# BUILDING A NATIONAL FAMILY TREE

Offer Drori, Ph.D.

SHAAM - Information Systems, P.O. Box 10414, Jerusalem, ISRAEL

offerd@cc.huji.ac.il

My interest in family trees was awakened many years ago, probably as a result of the consciousness of roots in our family. My grandfather Mordechai Salomon was the grandson of Yoel Moshe Salomon, and like some of my relatives I used to draw family trees and trace my past. In 1987, in the course of my work at SHAAM - Information Systems, the need arose to provide users of the service with information on the family trees of the entire population of Israel, for tax purposes. This was a challenge unlike the single-family tree I had dealt with hitherto: here were millions of people, and I needed to think and plan how best to build a family tree on such a scale. The lecture that follows will describe the structure of the national family tree, the way it was planned and built, and also the services obtainable from it.

The information presented in the system is available to its users (workers in tax departments) for their work purposes, but is not open for private research. According to the law (Population Census and the Protection of Privacy), a person is entitled to receive information about himself or herself, but not about other people (apart from name and address). Hence the family tree that was developed in the system is not open to public at large, according to present-day restrictions.

## 1. Introduction

A family tree is a technical, graphic way to portray the system of relationships between people of the same family. We distinguish two main types of family tree:

1) **A Table of Ancestors:** Beside the person in question are recorded the names of his parents, above each parent his or her parents, and so on. It is customary to build such tables as if to the fifth generation with a total of 16 ancestors (8 in the fifth generation above the person in question).

2) **A Table of Descendants:** Below the person in question (and spouse) are recorded their descendants of the first generation

(including spouses), below them are recorded their descendants, and so on. The table is limited only by the researcher's ability to trace and collect information about all the descendants. As a rule, family trees are drawn in a hierarchical form that enables one to see the different generations. Graphically represented, family trees spread outwards to include the various individuals in the family, with different generations shown at different heights. Family trees can include various facts about the individuals that appear in them: the person's name, identity number, indication of death, year of death etc. In most cases it is preferable to display additional details about individuals in a separate framework at the side, in order to preserve the picture of the tree as a whole.

Our national family tree was designed for tax purposes, and according to current tax laws, a person's income and property are assessed on the basis of the family unit to which he belongs. In order to help the system's users (i.e. the tax departments), it was necessary to build a computerized family tree of the entire population of the state which would provides the services required by the user and obviate the need to amass and investigate data on the family connections of millions of people.

## 2. Planning the Family Tree

The users of the system are accustomed to receive their basket of services in computerized form, and it was clear that we must give the same form to the family tree service, especially in view of the very great number of individuals included. We planners faced two central problems in our efforts towards this end:

1. How to construct a suitable mode that would enable us to build the appropriate family connections on the basis of the existing information in the national Population Register.

2. How to display the tree in a clear and easily usable form (since the intended users are not expert genealogists).

### 2.1 Constructing a model of family connections

A national family tree cannot be based on feeding by researchers: the sheer quantity of data and the need for reliability preclude this. It was necessary to build a model that would create the necessary connections on the basis of the information existing in the Population Register. Besides various personal particulars included in the Register, the family particulars include: identity number of the individual, identity number of the spouse, identity numbers of

the father and mother. On the basis of these four particulars alone, we constructed a model that enabled us to build a complete family tree. In the middle of the tree are the parents (on the basis of the identity number of the person in question and his spouse), at the bottom of the tree are their children (on the basis of the father's and mother's identity numbers which are recorded for each child), and at the top of the tree are the grandfather and grandmother (on the basis of the identity numbers of the father and mother of the man in the middle). Beside the man in the middle and his spouse are displayed his brothers, i.e. brothers and sisters born of the same mother and father.

### 3.1 Form of display of the tree

In order to make the display clear and convent for the user, it was decided to make it graphic, not textual. The aim was to display the whole family on one computer screen in a clear, graphic form, without the need to spill over on to other screens, which impair the total picture. One of the problems was how to deal with departures from the norm in family connections, such as divorce, remarriage, etc. These must be dealt with without complicating the information as a whole. We did not want to cram the computer screen with data, and so after a great deal of thought we decided upon a number of principles which in practice determined the way the data were displayed:

1. One computer screen should contain three generations of the family tree.

2. In the middle of the screen should be placed the person in question with his spouse by his side.

3. Dynamic building of a new family tree should be possible for every one of the Individuals on screen.

4. The information displayed on the first screen of the tree should contain a visual display of the connections, surname and first name of the central figure and his Parents, and first names only of the other persons shown in the tree (children and Brothers). In addition there should be an agreed symbol for a deceased person.

5. There should be an easy way to obtain additional information on each individual Tree, at three levels of detail: first, limited level consisting of full name, date of Birth and family status; second level containing full personal details, and a third More detailed level.

6. The tree should be constructed dynamically for every potential size of family, so that It can be organized graphically for a family of 3 children and adjusted for a family of 15 children. In practice, our program incorporated the Table of Ancestors with the Table of Descendants. Because of the dynamism in the creation of the tree, it was possible To obtain at any given moment the required form of display, both of the father's Generations above the person studied and of the descendants` generations, according to need.

## 3. Setting Up the Family Tree

The tree itself was built online by a program specially designed for the purpose. In the course of developing the program we had to find answers to problems connected mainly with the data in the system.

Firstly, not all the data accompanying the identity number of a given person exist in the whole population. For example, the identity number of the father of very old people is not always complete. A lack of this datum influences the ability to give a complete tree: even when the parents appear in the system of the Population Register, is no computer access to them. As explained above, “brothers” of an individual are shown on the basis of the identity number of their common mother and father. In this situation too, when the information is incomplete it is impossible to display a man's brothers, even though he does actually have brothers. Another problem was how to relate to different situations existing within a family, such as divorce and remarriage, etc. We were guided to a solution of this problem by the decision to display a family tree based on a person's biological parents. The decision means that the first display of the tree is based on the person's biological parents even if they are not married to each other at the time of display. Information on their present spouses is of course available to the user.

Development of the program several months. It was released to the users for their purposes at the beginning of 1987.

## 4. The Services Available from the Family Tree System

The system enables the user to locate a person (a resident of the state) and to create for him within seconds a graphic family tree that includes information on three generations: it relates to him, to

his brother(s) and to his spouse as one generation, to his parents as a second generation and to his children as a third generation. The system makes it possible to “stroll” between generations at the touch of a button. The age (i.e. comparative youth) of the State of Israel and its population records means that at present information is unlikely to be available for more than five generations. (This situation will naturally change in the course of time). The system makes it possible to receive a summary of data, three levels of detail, on every factor in the system. The dynamism of the tree enables the program to “stroll” through the family connections of all the residents of the state and to locate distant relationships too: for example the program enables one to locate a certain person and to receive his family tree including his brothers.

By pressing a button the user can create a new tree with one of the brothers in the middle: this new tree displays the brother`s spouse, whence one can navigate to the spouse's family, and so on all graphically on the screen. The ability to navigate within this tree is what makes it possible in practice to obtain a national family tree of all the residents of the state. Population increase in the state is dealt with “transparently” in the system, since the model upon which the system is based exists even when the population grows, and there is no disturbance or need for special treatment.

## 5. Summary

To create a national family tree is a challenge of enormous interest. There is a twofold difference between planning and developing a tree of this kind and creating ordinary family tree systems. Firstly, the challenge of size: in genealogical research a family tree that includes thousands of persons is considered large, whereas we had to deal with millions. Secondly, the information on which the tree is based is not controlled by the user or researcher, hence it was necessary to develop a model suitable for creating family connections on the basis of existing data. From its inception until the present day the system has undergone various improvements, but essentially it continues to provide the services it was designed for with great success. Alongside the program we developed in time an additional auxiliary tool consisting of a shortened family tree. The shortened tree is a “window” of information containing the name of the person investigated and the names of his parents, spouses and children. This auxiliary tool provides family tree service to other information systems that need it.

It is of course possible to “superimpose” our program on the population registers of other countries or on large population groups include the same four basic facts (identity numbers of a person, his spouse, his father and his mother). Thus one can create through the program national family trees for countries with populations far larger than ours.

**Offer Drori**, Ph.D. in computer science. Homepage: http://shum.huji.ac.il/~offerd/, Lecturer at the School of Business Administration, School of Engineering and Computer Science of The Hebrew University of Jerusalem. Head of information database systems, SHAAM - Information Systems. Chairman, Special Interest Group on Text Retrieval Systems (SIGTRS) - http://sigtrs.huji.ac.il

Area of professional engagement and research: Online information systems, system analysis and design, large information databases, information retrieval, human-computer interaction, family trees.

Membership: ACM (Association for Computing Machinery), Israel Association of System Analysts, IPA (Information Processing Association of Israel).