RELATIONSHIP SEARCHES ON FAMILY DATABASES: THEORY AND PRACTICE

It is a relatively simple task to create family group sheets or ancestral pedigree charts from a database of family relationships. Answering queries about the relationship of persons in the database is not easy. This article defines a structure for theoretical consideration of the variety of database searches for relationships. It describes the implementation of a blood relationship query answering system, using dBASEII (trademark of Ashton-Tate Corp.).

Suppose you have a database of your family relationships. What questions can you ask about the relationship of any two individuals? Is Solomon Butson related to Martha Johnston? What is the relationship? Is it a blood relationship? Are they 4th cousins, 6 times removed? Is one a descendant of the other? Who is their common ancestor? Or is it a relationship through marriage? or adoption? Who are the people relating them together? What other relationships, besides the closest relationship, are there between them?

We can organize these questions as follows:

I. Blood Relationships
  A. Closest Relationship
    1. Is X related to Y? Yes or no?
    2. Who is/are their common ancestor(s)?
    3. What is the relationship?
    4. Who are the people linking them together?
  B. Other Relationships (same questions)
II. Non-Blood Relationships
  A. Closest Relationship
    1. Is X related to Y? Yes or no?
    2. What is the relationship?
    3. Who are the people linking them together?
  B. Other Relationships (same questions)

To make this article manageable, we will only consider queries about blood relationships - first, the theory, and then the implementation.

PART I. THE THEORY

Our database has a record for each person. The record identifies the person and gives his or her parents. The identifying field is assumed to be a unique code. The record may have more information, such as the person's name and his or her date and place of birth. So a few typical records might look like this:

Person	Father	Mother	Surname	Given
127	18	1037	Harris	John
128	347	92	Jones	Mabel
129	2	212	Marek	Karel

Is X related to Y?

Our simplest query is: Is Martha Johnston related to Solomon Butson? They can be blood relatives in one of two ways. First, one could be the direct ancestor of the other. That is, Solomon could be Martha's 5th great grandfather. Second, they can be blood relatives if they have one or more common ancestors. Solomon could be Martha's 2nd cousin, 3 times removed.

Since we don't know which kind of relationship they may have, we must search for both types. Here's how:

1. Create 2 sets: one with Martha Johnston in it and one with Solomon Butson in it.
2. Find the records of the parents of each person, and add them to that person's set.
3. Check to see if anyone in Martha Johnston's ancestral set is also in Solomon Butson's ancestral set. If so, quit: the answer is "YES, they are related."
4. Repeat steps 2 and 3 until there are no more parents to be found. If that happens, then they have no known blood relationship.

Let's see how this works. We'll take a very simple case. We ask the computer if X (you) is related to Y (your brother). First, we create an ancestral set for you, X, and one for your brother, Y. There is only one record in each set, yours and his. That's step 1.

Next we use the father and mother codes from your record to find your parents' records. Then we add your parents' records to your ancestral set. And we add your brother's parents' records to his ancestral set. Now there are three records in each set. That's step 2.

Now, we check to see if there is anyone who is in both ancestral sets. If this is your half brother, there will be one such person. Your full brother will have two such people - your parents. That's step 3, and the answer is "Yes, X and Y are related."

Notice that this search is efficient. We stop adding records to the ancestral sets as soon as we know there is a relationship. A less efficient alternative is to find all of the known ancestors of X and Y and see if any match. If each person has a lot of ancestors in the database, the inefficient procedure

might search 50 or 100 records to answer a question that might take the efficient process only 4 searches.

Who are the common ancestors of X and Y?

To see if there is a relationship, we must see if there are any common ancestors. If there are, then we can easily answer the second question. The process is almost identical to the previous search. The only difference is that we print out the common ancestor(s) instead of simply "yes" or "no."

What is the relationship of X to Y?

Different cultures have different names for relationships. For example, Chinese culture distinguishes between "older brother" and "younger brother." European cultures do not make this distinction and refer to both as "brother." We will consider a European system.

What are the relationships? Father, mother, brother, sister, aunt, uncle, cousin, grandfather, grandson, grandmother, 5th great grand-daughter, great aunt, 7th cousin 4 times removed, and on and on. It seems overwhelming. But again a little thinking and organizing give us the solution. First of all, we can standardize things a bit by ignoring sexual distinctions. Instead of "father" or "mother", we will use parent. Instead of "brother" or "sister", we use "sibling", and so on.

The relationship is determined by how many generations the two people are from their common ancestor. We can show this in a table.

X is the ... Of Y.

Y	X = 0	X = 1	X = 2	X = 3
0	Self	Child	Grand Child	1st Great Grand Child
1	Parent	Full/Half Sibling	Niece/ Nephew	Grand Niece/ Nephew
2	Grand Parent	Aunt/ Uncle	1st Cousin	1st Cous, 1 Time Removed
3	1st Great Grand Parent	Grand Aunt/Uncl	1st Cous, 1 Time Removed	2nd Cousin

For example, suppose X is the child of the common ancestor. Then X is 1 generation from the common ancestor. And suppose Y is a great grandchild of the common ancestor - 3 generations away. Using X=1 and Y=3 in the table, we find that X is the Grand Aunt or Uncle of Y.

What patterns can you see in this table? First, there is the obvious symmetry about the diagonal. Corresponding to (X=0,Y=3) "1st Great Grand Parent" is (X=3,Y=0) "1st Great Grand Child." The symmetry reduces the table's complexity almost by half. We also can see four groups in the table. We'll label them A, B, C, and D.

	X=0	X=1	X=2	X=3	X=4
Y=0	A	B	B	C	C
Y=1	B	B	B	B	C
Y=2	B	B	D	D	D
Y=3	C	B	D	D	D
Y=4	C	C	D	D	D

Group B has the greatest variety of relationships. For group C, all we have to do is compute the number of "Great's" to use. For group D, we have to compute the cousin number (i.e., 1st, 2nd, etc.) and the number of times removed. Group A is not of interest. Now we can answer the question: What is the relation of X to Y?

1. Create 2 Sets: One with Martha Johnston in it and one with Solomon Butson in it. To each record add a field for counting the generation number. For Martha and Solomon, give this field a value of 0 (zero).

2. Find the records of the parents of each person, and add them to that person's set. In each record added, include the field for the generation number. Set the value to 1 more than the value of the field for the child.

3. Check to see if anyone in Martha Johnston's ancestral set is also in Solomon Butson's ancestral set. If so, compute and print their relationship and quit.

4. Repeat steps 2 and 3 until there are no more parents to be found. If that happens, then they have no known blood relationship.

Here's how it works. We'll again take the case of you, X, and your brother, Y. First, step 1, we create an ancestral set for you and one for your brother. There is only one record in each set. Each has an added field for the generation counter, with its value set to zero. Next, step 2, we use the father and mother codes from your record to find

your parents' records. We add your parents' records to your ancestral set. And we include the generation field, with a value of 1. Then we add your brother's parents' records to his ancestral set. Now there are three records in each set. Next, step 3, we check to see if there is anyone who is in both ancestral sets. There is. So, we use the generation numbers to generate the title of the realtionship. Relative to both you and your brother, your father's record has generation number '1' in each set. That is, X=1 and Y=1. So, the relationship is "Full or Half Sibling."

Let's take another example. You are X, and your father's sister is Y. We create your set and hers. Then we add both of your parents. Now your set has you, at generation 0, and your parents at generation 1. And your aunt's set has her as generation 0 and her parents as generation 1. But there is no one common to both sets. So, step 4, we add the parents' parents to each set, with generation numbers of 2. Now your set has you (generation 0), your parents (1), and your grandparents (2). Your father's sister's set has her (0), her parents (1), and her grandparents (2). But her parents are your grandparents. So your grandfather is generation 2 in your X set and generation 1 in your aunt's Y set. Using the table for X=2 and Y=1, we see that you, X, are the "Niece/Nephew" of Y.

Who are the people relating X to Y?

This query will be answered by a computer display like this:

Martha Johnston is the 1st cousin, 3 times removed, of Solomon Butson through Jane Keam:

Henry Butson	4	John Butson
Emma Butson	3	Solomon Butson
George Johnston	2	
Walter Johnston	1	
Martha Johnston	0	

Not only must we keep track of the generation numbers. We must also keep track of how each ancestor added to the set is related to the person we are interested in. Here's how:

1. Create 2 sets: one with Martha Johnston in it and one with Solomon Butson in it. To each record add a field for counting the generation number, set to zero. Also add a field for the chain that links to the person under search.

2. Find the records of the parents of each person, and add them to that person's set. Add the proper generation number. Also add the child's identification code to the chain linking the parent to the person under search.

3. Check to see if anyone in Martha Johnston's ancestral set is also in Solomon Butson's ancestral set. If so, compute and print their relationship and relatives on the linking chains, and then quit.

4. Repeat steps 2 and 3 until there are no more parents to be found. If that happens, then they have no known blood relationship.

Here's how it works. Once more, we use you as X and your brother as Y. We create your ancestral set and his. The one record in each set has the added field for the generation counter, with its value set to zero. Each record also has a field for the chain. Its value is blank to start. Next, step 2, we use the father and mother codes from your record to find your parents' records. We add your parents' records to your ancestral set, including the generation field, with a value of 1. We put your identification code into the chain field on your parents' records. Then we add your brother's parents' records to his ancestral set. Finally, we check to see if there is anyone who is in both ancestral sets. Since there is, we then use the generation numbers to generate the title of the relationship. And we use the chains of the common ancestors to tell us what names to print for the linking relatives.

Other Relationships beyond the closest

Once the first, closest, relationship is found, we may want to know if Solomon and Martha are related through any more distant relationships. So we remove the common ancestors from the ancestral sets and simply continue to search. We continue with exactly the same procedure as we were using. The removal of the common ancestors is the only modification.

Part 2 "THE PRACTICE" will appear in the JUL84 issue of GENEALOGICAL COMPUTING. It begins: "We now put to practice the theories of Part 1, using dBASE II... "