BYU Family Historian

Volume 5 Article 3

9-1-2006

Digital Roots of Human Relations: Enabling Technologies for Family History and Genealogical Research

William Barrett

DIGITAL ROOTS OF HUMAN RELATIONS:
ENABLING TECHNOLOGIES FOR FAMILY HISTORY AND GENEALOGICAL RESEARCH

WILLIAM BARRETT

Flowing out of a Computer Science research lab on the third floor of the Talmage Building is a wellspring of enabling technologies for family history and genealogical research. Here, computer science students, working under the direction of Dr. Tom Sederberg and Dr. Bill Barrett are creating software tools to help individuals with their family history research so that people everywhere can seek out their ancestors and perform vital ordinances in their behalf, as desired. These tools include visualization of an entire pedigree on a single (large) sheet of paper (Figure 1), the ability to automatically calculate if and how two or more individuals are related (Figure 3), and a variety of advanced technologies for automatically converting microfilm records into digital, accessible, searchable form over the Internet (Figure 5).

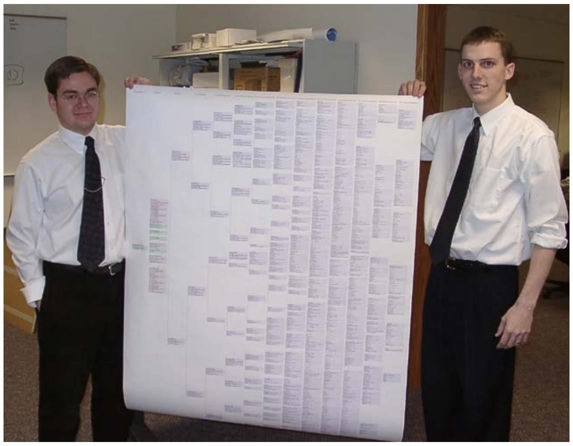

Figure 1. Nicholas North and Josh Jenny show a fourteen-generation pedigree on one page (One Page Genealogy). Conventional layout would require a piece of paper the size of Lavell Edwards football stadium.

OnePage Genealogy

Many people are familiar with a family pedigree displayed on a single 8 1/2” x 11” sheet of paper, which can accommodate at most five generations. But what if you want to display a fourteen-generation pedigree so that you can have a bird’s-eye view of your family tree? If the font size and information remains constant, a little arithmetic tells you that you would need a sheet of paper about 22” high (~ 2 feet) to get six generations on it—four feet for seven generations. Fourteen generations would require a sheet of paper about the length of Lavell Edwards football stadium (Figure 2).

Figure 2. Size of paper required for fourteen-generation pedigree in a conventional layout.

So how do we visualize a fourteen-generation (or larger) pedigree on a single 3 1/2' x 5' sheet of paper as depicted in Figure 1? Using a standard GEDCOM file for the input data, One Page Genealogy exploits the fact that family trees often have incomplete branches—areas of the pedigree that are incomplete. The program automatically optimizes the space by expanding other branches into these areas, while forcing every pair of sub-trees as close to each other as possible. The algorithm used to achieve these results is adaptive to each unique set of records and to each user's preferences. Users can access this software by visiting the OPG Web page found at http://roots.cs.byu.edu/pedigree/. Pedigrees are created in the popular PDF format for portability and shared access. A hard copy can also be ordered through this website for a nominal fee. 100 percent of these proceeds fund students to develop other family history technologies.

Relationship Finder

Family Trees can also be used to determine if, and how, two or more individuals are related. Relationship Finder, a software tool developed by Dr. Tom Sederberg, is accessible to anyone at http://roots.cs.byu.edu/digroots/. Relationship Finder is used to discover relationships between people using Ancestral File data (see http://www.familysearch.org/). Using it, individuals or groups can find out how they are related to Prophets, Apostles, Kings, Queens, as well as anyone else who is entered as part of the group. For example, the following excerpt of a Relationship Finder report shows how I am related to the Prophet Joseph Smith as well as several other people. Joseph Smith's Father's Father's Father's Father's Mother's parents are my Mother's Father's Mother's Mother's Mother's Father's Mother's Father's Mother's Mother's Father's parents. They are T. French and M. Scudamore, who lived in England in the early 1600s. I am a little more closely related to President Boyd K. Packer: His great-great-great-great grandfather, A. Anderson, is my great-great-great-great-great grandfather, who lived in Sweden in the early 1700s. And I am of "Royal Blood," as are many of you.

Relatives of William Barrett *******

Kimball: Becky's FFFFFFFFFFP = your MFFMFFFFFP = R. KIMBALL(1595,England); U. SCOTT(15

Lund: Ryan's MMFFMFFMMFMP = your MFFFFMMFMMP = D. BAXTER(1626,Massachusetts); E. (16

Larsen: Alan's FFMFMFP = your MMFFFP = C. JACOBSEN(1732,Denmark); K. JORGENS

…

*********** Relatives Among the Prophets:

Smith: Joseph's FFFFMP = your MFMMMFMFMMFP = T. FRENCH(1608); M. SCUDAMORE(1612,Eng

Young: Brigham's MFMMMF = your MFFFFMFFFF = F. PEABODY(1613,England)

…

Packer: Boyd's MMFFFF = your MMMFMFF = A. ANDERSSON(1730,Sweden)

…

Kimball: Heber's FFFFFP = your MFFMFFFFFP = R. KIMBALL(1595,England); U. SCOTT(15

…

****

William Barrett, you are of royal lineage. Here are some of your royal ancestors.

****

Henry III (1206-1272) King Of ENGLAND is your 21st greatgrandfather.

Louis VIII "the Lion" (1187-1226) King Of FRANCE is your 22nd greatgrandfather.

Phillipp II (1176-1208) King Of GERMANY is your 23rd greatgrandfather.

Figure 3. Sample report from Relationship Finder showing how William Barrett is related to other people.

Since this report makes use of the Ancestral File, it is only as accurate as the genealogical data in that file.

The website that hosts OnePage Genealogy (http://www.onepagegenealogy.com) and the Relationship Finder (http://roots.cs.byu.edu/digroots/) is owned and maintained within the Computer Science Department at Brigham Young University. Many students who have received their training in our computer science curriculum have been involved in the development of these tools. Tom Finnigan’s software expertise facilitated the “hand-shaking” between the Ancestral File and the Relationship Finder. Jeremy Robertson served as our Webmaster for the “Digital Roots” website and was the primary architect for the Relationship Finder interface. Brian Sanderson is now extending this interface and expanding its scope. Nick North and Josh Jenny implemented the first versions of the desktop application OnePage Genealogy and architected the PDF formatting that enables it to be easily shared, browsed and printed in our lab. Britton Quist and Moriah McClanahan are extending the formatting to allow OnePage Genealogy to be adapted to a more traditional layout and make the “holes” in the family tree more visible.

Many other technologies need to be invented and developed. For example, automatically converting microfilm records into a digital form for searchable access over the Internet requires the solution of several subproblems.

Digital Microfilm

The LDS Church began gathering genealogical records in 1894 and now maintains the world's largest repository of genealogical resources. Among these resources are 2.4 million rolls of microfilm, containing parish, census, and vital records, etc.—a total of three billion pages. The original microfilm rolls are stored in a granite mountain vault for preservation. Copies of these films are available for anyone to view through a network of over 4,000 family history centers worldwide. Development of the technologies for digitization and indexing of this vast collection is one of the major focuses of this lab.

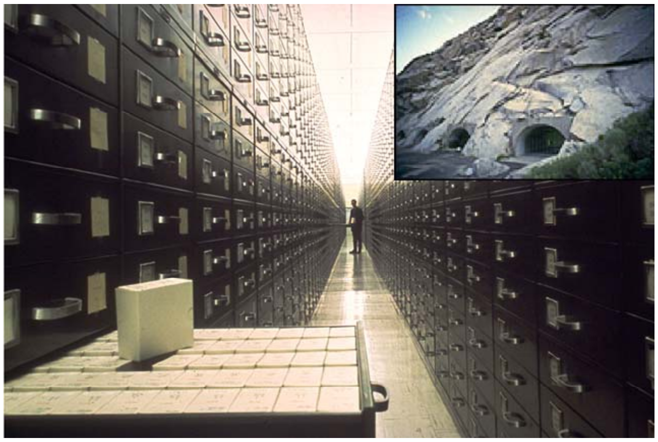

Figure 4. A granite mountain archive houses 2.4 million rolls of microfilm containing genealogical records.

To make full use of digital microfilm, individual frames must be deskewed, scaled, registered, cropped, enhanced, and indexed for random access and browsing. In addition to these geometric and grayscale transformations, the information *within* each frame must be zoned, labeled, and partitioned into fields for more efficient and targeted access across predominantly slower Internet connections. Individual fields containing machine print or handwriting must also be analyzed and recognized, automatically and with human assistance to identify document content. Extracted information (names, dates, places, etc.) must also be linked into a (possibly hybrid) database. Finally, enabling technologies for searching, linking and visualizing genealogical records are needed to assist patrons in finding names. We refer to this process as the Digital Microfilm Pipeline (Figure 5).

Many of these technologies do not exist and must be invented. Over the last ten years, our students have been involved in researching and putting together the puzzle pieces (cropping, enhancing, compressing, zoning, labeling, etc.) of the Digital Microfilm Pipeline as described below.

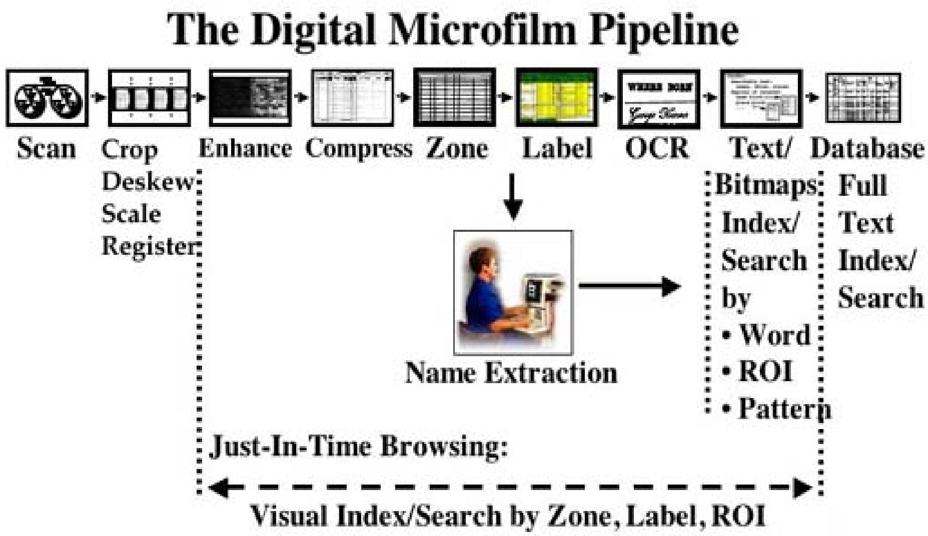

Figure 5. The Digital Microfilm Pipeline is a framework of technologies for converting scanned microfilm into searchable genealogical records.

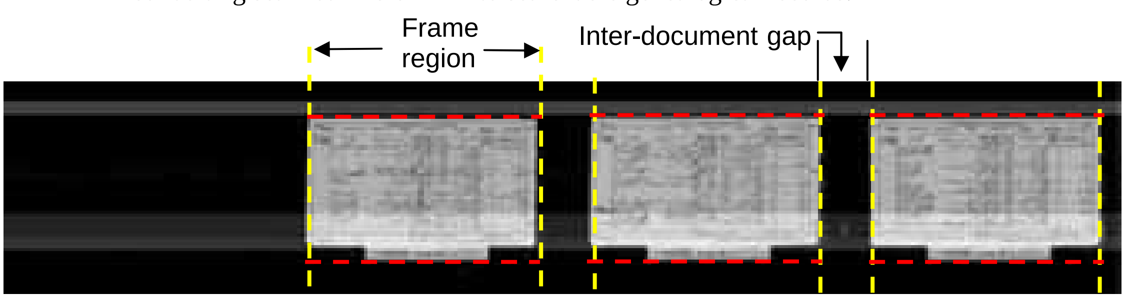

Figure 6. Automated frame cropping of scanned microfilm with stripe removal.

Graduate students Chris Nelson and Mark Pinson have created software to automatically detect and crop individual frames (dashed yellow lines) from scanned microfilm by processing an entire roll at a time. Automating this process greatly increases the accuracy, while dramatically reducing the time for frame cropping.

Figure 7 shows a small section of a tabular document obtained by averaging twenty similar documents, following the application of algorithms for correction for scale, rotation, and translation for each document individually. These algorithms were developed by Luke Hutchison as part of his MS thesis.

Digital image enhancement is used to improve the readability of films that through the acquisition process or other deterioration, are too light, too dark, too noisy, etc. In addition to the correction of grayscale transformation problems, image processing can also be used to selectively enhance patterns, strokes, lines, etc. (Figure 8) beyond what appears in the original document.

As part of his MS thesis, Chris Nelson is currently working on a compression technique that fits parametric curves to the outlines of handwriting, allowing documents to be incrementally downloaded in the order that they were written, revealing content to the user as needed. This will also enhance the novel Just-In-Time Browsing scheme developed by Doug Kennard as part of his MS thesis. Doug is now pursuing his PhD to integrate and extend several critical parts of the Digital Microfilm Pipeline.

As part of his MS thesis, Heath Nielson developed automated Zoning and Labeling algorithms to carve up and index (tabular) documents into meaningful Regions of Interest (Figure 9). This eliminates the time and tedium associated with manual zoning and labeling or downloading the entire (unzoned, unlabeled) document and searching for content sequentially. Extracting the document form also allows us to "cookie cutter" content for family history work.

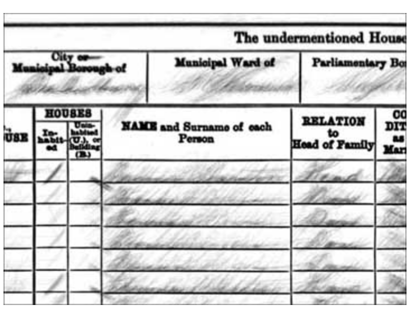

Figure 7. Average of twenty documents after correction of geometric distortion. Clarity of (registered) lines/text demonstrate correction to be highly accurate.

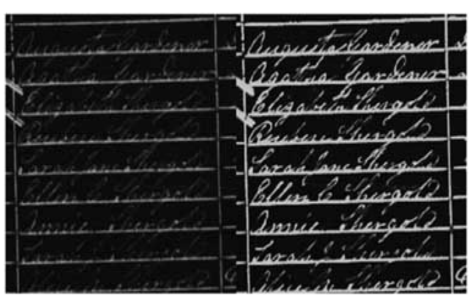

Figure 8. Selective Stroke Enhancement using a custom convolution kernel derived from the handwriting.

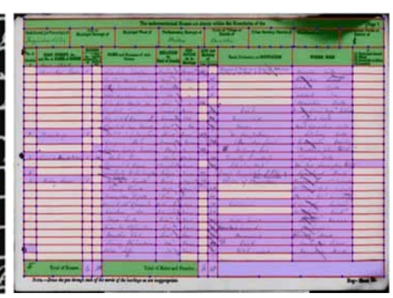

Figure 9. Document Zoning and Labeling. Green = machine printed text. Purple = handwriting. Pink = empty.

Family History Technology Workshop

Technologies such as those presented here are nourished by the Annual Family History Technology Workshop. Sponsored by the Department of Computer Science and promoted by our alumni, this workshop recently concluded its fifth year at BYU, bringing together technology and domain experts from industry, academia, and the LDS Church's Family and Church History Department (http://www.fht.byu.edu/). The workshop was created to provide a venue for the presentation of ground-breaking work in family history technologies and to catalyze interaction between the groups. Our students have been able to present their own work and rub shoulders with professionals from these areas.

This year was the strongest the workshop has ever been, with over 100 computer scientists and genealogists in attendance. The keynote address was given by Ransom Love, director of strategic relationships for the Family and Church History Department. The talk set the stage not only for the workshop, but for years to come, as it portrayed the convergence of technologies and inspired insights, while anchoring them to specific prophetic statements made over the past century.

Back to the Future

It is in the cultivating of inspired technologies that we are able to extend our reach back in time—to see our own roots more clearly, and, through that, to see our own future with deeper purpose and meaning. The flowering of technology has accelerated family history and genealogical research, but it is important to note that there is an ever-present spiritual component and impetus behind this work—even in the bits and bytes and novel algorithms. Certainly because it helps the LDS Church with one of its three major missions, and our students sense that. But they bring their own spirit of enthusiasm and inquiry to the work and are driven from within to make a contribution here. And they are becoming known for their polish and their work in this area. It is truly exciting to work with bright, young students who are providing timely solutions to such a meaningful problem domain.

At the 100th anniversary of the Genealogical Society of Utah, President Howard W. Hunter said, "The Lord has guided the development of information technology and accelerated its role in work for the dead, and will continue to do so. However, we stand only on the threshold of what we can do with these tools. I feel that our most enthusiastic projections can capture only a tiny glimpse of how these tools can help us—and of the eternal consequences of these efforts."

It is this prophetic vision that drives us. So we continue to dig and to plant, anchoring these digital roots in spiritual soil. These are the digital roots of human relations that will harvest hearts and turn them to each other.