

Find2000: A Search Tool to Find Date-Related Strings in SAS®

Sarah L. Mitchell, Qualex Consulting Services, Inc.
Michael Gilman, Qualex Consulting Services, Inc.

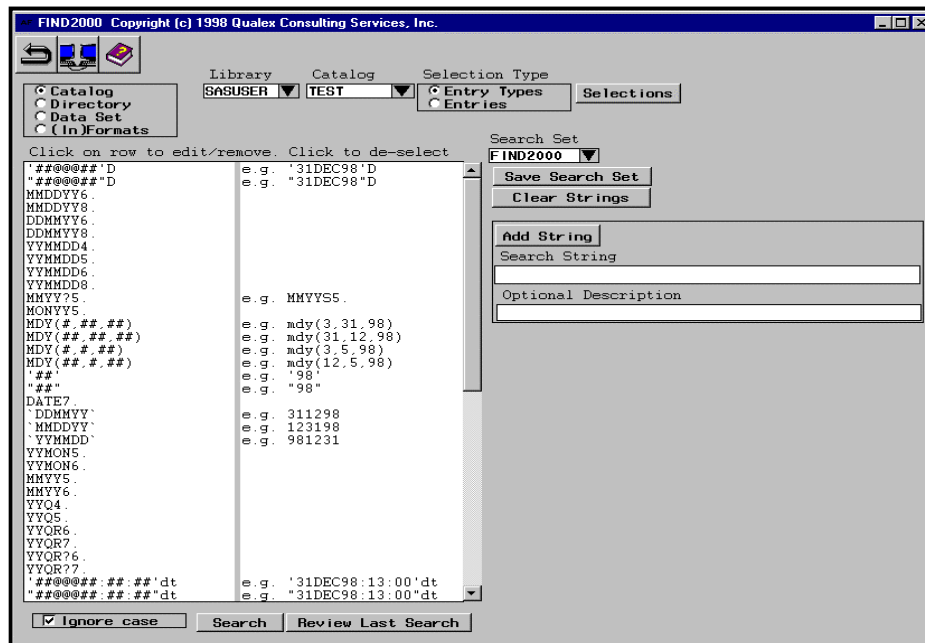


Figure 1

Abstract

Although SAS Version 6 is primarily year 2000 compliant, it may nevertheless be advantageous to find date-related strings that are associated with two digit years. Find2000 is a search tool written in SAS that facilitates finding date-related strings. Find2000 has the flexibility to customize the search criteria and save the search results in several ways.

Introduction

While the SAS Version 6 is primarily year 2000 compliant, end-user applications may not be. Though the YEARCUTOFF= option allows you to set the century window, it does not by itself identify date-related strings in your program or data. It may be important to identify date-related strings for the following reasons:

- The SAS system has the capability to access non-SAS data files from other systems such as Sybase and Oracle. These systems may not be Year 2000 compliant. This issue is more significant if you are using SQL pass-through.

- ASCII files that are used to store data may have date values stored with two-digit years; e.g. 980131.
- Many sites are still running SAS Version 5. SAS Version 5 is not year 2000 compliant, nor will it ever be. The YEARCUTOFF= option is not available. This is a particularly significant issue for those sites that have no intention to upgrade to later SAS versions.
- SAS Version 6 is primarily year 2000 compliant. The key word here is "primarily". There are variants of Version 6 such as 6.04 for DOS that are not compliant.

Find2000 is a SAS/AF® application that helps find date-related strings in SAS catalog entries, SAS data sets, and ASCII files. This paper describes the characteristics of the Find2000 tool by discussing the following points:

- Candidate date-related strings.
- Search string wildcards
- Code and data locations
- Search string sets
- Search techniques
- Search results

The paper does not discuss the general issues of SAS and year 2000 compliance.

Candidate Date-Related Strings

There are many strings in SAS that are date-related.

- Formats and informats
- Date functions
- Date literals

SAS has many date-related formats and informats such as:

- MMDDYYw.
- MMYXw.
- MONYYw.
- DATEw.

Many of the SAS date formats and informats are candidates for search strings. The width value (*w.*) primarily determines whether two-digit year values are a factor. For example, MMDDYY6. and MMDDYY8. imply two-digit years whereas MMDDYY10. does not.

The MMYXw. formats dates depicted such as 05-98. The *x* value represents the separator character and has values such as C for colon, D for dash, etc.

Other candidate search strings include date functions like MDY(), which can use a two-digit year as a parameter.

Date literals such as the following are also candidate search strings:

- '31DEC98'D
- "31DEC98:13:00"DT
- '311298'
- '981231'

Search String Wildcards

Find2000 provides special wildcard characters to facilitate the specification of candidate date search strings. The wildcards can be combined with any other wildcard or literal value. The following list of wildcards is provided:

?	Any one character or digit
#	Any one digit
@ @ @	Any 3-character month abbreviation (Jan-Dec)
`MM`	Any two digits between 01 and 12
`DD`	Any two digits between 01 and 31
`DDMMYY`	Six digit dates (311298)
`MMDDYY`	Six digit dates (123198)

`YYMMDD` Six digit dates (981231)

Note that the accent (`) character is used to delimit some of the wildcards.

Figure 1 displays the primary Find2000 screen with a view of the default search strings.

Code and Data Locations

Code and data can be stored in several different locations. These include the following catalog entries:

- SCL
- SOURCE
- PROGRAM
- HELP
- OUTPUT
- LOG
- CATAMS
- CBT
- FRAME
- SLIST
- SCREEN
- CLASS
- EIS
- FORMAT
- INFORMAT

Find2000 provides a screen (Figure2) that allows you to search all of these catalog entry types at one time. Note that you can search the SCL associated with Fsedit SCREEN entries.

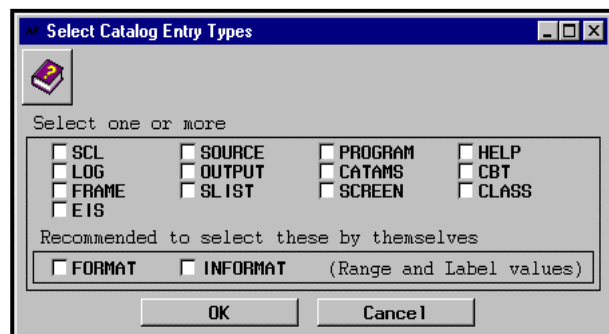


Figure 2

Find2000 also provides a screen that allows you to select specific catalog entries as shown in Figure 3. This screen also allows you to conveniently view the catalog entry source by double-clicking on the entry. The list of entries can be subset by entering a prefix or extension filter. The title of the list box displays the total number of entries available and the number of entries currently selected.

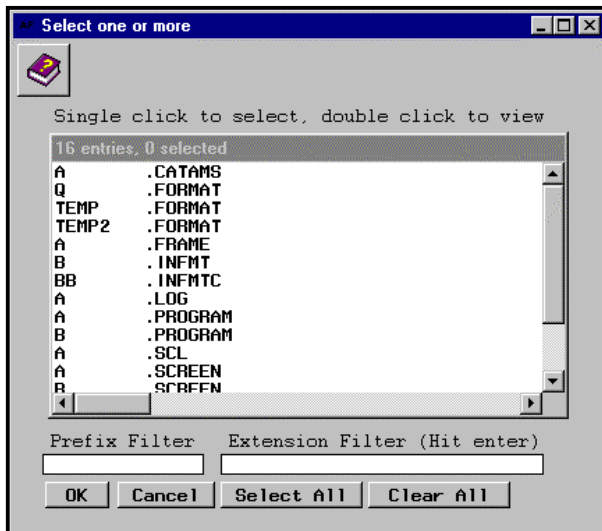


Figure 3

Find2000 can also search all ASCII files in a given directory. This is useful when you store your SAS code in “.sas” files or if your data is externally stored in ASCII files. The same selection box depicted in Figure 3 allows you to select ASCII files.

It may be useful to note that programs written in other languages can also be searched when they are stored as ASCII files. COBOL programs are an example of this. The search strings can be customized to look for keywords in different programming languages.

Find2000 also allows you to search the values of variables in SAS data sets. You can choose to search:

- All variables
- Character variables only
- Numeric variables only
- Selected variables

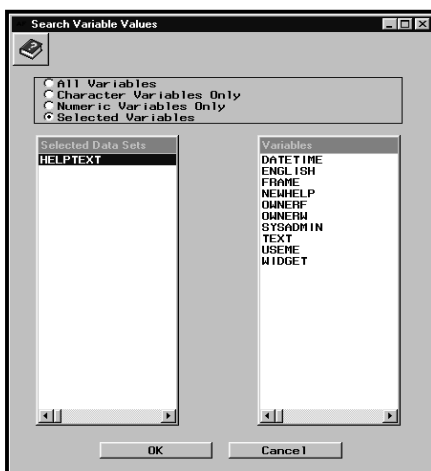


Figure 4

Figure 4 shows an example where the user wants to select specific variables from a data set

Search String Sets

Find2000 allows you to save different sets of search strings as separate files. This allows you to create your own set of search strings and retrieve them for later use. Find2000 provides a default set of search strings geared to finding date-related strings (see Figure 1). The default search set is saved as “Find2000”. This list can be modified to conform to your particular search requirements.

Search strings can be added, deleted, and modified. You can also entirely clear the search string set and start anew. Figure 1 shows this functionality.

Search string sets provides the advantage of using Find2000 for multiple search purposes. For example, companies may find that different projects have different coding standards. This is common for consulting companies where each client may have different coding and naming standards.

Search Techniques

To optimize performance, Find2000 reads in each line twice. The line is input into an both array of one-character items and one text variable of 200 bytes. The INDEX function is used for those search strings that have no wildcards. A more complex search technique using the array is employed when wildcards are encountered.

Input technique:

```
array line{200} $1 line1-line200 ;
length linetext $200;
input (line1-line200) ($char1.) @;
input @1 cc $char1. @1 linetext $200.;
```

Search Results

Find2000 displays the search results in a data table (see Figure 5). Each record in the data table has information about the search strings that are found:

- The found string
- The search string which triggered the find
- The text line in which the search string was found
- The line number at which the search string was found

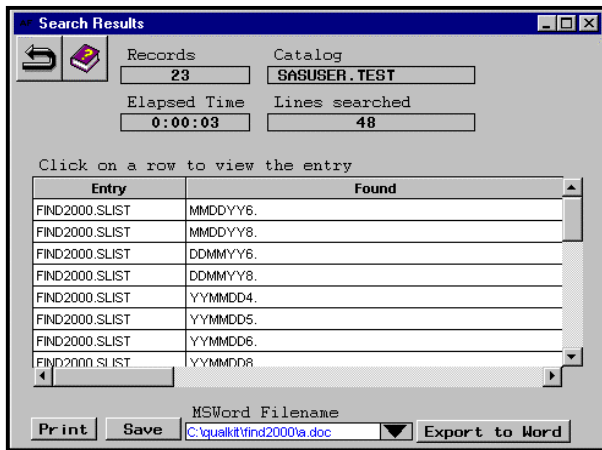


Figure 5

You can easily see the source of the entry in which the search string was found by clicking on a row in the data table. This opens a window to the entry at the precise line the search string was found.

The search results can be:

- Printed to the printer
- Saved as a SAS data set
- Exported to Microsoft Word.

Platform Issues



Find2000 has been tested on Windows 95/NT and UNIX, SAS Version 6.12 and Version 7.

Your data and code may reside on a mainframe running MVS. Find2000 includes a utility that allows you download and upload Partitioned Data Set (PDS) members from an MVS mainframe. (A

SAS/Connect session must be established for this to work.)

Generic Search Capabilities

Find2000 can be used for searches that are not related to year 2000. You can create a number of different search string sets and use each of them for different situations.

One specific other use concerns Europe's conversion to the Euro dollar. A search string set can be created to locate strings related to this.

Context-Sensitive Help



Find2000 provides context-sensitive help. To view help on a screen item, you select the help icon and then select the item on the screen.

Performance

Find2000 performance depends on:

- Number of records searched
- Number and type of search strings
- Computer performance

For the default set of Find2000 search strings, the program processes about 1000 lines per minute on a 233 MHz laptop computer with 64 Meg of RAM.

Searchable Entry Types

Find2000 is available in Standard and Advanced versions. The following is a comparative list of the types of entries that these versions can search:

Search	Standard	Advanced
SCL entries	•	•
SOURCE entries	•	•
PROGRAM entries	•	•
HELP entries	•	•
OUTPUT entries	•	•
LOG entries	•	•
CATAMS entries	•	•
CBT entries	•	•
ASCII files	•	•
SCREEN entries		•
FRAME entries		•
SLIST entries		•
CLASS entries		•
EIS entries		•
Range and Label values of Format and Informat entries		•
Values of SAS data set variables		•
Format and Informat names assigned to SAS data set variables		•

Software Requirements

FIND2000 requires SAS Version 6.12 and runs under Windows 95, NT, and Unix. It requires the following SAS products: SAS/FSP[®], SAS/AF[®], SAS/GRAPH[®].

Conclusion

It is important that business critical applications continue to function as expected come year 2000. Find2000 can help build an inventory of potential

year 2000 problems. The search capability of Find2000, the sources it searches and the search strings that can be specified make it useful for different coding environments.

Find2000 can be downloaded from the Qualex Consulting Services, Inc. web site: www.qlx.com

Author Contact Information

Sarah L. Mitchell
Qualex Consulting Services, Inc.
Phone: 919-380-8284
sarah.mitchell@qlx.com

Michael Gilman
Qualex Consulting Services, Inc.
Phone: 919-380-8284
michael.gilman@qlx.com

SAS® is a registered Trademark or Trademarks of SAS Institute, Inc. in the USA and other countries.
® indicates USA registration.