Using Visualization and Search to Locate Genealogy Holes

Daniel Zappala
Computer Science Department
Brigham Young University
Provo, UT 84602

ABSTRACT

A primary challenge for people working on their genealogy is where to focus their attention. Even experienced researchers can easily lose track of which ancestors need additional work, unless they keep extremely detailed records. This problem becomes significantly harder when working on a collaborative, online tree because of the volume of information involved and the difficulty in tracking who is doing what. We address this problem using a combination of visualization and search. Visualization guides users through graphical cues to easily spot ancestors that have missing information. Search helps a user to locate genealogy holes among more distant relatives such as siblings and cousins, as well as on collateral lines, that are not easily visualized using charts. We describe a web site we have built to investigate these techniques and describe user experiences with the system.

1. INTRODUCTION

One of the main challenges people face when conducting genealogical research is the task of figuring out where to focus their attention. Once a person's family tree begins to expand, it is difficult to keep track of what work has been done and what information or people are missing. With the advent of collaborative, online genealogy, this problem becomes even harder, because it is common to have large amounts of research already completed for one's ancestors, with varying degrees of quality.

Most genealogy software uses simple pedigree charts and descendant charts to show the discovered ancestors and some of their vital information in a standardize format. However, pedigree charts are limited to showing only direct ancestors and give no indication of how complete the research is for each person. Descendant charts, due to the large amount of information being presented, are often difficult to navigate. Research on visualization has produced new ways of visualizing trees of information, but the focus has generally been on static layout with few visual clues indicating research progress [7]. Significant work has been done on interactive visualizations of genealogical information [2, 5, 8, 1], but with little focus on helping researchers identify holes.

In this paper we describe our recent work to address this problem by using interactive visualizations and a search engine to help researchers find holes in their genealogy. We define a hole as any missing information, missing relatives, or, in the case of LDS ancestors, missing ordinances. We focus strongly on usability and user studies, working directly with both novices and experts to identify and address weaknesses in our visualizations and to refine our search algorithm so that it returns useful results. This builds on our previous work on the Twenty Minute Genealogy project [4], where we developed an interactive ring chart and rudimentary search.

In this paper, we make the following contributions: (1) an interactive fan chart, with improved coloring and annotation to show holes for direct ancestors, (2) an interactive descendant chart, showing holes among the descendants of a common ancestor, and (3) a search engine that can be customized based on the scope and direction that a researcher wants to search. Each of these tools can be customized so that the user specifies what kind of holes she is searching for. We also describe a tagging interface that enables users to keep track of the holes they have found, so that they can conduct research on these ancestors. An application demonstrating these techniques is available at `leaf.byu.edu`.

2. FINDING HOLES

Finding a hole in a large amount of genealogy data is a difficult process for a typical person. Consider Figure 1, which shows a pedigree view from the FamilySearch FamilyTree product. This tree is annotated with dates, gender, and an icon indicating whether temple work has been completed. This visualization helps the user find some holes in their direct line – missing ordinances and a missing birth or death date if the year is not known. However, the missing information is not readily apparent to the novice and the eye is not immediately drawn to this information – the user has to hunt for it. The hunt is even more difficult if the immediate pedigree has no holes – then the user must guess which ancestral line to expand. If there are holes among extended relations (siblings, aunts, uncles) or on collateral lines, the only way for the user to find them is to go through a laborious process of clicking and redrawing. Our discussions with users indicate that many get lost and discouraged in this process. Similar problems plague other online family trees, each with their own small indications of work that can be done (such as Ancestry.com's shaky leaf), but none that have focused strongly on helping users find holes.

Family History Technology Workshop @RootsTech, March 22, 2013, Salt Lake City, Utah, USA.
.

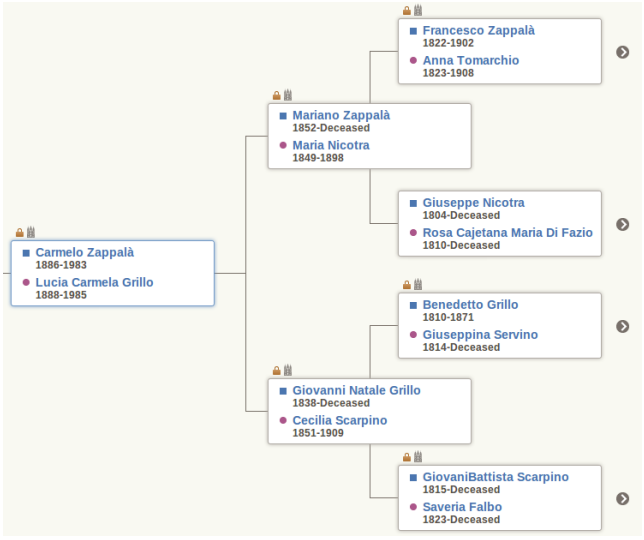

Figure 1: Pedigree view on FamilySearch FamilyTree

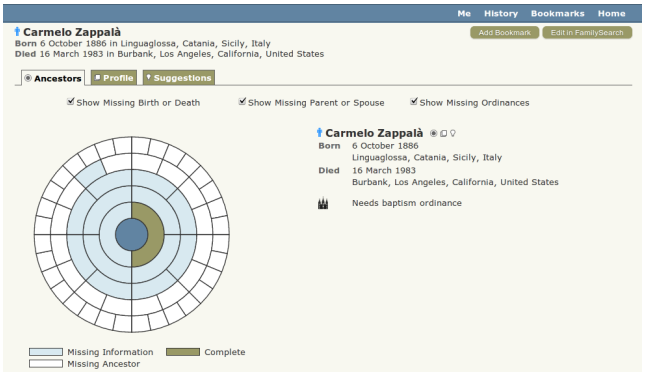

Figure 2: Ancestor view from Twenty Minute Genealogist

We made significant progress addressing these issues through the Twenty Minute Genealogy project at BYU (now part of Kinpoint, Inc.) [3, 4]. Figure 2 shows a ring chart we developed, with six generations displayed at once, and color coding indicating which are missing information, relatives, or ordinances. This visualization is focused on showing holes (blue color) that immediately catch the user's attention. Furthermore, a user can interact with the chart, clicking on a name to see the holes and clicking on the Ancestors button to redraw the chart with the highlighted person at the center. This project also developed a rudimentary search feature that shows holes for individuals in the direct lines, for those who prefer a search-engine style interface.

Our latest work, named *leaf*, revises the ancestor visualization based on user feedback, adds descendant visualization, and provides a more powerful search engine.

2.1 Ancestors

Figure 3 shows the ancestor view from *leaf*, displaying seven generations. The light colors indicate the person has a genealogy hole, and the dark colors indicate no holes were found. As compared to the ring chart, the fan chart provides a more familiar interface that our beta testers indicated was easier to navigate. Fan charts have been around for many

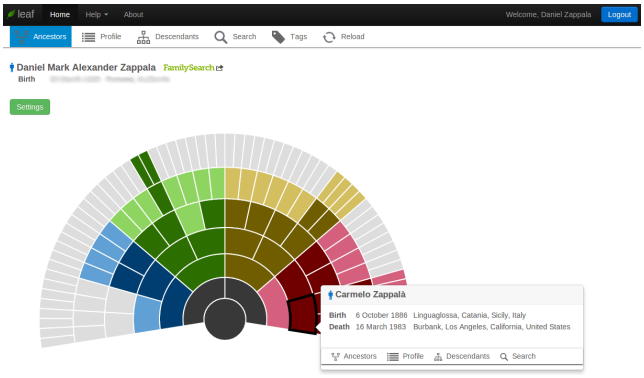

Figure 3: Ancestor view from *leaf*

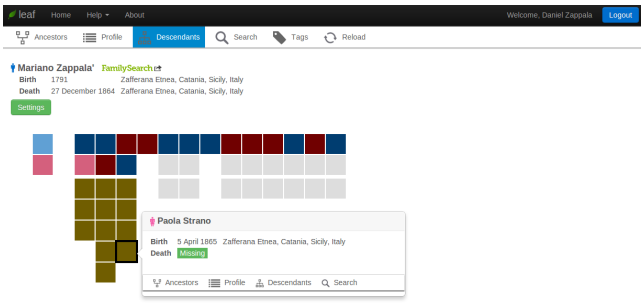

Figure 4: Descendant view from *leaf*

years, so the transition to an interactive fan chart with coloring cues is an easy one to make. As compared to a ring chart, users have an easier time orienting themselves, because paternal lines are always to the right and maternal lines on the left of each generation. In addition, the four primary lines of the chart, one per grandparent, are colored differently, providing additional visual cues for the user [1].

2.2 Descendants

Figure 4 shows the descendant view from *leaf*, displaying three generations at once. The blue and pink squares at the left indicate a married couple. Directly across from them are their children, shown horizontally in the top row, with the second row representing the spouses of their children. Arrayed vertically below each child are the grandchildren, in gold. As shown in Figure 5, the system also shows additional spouses on the same chart, by separating each family with some vertical space.

As with the fan chart, light colors indicate the person has a genealogy hole, and the dark colors indicate no holes were found. Descendants that have not yet been discovered are shown in gray. The system assumes that each person reaching adulthood had a spouse; though this is clearly not always valid it is usually a good indication of an avenue for further research. Likewise, the system assumes that each person had one or more children. The system also assumes that any children dying before adulthood did not marry or have children.

3. SEARCH

[1]This was inspired by recent work at `createfan.com`

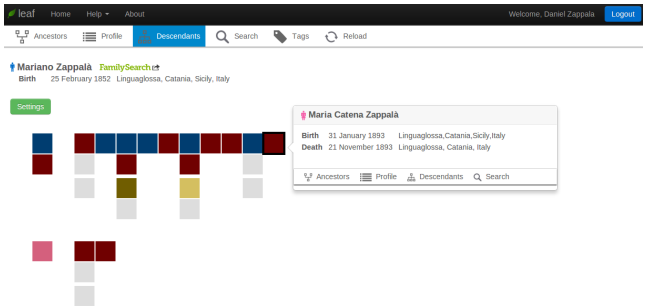

Figure 5: Descendant view with multiple spouses

In some cases, the initial ancestry view shows no pedigree holes. The user can choose to redraw the chart with any of these ancestors as the new center, but for a beginning user this brings them to ancestors so far in the past that research may be difficult for them. One modification we made was to add additional options in the settings for the fan chart, allowing the user to choose to search for single-child families. This option found one couple in the ancestor chart with only one child, indicating an avenue for research that was not otherwise apparent.

A more powerful alternative is the search engine we developed for *leaf*. The user can customize what to search for:

- missing birth date or place,
- missing death date or place,
- missing marriage date or place,
- missing parent,
- missing spouse, or
- missing children (including single children families).

The user can also specify the scope of the search:

- direct line ancestors,
- direct line ancestors plus all siblings,
- descendants, or
- all ancestors, including collateral lines.

With the search engine, the same beta tester discussed above was able to find several genealogy holes that would not otherwise have been visible. For example, one of her direct ancestors had two wives, and the wife not visible in the ancestor chart was missing siblings and parents. A quick census search yielded several hits on this family.

4. FUTURE WORK

Significant future work remains to be done with interactive visualization. The *leaf* site is limited to pulling in genealogical records from FamilySearch, so there can be a delay of 10 to 30 seconds before a chart can be drawn showing multiple generations at a time. Operators of genealogy sites could improve on the user experience by allowing the user to dynamically expand the chart while exploring stored records for her family lines. Visualizing large descendant trees is a challenging problem. The chart we show here has the advantage of showing three generations in a compact form that is easy digest at a glance. Further research is needed to determine if even more generations can be shown without overwhelming the user or requiring too much screen space.

Likewise, additional work is needed on genealogical search engines. Particular areas of research could include refining the search algorithm to find useful results, prioritizing results, and providing multiple, on-demand pages of results as with an online search engine. These techniques are best explored by operators of genealogy sites, where the genealogical data is stored. Current sites primarily focus on searching collections of extracted records, rather than on searching through a user's ancestral database to find genealogical holes.

5. REFERENCES

[1] A. Bezerianos, P. Dragicevic, J.-D. Fekete, Juhee Bae, and B. Watson. Geneaquilts: A system for exploring large genealogies. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1073 –1081, nov.-dec. 2010.

[2] G.M. Draper and R.F. Riesenfeld. Interactive fan charts: A space-saving technique for genealogical graph exploration. In *8th Annual Family History Technology Workshop*, 2008.

[3] Charles D. Knutson and Jonathan Krein. The 20-minute genealogist: A context-preservation metaphor for assisted family history research. In *Family History Technology Workshop*, March 2009.

[4] Charles D. Knutson and Daniel Zappala. The twenty minute genealogist: Assisting family history research through navigation and context preservation. In *Family History Technology Workshop*, March 2012.

[5] M.J. McGuffin and R. Balakrishnan. Interactive visualization of genealogical graphs. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 16 –23, oct. 2005.

[6] Manojit Sarkar and Marc H. Brown. Graphical fisheye views of graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 83–91, 1992.

[7] C. Tuttle, L.G. Nonato, and C. Silva. PedVis: A structured, space-efficient technique for pedigree visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1063 –1072, November 2010.

[8] Janet Wesson, MC du Plessis, and Craig Oosthuizen. A ZoomTree interface for searching genealogical information. In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, AFRIGRAPH '04, pages 131–136, 2004.