

How to become a better researcher and help others at the same time

*Richard D. Rands, President
Silicon Valley Computer Genealogy Group*

We have a delightful opportunity for you as members of our group to fine tune your record searching skills and at the same time provide a great service to the genealogical world. You may have already heard about the FamilySearch Indexing project that was introduced nearly two years ago at the FGS conference in Salt Lake City. We are now part of that project.

As part of the huge undertaking initiated by the Family History Department of the LDS Church to digitize the enormous collection of microfilmed records in their vaults, they have developed a new methodology for indexing records. This system allows anyone with a computer to participate. So many old records are being transferred to digital format that at the former indexing speed, they wouldn't have been available for searching for several thousand years. So a new method has been introduced that will engage many thousands of new volunteers throughout the world in indexing these important records. The only prerequisites are that participants have a very basic level of computer skills, access to the Internet, and a commitment to spending at least half an hour or so per month doing the indexing.

The concept of indexing may be new to you. The easiest way to explain it is to imagine sitting at an old microfilm reader and cranking through image after image for hours searching for the record of the death of your third great-grandfather, Jeremiah. The process is slow, tedious, eye wrenching, and frustrating. The print is small, faint, and at times illegible. You spend hours going through a single reel of film and don't find a single piece of useful information. Sound familiar?

Now imagine that you can enter Jeremiah's name and basic information into a search engine online, and within a second it points to a digital picture of the precise image of his death record, scanned from the microfilm where Jeremiah is listed. In a short time, you find what you are looking for. You may have already experienced that kind of discovery using the search engines at FamilySearch.org, Ancestry.com, the NEHGS web site, or other websites with indexed records. This amazing feat is only available because those organizations have arranged for a team to scan every entry in their records and transcribe the key data into a computer file that then becomes accessible to the search engines. The key information becomes an index that is searchable by computer programs and contain pointers to the exact image where the information was extracted.

Sometimes these teams are paid for their work, making it very expensive. Often such teams are located in countries where wages are low, but where the ability to read historical records from different cultures is limited. Sometimes teams have

been recruited from prisons. It is not hard to surmise that people with some genealogy experience will produce higher quality indexes.

One might wonder why they have not developed computer programs that are able to scan old documents and automatically build the indexes for searching. The reality is that many millions of dollars have been spent for many decades trying to find a method for interpreting myriad handwriting formats, in dozens of languages, and in a variety of document styles. I recall the amount of technology tried in the 1950's and 1960's to develop what was called "optical character recognition" or OCR. The most effective success that has ever been utilized for OCR is to have a rigid set of numbers and letters that can only be created by a special printer. The numbers across the bottom of your checks are an example of an OCR font. If you have ever used the Graffiti program for entering information into a handheld computer, you will understand the challenge of computer recognition of free style handwriting. The census enumerators in the last century were not expected to fill out their census forms so that a computer would be able to scan them, nor were parish priests in the 1700s and 1800s planning to have their registers digitized and searched.

What is this new methodology for indexing? Record extraction has been going on for decades by members of the LDS Church. How is this different? And how will this improve your searching? How much time will it take?

The new method is very similar to the record extraction system now in use within the LDS Church, but with a few essential differences. First, a new software program must be downloaded onto your computer, with features that make it easier to use and make it more effective for participants, who will not be as closely supervised as LDS extractors have been. The program is written in the programming language JAVA, which makes it more versatile for different types of computers. For the time being, they are focusing on PC computers, but those with Macs and Unix computers will soon be invited to participate.

Second, the program allows the FamilySearch Indexing project server to give each volunteer one small segment of a document at a time, so that it will not take more than about 30 minutes to index each batch.

Finally, the entire process is handled over the Internet, so that you can request a single batch at any time of the day or night; once you have completed indexing the batch, you can return it to the project server at any time. All of these features add up to a powerful means for engaging enough volunteers to reduce significantly the time needed to make the vast collection of digitized documents searchable via the Web.

The Silicon Valley Computer Genealogy Group has been designated as a team by the management of the FamilySearch Indexing project. We have been given permission to add people to the team and provide assistance for our members.

It is our goal to motivate our members to join the project and to commit to index at least one batch per month. Those of us who have been indexing for a month or so have concluded that doing indexing teaches us to use search engines more effectively for locating our family records. We become better at choosing alternate variations of spelling, and better at selecting key variables in the search criteria. When working with census records, we notice patterns of family relationships that are only apparent when working with several pages of an enumeration district. Also, you will increase your ability to read difficult-to-decipher handwriting. Some of our volunteers have called me to say how much fun they are having, and how good they feel about contributing to such a valuable project. I can assure you that you will be a better genealogist if you spend some time on the indexing project.

Please join our team on the FamilySearch Indexing project. If you would like to find out more information, you can check out the web site at:

www.familysearchindexing.org

Although you can volunteer at the web site, we hope you'll register through us instead, particularly because it may take FamilySearch Indexing considerable time to process your registration; if you join our team, I can register you immediately and get you started. By joining our team, you will help us reach our goal of 10,000 entries per month. Please email me the following information so that I can enter your registration as part of the SV-CGG team:

1. A short user name that will be your logon name. It should be something easy to remember and not too long – perhaps 5-8 characters long.
2. Your first and last name
3. Your email address
4. Your telephone number
5. Your postal address

Please send your information to me at rands@earthlink.net. I will send you a confirmation message indicating that you have been registered, along with the instructions for downloading the program and logging on to retrieve a batch for indexing. I will also send you some hints for efficient indexing. Those of you who are able to attend the monthly meetings can attend a training class during the January 13 meeting in Santa Clara, CA. If there is sufficient demand for additional classes later, I will provide them.

We hope you will join us to make genealogy history in this worthwhile project.