか」，「呉羽山ではどのようなシダ植物が採集されているか」などの検索処理には，ＭＳ－ＤＯＳのＦＩＮＤという命令を使う。ＦＩＮＤはディスク上の指定したファイル全体を片っ端から検索し指定文字を含む行を探しだす機能である。MS－DOSのＡ：が表示されている時に入力を行う。ＦＩＮＤ“ワラビ” TESTと入力すると，フロッピィディスクが回転し，TESTファイルに存在する「ワラビ」という文字を持つレコードがすべて画面に出力される。ＦＩＮＤの結果を画面ではなく，任意のファイルあるいはプリンターに出力するには，上記の命令を続けて＞WARA－BI，＞ＰＲＮのように入力する。

ＦＩＮＤの所要時間は，図1の様式で書かれた2411件のデータを持つファイルで４０秒，約６０件／秒である。この速度はファイルの大きさにほぼ比例する。すなわち，図１の１レコードの平均の長さは８５バイトであるから，約5100バイト／秒がＦＩＮＤ命令の性能と考えられる。つまり１レコード長が短かければ，単位時間当りの検索件数が多くなるため，より高速の検索を行うためにはデータベースは読み易さを損なわない程度に，あらかじめ簡略な仕様にしておく必要がある。ＦＩＮＤのように片っ端から検索していく速度を他の言語によるものと比較した場合，簡易言語のｄＢａｓｅⅡで３０件／秒，ＢＡＳＩＣで14.６件／秒と，ＦＩＮＤはまずまずの速さである。

ただしｄＢａｓｅⅡの場合,あらかじめ検索する項目を指定してＩＮＤＥＸファイルの作成を行っておいた場合には，ＦＩＮＤの２０倍近い速度が得られ，データ数が多くなるにつれこの効果が大きくなる。しかし，ｄＢａｓｅⅡの高速検索は，条件指定文字と項目の先頭文字とが一致しなければならず，学名や人名の検索には適しているが採集地の細かな検索には不向きである上，INDEXファイルの容量が以外と大きく，1ディスクに収容できるデータ量が少なくなるという短所がある。また，検索にあたっては，INDEXの有無に関わらず，検索しようとする項目名を指定する必要があり，あらかじめ各項目の内容と項目名を使用者が記憶しておかなければならない。一方のＢＡＳＩＣにおいては，ＩＮ－ＤＥＸファイル作成やデータのランダムファイル化を行い，さらにプログラムを工夫すれば，上の条件で数秒程度の検索速度達成も可能であるが，データあたりのディスク使用部分はさらに多くなりプログラムはかなり複雑なものになる。

上記の速度はいずれも該当データの表示を行わないで比較したものであり，実際の使用時には，画面やプリンターに出力するので，検索によって出力されるデータ数が多くなると，これにかかる時

間がコンピュータ使用時間の大部分を占めるようになる。したがって，容量の小さなデータベースを使っている範囲内においては，ソフトによる検索速度の差はあまり大きな時間差となって表れないのである。

ＭＳ－ＤＯＳのＦＩＮＤは，検索指定文字を行のどこかに含む全ての行を表示するので，検索項目の内容等を全く考慮しなくてもよいという利点があった。半面，先のＦＩＮＤ“ワラビ” ＴＥＳＴの例では，イヌワラビやヒメワラビ等も出力されてしまう。これを避けるために「ワラビ」の始めに１文字空白を奈けて「　ワラビ」としたり，「ワラビ」と科名「ＰＴＥＲＩＤ」の重複検索を行うなどの工夫が必要である。重複検索はＡ：に続けて，ＦＩＮＤ “ワラビ” TEST｜FIND “ＰＴＥＲＩＤ” と｜を入れてＦＩＮＤ命令を２回入力する。２回目にはファイル名は不要である。

ＦＩＮＤ命令のオプションとして－Ｖ，－Ｃ，－Ｎ（メーカーによって－が＞となる）があり，－Ｖは指定文字を含まないものを出力する時に，－Ｃは指定文字を含む行の数をカウントする時に，－Ｎは指定文字がファイルの何行目の位置にあるかを見る時に，ＦＩＮＤ－Ｖ“ワラビ” TESTのように入力して用いる。

分　　類（SORT）

入力されたデータや検索によって出力されたデータは，順不同で見にくいため任意の項目で配列し見やすくする必要が生じる。この作業をソートという。ソートは数値あるいはアルファベット順に並べ代えるだけの処理であるが，この結果隣接する行の文字が揃うので一応の分類がなされるのである。漢字の場合はＪＩＳコード順（およそ音読みの順）に配列される。

ソートは，ＭＳ－ＤＯＳのＳＯＲＴという命令を使う。ＳＯＲＴは行の指定した位置から右側にある文字列をソートの対象とするため，ソートを必要とする項目はあらかじめ項目ごとにデータの先頭位置を揃えておかなければならない。図１のデータを学名順にソートする場合は，Ａ：のプロンプトに続けてSORT －＋６ ＜OSMと入力する。－＋６はレコードの左端から６文字目のデータを意味し，＜はソートの対象となるファイルを意味する。ソートの結果を任意のファイルあるいはプリンターに出力するには，上記の命令に続けて＞OSM２，＞PRNのように入力する。

ソートの処理速度は，ファイルの大きさの比の２乗にほぼ比例し，図１の様式のデータ ４００件（ファイルの大きさで３４キロバイト）で約４０秒である。

SORTには，一度に扱えるファイルの大きさが約６４キロバイトまでという限界がある。このため大きなファイルをソートする際には，まずいくつかのファイルに小分けし，おのおのにＳＯＲＴを実行し，最後にファイルをつなぐという一連の作業が必要である。この作業はＦＩＮＤ，SORT，COPY等の繰り返しであるため，大型ファイルのソートのたびに同一の命令を入力する手間を省くため，あ

```
A:EDLIN OSMSORT.BAT
ファイルをすべて転送しました
*L
    1:*REM ｵｵｼﾏ ﾃﾞｰﾀ OSM ｦ SORT ｽﾙ。
    2: FIND "ASPIDI" OSM >ASP1
    3: SORT /+29 <ASP1 >SOSM1
    4: DEL ASP1
    5: FIND /V "ASPIDI" OSM >ASP2
    6: SORT /+29 <ASP2 >SOSM2
    7: DEL ASP2
    8: COPY SOSM1+SOSM2 SOSM
    9: DEL SOSM1
   10: DEL SOSM2
   11: REM TITLE ｦ ｺﾞｳｾｲｾﾖ。
   12: EDLIN SOSM
   13: DEL SOSM.BAK
*
```

らかじめ一連の命令を簡単な自動実行プログラムとして作成しておくとよい（左図）。このプログラムは入力の項で述べたEDLINを用いて作成する。EDLIN OSMSORT.BAT と任意のファイル名の最後に自動実行プログラムを意味する。.ＢＡＴをつけて入力する。左図の1から10行目までがさしあたり必要なプログラムで，図1の様式のOSMという約100キロバイトのファイルを，ASPIDI（科名）という文字を含むか否かで2つの仮のファイルに分け，それぞれを和名順（和名は左から29文字目）にソートとし，8行目のCOPYによってSOSMという一つのファイルに合成している。REMは注意書きを意味し，これから右側にメモ等を書くための記号，DEL はファイルを消去する命令である。このプログラムの実行は，Ａ：続けてOSMSORTと入力するだけである。

図2の自動実行プログラムは，ファイル名や文字列，数値を置き換えることによって種々の目的に検索やソート等を自由に操作する応用が可能である。

ＭＳ－ＤＯＳの関連命令

COPY：ファイルを複写する。例えばドライブAのファイルOSMをドライブBへAA というファイル名でコピーする場合はCOPY Ａ：OSM Ｂ：AAと入力する。同一名でコピーする場合，AAは不用。データベース専用のディスクには，あらかじめこのCOPYを用いて，システムディスクからEDLIN.COMとＦＩＮＤ．EXE，SORT．EXEの3つのファイルを複写しておくとよい。

DIR： ディスク上のファイルの大きさや作成年月日をリストにして表示し，ディスクの残り空間を計算する。

TYPE： ファイルの内容を先頭から全て表示する。

学名辞書について

標本データベースを利用していく際，科や属などの分類群ごとにデータをまとめる必要が生じることがある。この場合には，どうしても学名辞書を作成しなければならず，この入力にかなりの時間がかかる。筆者は一応フルセンテンスの学名・和名対照表を作成しているが，植物目録など特別に体裁を整える必要がある場合を除き，種名の区別や検索とソートに大きな不自由がない範囲で字数制限を行ったものを使用している（図１）。データベースへの学名自動挿入は，ＢＡＳＩＣのプログラムを作成して行っており，処理速度は今のところ3件／秒程度である。

最　後　に

本紹介は，ＭＳ－ＤＯＳを使って標本データベースを作成し利用することを目的として述べたが，同様の操作で名刺管理や文献整理、数値データベースの作成，メモ書き等幅広く応用が可能であることは勿論であることを付記させていただく。

本紹介をお読みいただき，使用感，問題点，要望等をお聞かせ願えれば幸いである。本方法が諸兄の研究にささやかな手助けとなれば幸いである。

参　考　資　料

アスキー出版局編著，1984．標準ＭＳ－ＤＯＳハンドブック．アスキー出版局．

アスキー出版局テクライト編著，1986．ＭＳ－ＤＯＳが見えてくる本．アスキー出版局．

金井弘夫，1982．植物地理学的情報蓄積のための基礎的技法．国立科学技術博物館研究報告Ｂ類 8(1)：1－14．

金井弘夫，1983．長野県フロラ作成資料の電算機処理，長野県植物研究会誌16：1－7．

太田道人，1987．富山市科学文化センター収蔵資料目録第1号　進野久五郎植物コレクション．

清水建美，1984．コンピュータ利用によるフロラ作成と地理学的解析．昭和57・58年度科学研究費補助金（試験研究）課題番号57840038成果報告書．

田中一郎，大橋均，1986．はじめてのＭＳ－ＤＯＳ．新星出版社．

脇英世，1986．日本語ファイルの作りかた．科学朝日46(1)：115-119．

脇英世，1986．データ整理を効率化する．科学朝日46(2)：131-135．

脇英世，1986．ＭＳ－ＤＯＳを使いこなす．講談社ブルーバックス．