A DATA SYSTEM FOR TAIWANESE GRASS SPECIES (PART I): SPECIMEN LABEL DATA STORAGE AND RETRIEVAL[1]

CHANG-FU HSIEH[2]

(Received for publication February 16, 1986 and in revised form March 17, 1986)

ABSTRACT

A data storage and retrieval system has been developed for assembling and manipulating specimen information concerning Taiwanese grass species. It was designed for use on an affordable and widespread microcomputer, and is most suitable for small colleges and universities with small herbaria or for personal collections. Steps involved in the creation of item fields, character sets, file structure, and their updating are briefly described. The query system was so designed that it can be used to retrieve records with desired attributes and from them to give results in a variety of forms, including specimen labels, distribution map, and other relevant information.

INTRODUCTION

The continuous botanical surveys since the establishment of the Taihoku Imperial University (the Predecessor of the present National Taiwan University) in 1928, have resulted in the housing of about 15,000 specimens of Gramineae in the Herbarium of the Botany Department (TAI). While TAI's special emphasis is on the Taiwan Flora, it also contains plenty of specimens from South Mainland China, Pacific islands, Japan, and many other countries.

For a maximal utilization of the label information, a computerized data storage and retrieval system has been developed, which will incorporate the Gramineae collections of the TAI into a complex data base.

The principal aim of this system is to cater for a variey of needs, such as printing of labels, generating distribution maps or extracting the relevant data according to various kinds of queries.

The importance of electronic data processing in herbaria has been discussed by several authors (Crovello, 1967; Crovello & MacDonald, 1970; Rensberger & Berry, 1967; Soper & Perring, 1967; Hall, 1972, 1974; Morse, 1974) and many systems have been sucessfully used. The EDP-IR in the Columbian National Herbarium (Forero & Pereira, 1976), the EDP technique designed for Florida's Central-East coast vegetation (Sweet & Poppleton, 1977), the Precis of the National Herbarium of South Africa (Morris & Glen, 1978), the optical-scan data encoding system at the University of Georgia Herbarium (Jones et al., 1983) are examples among them.

The present system is designed primarily for handling a small set of label data. It takes the advantage of the affordable and widespread microcomputer (Apple II),

(1) This research has been supported by National Science Council (NSC 72-0201-B001a-62).
(2) 謝長富, Associate Professor, Department of Botany, National Taiwan University.

and is highly suitable for local flora study. In the case of processing a larger number of specimens, a few modifications are necessary to allow remote terminal control of the system on a large time-sharing computer.

LABEL DATA FORMAT

1. Field creation

The first step taken by the system is field creation that specifies each item (field) name and it's width to be contained in the database. For convenience, the label items are divided into two classes. The first class contains information that are commonly fixed for all specimens. These include taxon number, family name, species name, locality, collector, collector's number, and date of collection. The second class is devoted to the citations of conditions that seem very inconsistent and are varied with personal interpretations and countries. The program ITEM DEFINING makes it possible to set out a hierarchic structure, that means each ecological item can contain several levels of subitems. The data recorded from each specimen are shown on Table 1. The underlines following each item indicate the width of that item. Some special features of the data are given below:

(1) Taxon number

To facilitate the handling of scientific names, a number of nine digits has been assigned to each species. The first three digits are family number, which sequentially ascendes corresponding with the systematic sequencce of the family following that of the Flora of Taiwan (1975-1978). Generic name and specific epithet are coded with numbers of three and two digits respectively. Both are arranged in alphabetical order. The last digit is assigned to infraspecific categories.

(2) Grid reference

For easiness of map construction and distributional data plotting, a grid code system was adopted, which divides the whole Taiwan area into 97x165 units. Each grid unit is approximately 6.3 square kilometers. The grids so divided are to fit into the coordinate system of the high-resolution graphics mode of the Apple II computer.

(3) Loan

The loan field will be used to keep track of incoming and outgoing loans of specimens. A sequential number will be automatically generated, which is linked to an on-loan file.

(4) Ecological information

Ecological information includes altitude, physiographic regions, land feature and land use, substrate, moisture regime, aspect, and light. The altitude is quoted in meters and with a conventional indication of range. The other items need to be encoded from personal observation in the field or only if the relevant data can be discerned from the backlog specimens.

(5) Notes

In this field,up to 80 characters can be used to further describe the locality of collection and habitat conditions. They are all entered as text.

Table 1. Data encoding form.

(01) Taxon number: ---------
(02) Family: ----------
(03) Plant name: ----------
---
(04) Coordinate: X: --
Y: ---
(05) Collector's name: ----------
(06) Collector's number: -------
(07) Date collected: Year: ---
Month: --
Day: --
(08) State of the specimen
1 Vegetative phase only
2 Other
WHICH ONE? -
(09) Type status
1 Holotype
2 Lectotype
3 Neotype
4 Isotype
5 Syntype
WHICH ONE? -
(10) Label language
1 English
2 Chinese
3 Japanese
4 Other
WHICH ONE? -
(11) Loan: ----
(12) Altitude: ----
---
(13) Physiographic provinces
1 Coast
2 Volcano
WHICH ONE? -
(14) Land features
1000 Natural or semi-natural vegetation
1100 Herbaceous types
1200 Shrub/scrub types
1300 Forest types
1310 Conifer forests
1320 Conifer/broad-leafed mixed forests
1330 Broad-leafed forests
1340 Littoral forests
2000 Disturbed lands
2100 Urban
2200 Agriculture
2210 Paddy field
2220 Dry farmland
2230 Orchard
2240 Garden
2250 Pasture

Table 1. (Continued).

2300 Plantation
2400 Planted meadow
2500 Road/railwayside
2600 Recently burnt

WHICH ONE? -----

(15) Substrate
10 Soil
11 Sand
12 Loam
13 Clay
14 Gravel
15 Laterite
16 Humus-rich
20 Stone
21 Coral reef
22 Limestone
23 Sandstone
24 Shale
25 Slate
26 Schist
30 Water
31 Lake/pond/dam
32 River/stream
33 Ditch
34 Estuary/sea

WHICH ONE? --

(16) Moisture regime
1 Well drained
2 Poorly drained
3 Marsh
4 River/stream bank
5 Dry river/stream bed

WHICH ONE? --

(17) Aspect
1 N
2 NE
3 E
4 SE
5 S
6 SW
7 W
8 NW

WHICH ONE? -

(18) Light
1 Full sun
2 Light shade
3 Dense shade

WHICH ONE? -

(19) Notes: ----------
------

2. Map construction

The program MAP CONSTRUCTION permits one to plot his own map. Owing to the limitation of the high-resolution graphics mode, the map constructed for Taiwan area only includes the boundaries of all counties. Other detailed features are completely omitted.

3. Character generation

A characte generator was designed for the reason that there is no provision for generating lower case on Apple II computer. The program CHARACTER GENERATOR will provide for the full ASCII codes, including both standard and emphasized sets.

DATA ENTRIES

Once the items, map, and character sets have been entered into the computer, three files are created on the disk. Then the following program LABEL DATA INPUT will help to create the label data file. It shows items line by line using the same format as Table 1. The user merely responds to it through the keyboard and skip over those not available. Since taxon names and taxon numbers should be repeated continuously when a successive set of backlog specimens from the same species are recorded, considerable effort can be saved by having a skip option set up. The same is the case for ecological information, which is unobtainable from many specimens.

RETRIEVAL SYSTEM

The partially inverted file structures used for pollen data system (Hsieh & Huang, 1983) are the basis for the current system, in which all specimen records containing a given item attribute will have their identifiers (sequence numbers when records are entered on the disk) listed in a monotonic sequence within a variable length record, the address of which is an element of the key directory for that attribute. That is three linked files, namely specimen data file, list file, and key directory file, are contained in the retrieval system. The program INVERT allows one to create key directoris for all label items except the notes, for which a search argument can be expected. The system designed in such a way can meet the requirement, in which the choice of a subset of item attributes for query may be at the discretion of the user.

The results of the search are produced on the screen or on a printer. The program PRINT specifies the ways they are to be printed.

1. Label printing

Two pages of high resolution graphics mode (HGR and HGR 2) are set up for a label display. The first page HGR is used for label text and the second for distribution map. The text page consists of a maximum of 20 lines, each of 40 characters. Before the text is shown on the screen, the whole characters defined by the character generator must have been loaded. The starting display location for each word or sentense should be specified and their length under control. The heading of label and scientific name are in emphasized form and others in standard form. When the screen displays are finished, a label printing option allows the user to obtain a combined physical copy of the two graphics pages (Fig. 1). The resulting label is 6.5x10.5cm.

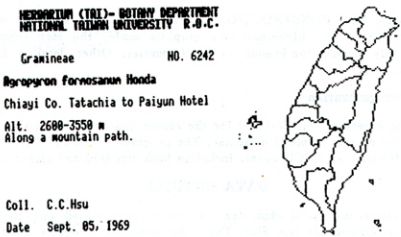

Fig. 1. Example of a printed label.

2. Distribution maps

In present version of the system, it is possible to plot distribution maps of the species specified by the user. A matrix printer was used. The distortion due to the aspect ration of the screen picture, which is not 1:1, was corrected. The resulting map is shown in Fig. 2.

Fig. 2. Distribution records of *Bothriochloa ischaemum* (L.) Keng

3. User's defined layouts

It is possible for the user to define his own conditions. That means he can select an item attribute or multiple item attributes for input. The data to be printed out can be also specified. These options are very useful when some relevant information is desired. For example, a list of all plants collected in Mt. Alishan, or all places Sasaki visited in 1930, or all species of Yamingshan area flowering in May can be easily prepared.

FILE MAINTENANCE

File maintenance can be put into three categories: addition, modification and deletion. It enables the computer to add new specimens to the data bank and to make changes in the information already recorded. Once a record has been added or an item attribute, from which the key directory was made, been modified, all files involved in this system must be updated.

DISCUSSION

Although several general use systems have been designed for data storage and retrieval on microcomputers (For exampler, dBASE II and dBASE III), they appeared too complex for use or there are some limitations in handling specimen information. The present data bank serves as an aid for small colleges and universities with small herbaria or for personal collections. It possesses a high degree of flexibility and is easy to use. A variety of useful outputs can be obtained by means of different options. It is expected that this system may be even more valuable if we add some morphological, anatomical and other information such as economic narrative to the data bank. Another extension being considered at present is the inclusion of type specimens and specimens other than Gramineae. By the time large numbers of label records have been accumulated, the real advantages of the system will become apparent (e.g., to produce as many regional or local guides and lists of plants that one wants; to show the plant richness of areas where habitat destruction should be avoided; to provide the list of conserved rare and endangered species).

REFERENCES

Crovello, T. J., 1967. Problems in the use of electronic data processing in biological collection. Taxon 16: 481–494.

---------- & R. D. MacDonald, 1970. Index of EDP-IR projects in systematics. Taxon 19: 63–76.

Forero, E. & F. J. Pereira, 1976. EDP-IR in the National Herbarium of Columbia (COL). Taxon 25: 85–94.

Hall, A. V., 1972. Computer-based data banking for taxonomic collections. Taxon 21:13–25.

---------- , 1974. Museum specimen record data storage and retrieval. Taxon 23: 23–27.

Hsieh, C. F. & T. C. Huang, 1983. A data storage and retrieval system for Formosan pollen grains. Taiwania 28: 42–67.

Jones, S. B., N. C. Coile and R. Martin, 1983. The University of Georgia Herbarium (GA) optical-scan data encoding system. Taxon 32: 47–50.

Morris, J. W. & H. F. Glen, 1978. PRECIS, The National Herbarium of South Africa (PRE) computerized information system. Taxon 27: 449–462.

Morse, L. E., 1974. Computer-assisted storage and retrieval of taxonomy and systematics. Taxon 23: 29–41.

Rensberger, J. M. & W. B. M. Berry, 1967. An automated system for retrieval of museum data. Curator 10: 297–317.

Soper, J. H. & F. H. Perring, 1967. Data processing in the herbarium and museum. Taxon 16: 13–19.

Sweet, H. C. & J. E. Poppleton, 1977. An EDP technique designed for the study of a local flora. Taxon 26: 181–190.

The editorial committee, 1975–1978. Flora of Taiwan, Vol. 1–6. Epoch Publishing Co., Ltd., Taipei.

臺灣禾草之資料系統(I)：標本採集資料之處理

謝長富

摘　要

本資料系統建立之目的在使標本之標籤資料達到最有效之運用，並以臺大植物系標本館之禾本科標本爲設立範本。全套系統係建立於廣泛流行之個人電腦上，以提供小型標本館或個人採集之用。系統之設立可區分爲五步驟：(1)資料格式之設立，包括資料項目、字元產生器、及臺灣圖形等；(2)標籤資料之儲入；(3)檔案結構之建立；(4)查詢及其結果之輸出，此包含印製標籤、種類分佈圖、及符合使用者需求之各種資料；(6)資料及檔案之更新。