XI   (F3-F) CREATING INDEX FILES - (an Extended Option for Registered Users)

For all files of information, there is a natural order for processing records in the file. The "natural" order is usually the order in which the information was originally entered. For your family files this is the order of the ID numbers assigned to each name record. In addition, the family files have been designed to permit following lines of ancestry and descendancy and listing information in these orders as well. However you have no doubt found yourself wanting to list your family information in other orders; for instance in alphabetical sequence of last name (surname) and given name, or in birth date sequence.

There are two approaches to writing a program to list information in an "unnatural" order. One is to reorder or "sort" the information each time the listing is produced. For small sets of information, such as one's ancestors on file, this is a reasonable approach. However for larger groups of information, such as for all descendants of a 6th or 7th generation ancestor, or for all relatives of a given individual, this becomes unsatisfactory because of the time required for the "sorting" of the information. The second approach is to sort all records in the file into the desired order, and then build a second file, called an index file, containing nothing but the original ID number of each record stored in the sorted order. This "index" file may then be used to process all or part of the records in the sorted order, and may be re-used many times for different reports. The Family History System has been designed to use the latter approach.

NOTE: *What might be thought of as a "third" approach to the problem of processing records in sorted sequence would be to provide an index file as a part of the family file and maintain it in the proper sequence as the file is being updated. In that case there would be no need to perform a full sort operation to create a usable index file. In FHS, index files are <u>not</u> automatically updated when you make changes to your family file that would affect the order or number of records in the index file. Whenever such changes are made it is necessary to return to this program to Recreate the Index file to reflect the changes before performing any operations that would use the index file.*

The program for creating index files is invoked by selecting Main Menu option F3-F. When you select that option the screen is reformatted as shown at the top of the next page. It follows the same format as most of the report programs with an identification of the current Family File and Printer Setup being used. In addition, it shows the name of the "current" index. The "default" name for the index file uses the same prefix as the .NAM dataset of the family file and a suffix of .NDX. If the index file has been previously created, the number of ID's in the index file will also be shown. If you have added records to your family file since you created the index, the number of ID's in the index file may be different than the number of Names in the Family File.

The lower right portion of the screen shows a list of names for data items that may be used as sequence fields for the index file. They include: Surname, Given name, date and place of Birth, (most recent) Marriage date, date and place of Death, Gender Code and ID number. Next to each field name are three pieces of information describing how the sequence data is to be used. First is a number indicating the order of precedence when more than one sequence field is used. (You will find that you cannot change this for the ID-number field. The ID-number is always the last sequence field for each index.) The second piece of information is an "A" or "D", depending on whether the sorting is to be in "ascending" or "descending" order on that field. (Normal alphabetical or chronological sequence is "ascending".) Finally, the length of the sequence data is given. This "length" is fixed for all sequence fields except surname, given name and place names. Because the use of disk work areas slows the sort procedure when it is not possible to hold all sequence data in the program work area, you are permitted to adjust the length of the name fields to reduce the size of the sequence data and therefore permit sorting more records at a time in memory. (I find that the first 8 characters of the name fields are sufficient to uniquely sort the information in my own family files, which now contain almost 2200 name records.)

FAMILY HISTORY SYSTEM

```
                        Family History System
                  * * * Family File INDEX Program * * *

Family File Setup: _____  _____    # Names:_____
        Index File: _____     Created: _____    # Names:_____
    Printer Setup: _____  _____
            Form: Width:____   Length:___

                      SORT Names: 1)Unchanged 2)CAPS 3)SNDX 4)SNDX+Text
                           Dates: 1)YMD 2)MDY 3)DMY 4)YM 5)MD 6)Y 7)M 8)D
     Select Program Option          +----------------------------------
       F1  Change FILE Setup        | Index Fields      Seq   A/D   Lth
       F2  Change PRINTER Setup     | Surname             1     A    10
       F3  Change Index FIELDs      | Given Name          2     A    15
       F4  Change Report Options    | Birth Date          3     A     4
       F5  CREATE Index File        | Marriage Date       0     A     4
       F6  PRINT Indexed List       | Death Date          0     A     4
                                    | Birth Place         0     A    12
       F9  Return to Main Menu      | Death Place         0     A    12
                                    | Sex Code            0     A     1
                                    | ID Number           4     A     2
```

View 1: Screen Display for Family File Index Program

Above the area for the Index Fields are two lines that indicate different rules that may be applied for sorting text fields and date fields. For *text fields* you may choose to have each text field be sorted just as it was entered into the file or you may request that it be converted to ALL CAPS. The use of capitalization may avoid some inconsistencies in sorting resulting from variations in spelling of some names. You may also alternatively request that text fields will be sorted by Soundex code (SNDX) or using both Soundex code and actual text field. The use of the Soundex code may result in differently spelled, but similar sounding, names being grouped together. *Date fields* can be sorted in Chronological sequence (YMD) or Calendar Sequence (MDY). You may also use partial dates to produce alphabetic listings of individuals grouped by Year, Month, Day, etc. The selection of these options is given in response to a prompting message when the index file is created.

As with other programs in the Family History System, program options F1 and F2 are used to select different Family File and Printer Setups respectively. Program option F4 is used to change option settings for an indexed report that may be produced by the program. Other options will be described in the following sections.

XI.A   (F3) Change Index Fields

Program option F3 allows you to make changes to the "Index Fields" information prior to creating an index file. Use the TAB key to move to the field that you want to change, make the change and, after all changes are complete, press the F1 key to terminate the operation.

The sequence numbers for the fields that you have entered may be adjusted to use standard counting numbers. (For instance if you have sequenced the fields using the numbers "7 2 4" they would be adjusted to "3 1 2".) If you have entered duplicate sequence numbers, the duplicates (after the first occurrence of the number) will be ignored. The "Seq" value for the ID number will be adjusted to reflect its being the last sort field.

XI.B   (F5) Create an Index File

Program option F5 is used to create an index file using the settings for Index Fields that you have established. If Surname is one of the index fields, you will be asked if you want to:

Substitute Husband's Surname for Wife's (Y/N)

This may be appropriate for producing birthday or anniversary lists using the more familiar last name for female relatives. When you reply "Y" to this option, a married woman's surname will be replaced with the surname of her most recent spouse when gathering information for the sort fields. The substitution will *not* be made for women who have the "Surname..Use" field value of "Y" in their name record. (A value of "Y" in the Surname…Use field of a married woman indicates that her own surname is to be used in all reports.) If you choose to make this substitution during the creation of the Index File then other programs that use the index for producing reports will also make the substitution of surnames in the report.

You will next be asked your preference for sorting text fields by the message:

Sort Text: 1) Unchanged   2) in CAPS   3) Soundex   4) Soundex + Text

Your response will be indicated by a hilited value for the "Sort Text" line near the middle of the screen. If you simply press the enter key in response to the prompt, then the previously hilited value for the "Sort Text" option will be used as the default.

You will also be asked if you want to:

Use SORTSEQ Table for Text Character Sequence? (Y/N)

The SORTSEQ system table will provide a more natural ordering of international characters and will eliminate the influence of special characters (for example, "(" and ")") on the sort order. But it will add to the time it takes to prepare an the key fields for sorting.

You will next be asked to identify your preference for sorting dates in response to the message:

Sort Date as: 1) YMD 2) MDY 3) DMY 4) YM 5) MD 6) Y 7) M 8) D

Your response will be indicated by a hilited value for the "Sort Date" line near the middle of the screen. If you simply press the enter key in response to the prompt then the previously hilited vale for the "Sort Date" option will be used as the default.

The sort process then begins. It will consist of several phases. First the information for the sort fields will be read from the family file. If there is insufficient space to hold all of the information in memory, then some information will be written to a temporary disk file. In that case, the sort procedure will involve sorting two or more groups of records (each group is called a "sort segment"). After all groups have been sorted, they will be merged together into a fully sorted sequence. Finally, the Index File will be created by writing out the ID numbers of all records in the final sorted order.

NOTE: *The amount of memory available for holding information for sorting is not based upon the amount of memory that you have in your personal computer. Instead, it is limited to the size of a work area that is provided by Basic. That area holds about 50,000 characters of information.*

When the sort process begins, the viewing area in the lower right portion of the screen will be cleared and reformatted as shown at right. This area will be used to record the progress of the sort procedure. The information shown here includes the number of ID #'s from the family file, the total size of all index fields that are to be sorted, the number of characters of sort information that can be

```
         Sort Progress Report
#ID's =_____  Data (Bytes) =_____
in Memory =_____  on Disk =_____
 #SS=___    SSCT=_____   XSCT=_____

I/O:       IN      Work       OUT
SS=___  _____  _____  _____
Time:  IN    Sort   Work   Merge  OUT

       _____  _____  _____  _____  ____

Begin: _____        Last: _____
```
View 2: Sort Progress Report

held in memory and the number of characters that will be placed into a disk work area. If a disk work area is being used, the #SS field indicates the number of "sort segments" or groups of records that will be written to disk, SSCT indicates the number of ID's in each sort segment and XSCT indicates the number of additional records that will be sorted (a "short" segment of records).

The two lines that begin with the label "I/O" show how many "Input/Output" file operations have taken place. Under "IN" is recorded the number of records that have been read from the family file while gathering the information to be sorted. Under "Work" is recorded the number of records that have been written to the work file during the Input phase, read from the work file and rewritten to the work file during the sort phase, and read from the work file during the Merge

phase. Under "Out" is recorded the number of records that have been written to the Index File. The "SS=" field on the second line shows the number of the sort segment that is being processed (during the sort phase).

The two lines that begin with "Time" record the amount of time that it has taken in each of the phases of the process for creating the Index File.

After the Sort/Merge process is complete you will be asked to:

<div align="center">Enter INDEX File Dataset Name: _____</div>

The default name for the index file will probably be the prefix for the .NAM dataset in your family file, followed by a suffix of .NDX. You may change this as appropriate. You may want to use the default name for an index file which processes records in Surname, Given Name, Birth Date sequence, and perhaps have a BIRTHDAY.NDX file which processes records in Birth Date (MDY), Surname, Given Name sequence.

If a file already exists of the name that you have entered, you will be prompted to confirm that it is okay to Delete the file.

After the creation of the index file is completed, the message:

<div align="center">Index File Successfully Created…</div>

will be written on the bottom line of the screen. The "Begin" and "Last" fields in the "Sort Progress Report" will show the times that the process began and was completed. When you press a key on the keyboard, the message will be cleared and the viewing area in the lower right portion of the screen will be reformatted with information about the sort sequence fields for the index file that was just created.

XI.C (F6) Printing a List of Names in INDEXed Order

Program Option F6 permits you to produce a simple full file listing using the index. You will be prompted to enter the name of the Index File that will be used to produce the report and the destination of the report (Screen, Printer or File). The report is in columnar format and the only information that will be included in the report is that which can be used as an Index Field in an index file. To show where breaks in the primary sort field occur, repeated values for that field are not shown.

Although this report should prove convenient for verifying the results of the index building procedure, more useful indexed listings are produced by other system options. In particular, Main Menu option F2-E (Search/Select/LIST) is the option that I've provided for printing most listings that you are likely to want, such as alphabetic sequenced lists of all or selected records or Birthday or Anniversary lists. That option allows you to include such information as relationship, number of children, places of birth/marriage/death, comments and most recent residence address.