

UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

REFS2DB--A PC Program to Reformat Publication-Format
Bibliographies for dBASE or PARADOX

by

Carol N. Gerlitz¹

Open-File Report

89-560A Program Documentation (paper copy)
89-560B Source Code, Executable Program, and Documentation,
(5-1/4" floppy disk)

1989

This report is preliminary and has not been reviewed for conformity with U.S. Geological Survey editorial standards. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Although this program has been used by the U.S. Geological Survey, no warranty, expressed or implied, is made by the USGS as to the accuracy and functioning of the program and related program material, nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection therewith.

¹U.S. Geological Survey
Denver, Colorado 80225

CONTENTS

Page

Introduction	1
Program disk	1
Restrictions on original ASCII data file	2
Database fields	2
Author	2
Year	2
Title	2
Citation	4
Large bibliography files	4
Import the delimited file into a database manager	4
dBASE	4
PARADOX	5
Acknowledgments	5
References	6

ILLUSTRATIONS

Figure 1. Format of ASCII reference files prior to reformatting by Program REFS2DB or Program REFS2PAR	3
---	---

INTRODUCTION

A bibliography stored in a database is more accessible and can be searched more quickly than an ASCII file that contains the same bibliographic information in publication format. For this reason, I wrote Program REFS2DB. The program reads an ASCII file of paragraph-formatted references and rewrites it into an entry-delimited file you can import into a database manager.

REFS2DB breaks each reference into author, year, title, and citation fields for the database record. The fields are enclosed in double quotes and separated by commas (i.e., ".....", ".....", ".....", "....."). Both dBASE and PARADOX accept this entry-delimited format. Once the references are in a database, you can apply the rapid search and sort capabilities of the database manager to the bibliography. You can also extract subsets of references by searching on words or phrases in any of the fields. If you want to add a separate keyword field to the database, both dBASE and PARADOX have provision for doing so.

REFS2DB is written in Turbo PASCAL and runs on an IBM PC or PC-compatible computer. Delimited files written by the program were tested in dBASE III PLUS and PARADOX, version 3.0.

PROGRAM DISK

This report includes a 5-1/4-inch floppy disk with program documentation, REFS2DB.PAS, REFS2DB.EXE, REFS2PAR.PAS, REFS2PAR.EXE, MAXLEN.PAS, and MAXLEN.EXE.

The disk contains two versions of the reformatting program. REFS2DB creates delimited "title" fields up to 254 characters long. Use files written by REFS2DB with dBASE, which allows a maximum of 254 characters for character fields. REFS2PAR creates delimited "title" fields up to 255 characters long. Use REFS2PAR with PARADOX, which permits character fields up to 255 characters in length. REFS2DB.EXE and REFS2PAR.EXE are the executable programs. REFS2DB.PAS and REFS2PAR.PAS contain the PASCAL source code for these programs.

Program MAXLEN evaluates a file created by REFS2DB or REFS2PAR and determines the maximum length of each database field (author, year, title, citation) in the file. MAXLEN.EXE is the executable program, and MAXLEN.PAS is the PASCAL source code. Because dBASE requires that you set up your record structure, including field lengths, before you import the data, you should run MAXLEN to determine the required field lengths for each field in your reference database.

MAXLEN can also be useful for debugging an ASCII file of references in paragraph format. During testing of REFS2DB, MAXLEN called attention to references that were written incorrectly. References with errors had extremely long field lengths, much longer than they should have had.

RESTRICTIONS ON ORIGINAL ASCII DATA FILE

References in your ASCII file can be in one of three paragraph formats: hanging indent, flush left, or standard indent (fig. 1). Indented formats may be indented any number of spaces; the program always looks for the first non-blank character.

The entire file can be indented one space along the left side (can have a column of single spaces along the left edge) of the file. The file can also have one or more blank lines at the beginning of the file and between references. "Flush left" files MUST have a blank line between references.

The original reference file should be ASCII and should contain only the references--no page breaks, no formatting characters, and no title at beginning of file. Each line of the original file should have no more than 80 characters.

DATABASE FIELDS

REFS2DB and REFS2PAR reformat each reference into a single-line record that contains five database fields--author, year, title1, title2, and citation (see the following sections for a description of each field). The fields are enclosed in double quotes and separated by commas. Because dBASE and PARADOX limit character field length to 254 and 255 characters, respectively, fields longer than this will be truncated if you import the file into your database manager. dBASE will also truncate any field in your delimited file that is longer than the field length you set when you create the file structure.

AUTHOR

The author field may be any length. The program will find and delimit an author field longer than 254 characters, but if it is longer, it will be truncated at 254 characters in dBASE and at 255 characters in PARADOX.

YEAR

The program will delimit a year field of any length. This allows you to make entries such as "1987a", "1970-1972 [1972]", "19xx [no date]", and "1873-1880 and 1885-1900". The length of the information given with the year is immaterial, but the year field will also be truncated if it is longer than the number of characters allowed by your database manager for a character field. In your original ASCII file, year information must begin with "19" or "18", must contain no comma, and must be followed by a comma. The program distinguishes between author and year on the basis of "19" or "18"; it distinguishes between year and title on the basis of the first comma following "19" or "18".

TITLE

Titles longer than 255 characters are not uncommon. For this reason, the program creates two title fields. If the title in your reference is long, the program will break the title between words, at or just before 254 (REFS2DB) or

a. Example of HANGING INDENT paragraph format.

Anonymous, 1952, Homestake--A South Dakota enterprise, in Sonnenberg, F.P., ed., Black Hills-Williston Basin: Billings Geological Society, 3rd Annual Field Conference Guidebook, p. 128-131.

Anonymous, 1985, Black Hills with a golden future: Mining Monthly, p. 32-33.

Babb, C.C., and Hoyt, J.C., 1905, Report of progress of stream measurements for the calendar year 1904, Part VII, Hudson Bay and Minnesota, Wapsipinicon, Iowa, Des Moines, and Missouri River drainages: U.S. Geological Survey Water Supply Paper 130, 204 p.

b. Example of FLUSH LEFT paragraph format.

Anonymous, 1952, Homestake--A South Dakota enterprise, in Sonnenberg, F.P., ed., Black Hills-Williston Basin: Billings Geological Society, 3rd Annual Field Conference Guidebook, p. 128-131.

Anonymous, 1985, Black Hills with a golden future: Mining Monthly, p. 32-33.

Babb, C.C., and Hoyt, J.C., 1905, Report of progress of stream measurements for the calendar year 1904, Part VII, Hudson Bay and Minnesota, Wapsipinicon, Iowa, Des Moines, and Missouri River drainages: U.S. Geological Survey Water Supply Paper 130, 204 p.

c. Example of STANDARD INDENT paragraph format.

Anonymous, 1952, Homestake--A South Dakota enterprise, in Sonnenberg, F.P., ed., Black Hills-Williston Basin: Billings Geological Society, 3rd Annual Field Conference Guidebook, p. 128-131.

Anonymous, 1985, Black Hills with a golden future: Mining Monthly, p. 32-33.

Babb, C.C., and Hoyt, J.C., 1905, Report of progress of stream measurements for the calendar year 1904, Part VII, Hudson Bay and Minnesota, Wapsipinicon, Iowa, Des Moines, and Missouri River drainages: U.S. Geological Survey Water Supply Paper 130, 204 p.

Figure 1. Format of ASCII reference files prior to reformatting by Program REFS2DB or Program REFS2PAR.255 (REFS2PAR)

255 (REFS2PAR) characters. The first character in the second title field will always be the blank space between the words; you will thus be able to concatenate the two fields during a retrieval without adding a space. In your original ASCII file, the title and citation must be separated by a colon; this colon should be the only one in the reference. The program distinguishes between title and citation on the basis of the colon. If the program finds no colon in your title/citation information, it will consider everything following the year to be title information and will write the data across the two title fields, and, if necessary, into the citation field, always breaking the information between words. If there is a colon in your title, the program will assume that the citation begins after the colon.

CITATION

The citation may be any length. The program will delimit a citation longer than 255 characters, but it will be truncated in dBASE at 254 characters and in PARADOX at 255 characters if it is longer.

LARGE BIBLIOGRAPHY FILES

Your bibliography may be so large that you have broken it into two or more files for easier handling. If you want to import all the files into the same database, you should first reformat each ASCII file separately using REFS2DB or REFS2PAR.

Next, concatenate the files using I/O redirection in DOS. For example, if you have 3 files to be put together (FILE1.INP, FILE2.INP, and FILE3.INP):

```
C> TYPE FILE1.INP>>BIGFILE.INP
```

```
C> TYPE FILE2.INP>>BIGFILE.INP
```

```
C> TYPE FILE3.INP>>BIGFILE.INP
```

This procedure will result in one large .INP file, BIGFILE.INP, that includes all the references in FILE1.INP, followed by all references in FILE2.INP, and finally, all references in FILE3.INP.

Creating one large file of delimited references allows you to set correct field lengths for the entire bibliography as you import it into your database manager. If you run Program MAXLEN on the large, final delimited file, you will know how long to set field lengths to avoid truncating any data in dBASE. If you import the complete delimited file into PARADOX, the database manager will automatically set field lengths long enough to accommodate data in the entire bibliography.

IMPORT THE DELIMITED FILE INTO A DATABASE MANAGER

dBASE

To convert an ASCII file of references into a database in dBASE:

1. Run Program REFS2DB to create the entry-delimited input file for dBASE.
2. Run Program MAXLEN to determine the database field length you need to accommodate the maximum length for each field in the delimited file.
3. At the dot prompt in dBASE, create the file structure:

.CREATE <database name> (see dBASE manual for details)

4. Bring your delimited reference file into dBASE:

.APPEND FROM <delimited filename> DELIMITED WITH "

You can modify the dBASE file structure to add a "keyword" field either before or after you append the delimited data file.

PARADOX

To convert an ASCII file of references into a PARADOX database:

1. Run Program REFS2PAR to create the entry-delimited input file for PARADOX.
2. From the PARADOX main menu, select "Tools".
3. From successive PARADOX menus, select "Export-Import", "Import", "ASCII", and "Delimited".
4. As PARADOX prompts you to enter a filename, use the name of the delimited file created by REFS2PAR. To the prompt "Enter table", respond with the name you have chosen for your database.

You can add a "keyword" field to a PARADOX database after importing the delimited data file.

ACKNOWLEDGMENTS

Program REFS2DB arose from a discussion with Ed DeWitt (Branch of Central Mineral Resources) as he prepared to publish a file of 3,676 references on geology of the Black Hills (DeWitt and others, 1989). I am indebted to Ed for testing the program to obtain the entry-delimited files included in the Black Hills Open-File Report.

The original version of this program was written in VAX PASCAL for the Delta CUSMAP DATATRIEVE database system (Gerlitz, 1987). Files created by the VAX version are in tabular format.

REFERENCES

- DeWitt, Ed, Clemensen, Margaret, and Barari, Rachel, 1989, BHBIB-- Bibliography of Black Hills geology, 1852-1988: U.S. Geological Survey Open-File Report 89-443: A, Bibliography (paper copy); B, ASCII datafile (publication format--three 5-1/4" 360 Kb floppy disks); C, ASCII datafile (publication format--one 5-1/4" 1.2 Mb floppy disk); D, ASCII datafile (entry-delimited-- three 5-1/4" 360 Kb floppy disks); E, ASCII datafile (entry-delimited--one 5-1/4" 1.2 Mb floppy disk).
- Gerlitz, C.N., 1987, REFS-DTR, a reference database for VAX DATATRIEVE: U.S. Geological Survey Open-File Report 87-0637, 31 p.