The Twenty Minute Genealogist: Assisting Family History Research through Navigation and Context Preservation

Charles D. Knutson
Computer Science Department
Brigham Young University
Provo, UT 84602

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

ABSTRACT

What would software have to do in order to help someone be productive as a family history researcher in as little as 20 minutes per week? Our previously published research suggested that for people who believe that family history work is very important, more than 90% spend essentially no time actually doing anything about it. Our results further suggested that the cause of inaction among the otherwise converted rests with the fundamentally frustrating experience shared by most would-be genealogists – they spend some time, they become frustrated or confused, and they stop. Our research suggested that the predominant inhibitor of research productivity for this group of individuals lies with the overhead involved in navigating pedigree charts and the cognitive burden of preserving a sense of context and research objectives between research sessions.

The Twenty Minute Genealogist is a student-driven project in the Computer Science Department at Brigham Young University that has wrestled with these fundamental challenges since the fall of 2007. Our efforts have yielded the software system described in this paper. The first release of the Twenty Minute Genealogist is a browser-based application that utilizes the FamilySearch API to provide a higher-level interface and search capability for users while drawing on live genealogy data from the NewFamilySearch database. In this paper, we describe this software system and explain how it enables family history researchers to easily navigate the complexities inherent in genealogy work.

1. INTRODUCTION

The Twenty Minute Genealogy project was launched with the goal of answering this fundamental question: *what can software do to help people be productive as family history researchers in as little as 20 minutes per week?* While software has been a boon to genealogy research, we believe current genealogy applications fall short in helping people achieve this particular goal. As a result, this question drives us to develop innovative technologies to bridge this gap. This on-going project is managed and developed primarily by students at Brigham Young University as part of a senior Software Engineering course in the Computer Science Department. Product development and marketing functions have been carried out by students, with Prof. Charles Knutson acting as CEO and Prof. Daniel Zappala acting as CTO.

Over the past five years, we have conducted market research and developed multiple iterations of web and mobile technologies. Our market research suggests that the vast majority of survey respondents who believe it is important to work on their genealogy do almost no work with any regularity, primarily because they don't know how to start or because they perceive that it takes too much time [3]. Informed by these findings, we narrowed our focus to two principle areas where innovation is needed to help the general public be more productive in a relatively short period of time: navigation and context preservation.

Good navigation helps people to quickly determine which of their ancestors they should work on, and the nature of the work that needs to be done for that ancestor. Navigation of family records is traditionally done through pedigree charts and family group sheets. While these tools can be effective for viewing small parts of a family tree or members of an individual family, they fail to reduce the complexity of navigating a large, exponentially expanding graph structure (i.e., a pedigree chart). Traversing one's pedigree by navigating a few generations at a time is not unlike admiring the Grand Canyon while gazing through a straw – it's very difficult to acquire and maintain the big picture. In more pedestrian terms, once a researcher has clicked through several generations of a pedigree, visited children, or moved off to a cousin, it can be difficult to locate where they are in the graph, even within just a few generations. More importantly, traditional navigational metaphors provide at best small clues or hints as to work that needs to be done, and almost no sense of completion for work once finished.

Good context preservation enables a researcher to easily resume a research project without suffering from what we call the "threshold of thrashing" – the significant time that must be invested each week, below which little to no actual work gets done. Our research suggests that a significant level of immersion is required by family history researchers in order to maintain a current and accurate mental model of research objectives, prior accomplishments, and upcoming goals or quests. When time spent each week falls below this threshold, the researcher's mental model suffers, and significant time must be invested to restore the researcher's mental model before she can resume productive research.

Family History Technology Workshop @rootstech, February 3, 2012, Salt Lake City, Utah, USA.

For amateur genealogists with precious little time, we find that with only a few hours each week to dedicate to family history research they typically spend as much or more time remembering why and what they were doing than they do actually searching for their ancestors. As time between sessions rises and available time shrinks, they reach a point where all time spent is context switching with no actual work accomplished. In operating systems design, this phenomenon is known as "thrashing" – a situation in which the amount of time a system spends context switching exceeds the time spent actually doing work. Needless to say, amateur genealogists who suffer through several weeks of thrashing soon find their motivation for family history research waning. In our interviews, we find this phenomenon among busy would-be genealogists to be nearly universal.

Even for experienced genealogical researchers, our interviews suggest that a layoff of as little as two weeks can require a researcher to invest considerable time reconstructing a mental model of their research space. Of course, for experienced researchers with significant time to spend, the cost of reconstructing a mental model is amortized across the ensuing multiple weeks of research immersion, and thus is viewed as a necessary evil, but one that can be overcome. Unfortunately, the pool of experienced genealogical researchers with considerable time to spend each week is a very small set indeed. Our research suggests that the set of family history researchers who have this level of time available accounts for as little as 1-3% of the population of those who claim to care about genealogy.

Effective context preservation must enable a user to quickly jump into a genealogical quest, despite significant time lags and lack of immersion time. While helpful hints, such as a list of research tasks or a set of bookmarks, may provide some assistance, they don't provide a sufficiently rich user experience to enable an amateur researcher with very little available time to retain sufficient context during infrequent forays into their ancestral tree.

We have found that an innovative combination of effective navigational metaphors and context preservation methods can provide the rich experience that amateur genealogists crave. In this paper we describe the software system that has emerged from almost five years of working with senior Computer Science students at Brigham Young University. Our client-facing products include a web application as well iPhone and Android applications. We believe this system represents a major step toward enabling people to instantly navigate the complexities of their family history data, allowing them to focus their efforts on the actual task of genealogical research.

2. USER INTERFACE

Our user interface design is influenced by emerging insights into gamification in software design [5]. The basic idea of gamification is to apply concepts from game design to non-game applications in order to make them more engaging. The Twenty Minute Genealogist employs a "command center" metaphor in conjunction with a task or puzzle completion grid. The user presents FamilySearch credentials through a pop-up window and then is presented the interface shown in Figure 1 with himself or herself in the center of a ring chart.

A ring chart is a well-known technique for displaying genealogical information in a compact form [2, 1, 4]. The

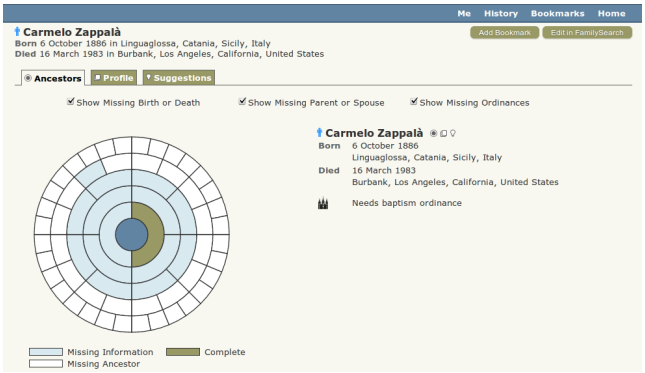

Figure 1: Ancestors page for Carmelo Zappalà

interface we use has a ring chart showing five generations at a glance, comprising 63 individuals. Each cell in the chart is color-coded with white indicating an unknown individual, light blue indicating a known ancestor with missing information, and olive indicating an individual for whom all desired work is completed. Three check boxes allow the user some configuration over the criteria for considering a cell to be complete. One can "Show Missing Birth or Death," "Show Missing Parent or Spouse," or "Show Missing Ordinances." Unchecking all boxes equates the presence of an ancestor with completion of the work for that individual. By checking one or more of the three boxes, the user can determine the completion criteria for individuals in the chart.

One of the effects of this interface is that a user can see within literally one or two seconds where work needs to be done. Viewed from a gaming perspective, the interface invites the user to consider setting off on a quest to find the missing ancestors colored white, or to find more information about those colored light blue. The olive colored cells provide an immediate sense of accomplishment for the user as all cells in a graph are completed over time. Hovering the mouse over the ring chart reveals the names of the individuals in the chart, providing a painless way for the user to quickly build a mental model of the ancestors in the particular chart. Clicking on any non-empty cell in the chart causes a "mini-profile" to be displayed with summary information for that individual. The "mini-profile" includes the person's name, gender, birth and death dates and places (if known), and a description of the work that is missing for that individual.

While the default view for a user is indicated by the "Ancestor" tab, two other tabs provide alternative views for the individual at the center of the chart. The "Profile" view provides detailed information from the FamilySearch database for this particular ancestor, including incomplete fields tagged as "missing" (see Figure 2). The "Suggestions" view provides a list of suggested tasks associated with ancestors of the individual at the center of the chart (see Figure 3).

The "Profile" view of the person at the center of the ring chart can be accessed by clicking the "Profile" tab below that person's name. For any individual whose name appears anywhere in the interface, whether in the ring chart of the "Ancestors" view, the detailed information of the "Profile" view, or the list of individuals in the "Suggestions" view, a "Profile" icon permits the user to make that person the central focus, and move to the "Profile" view for that person. In

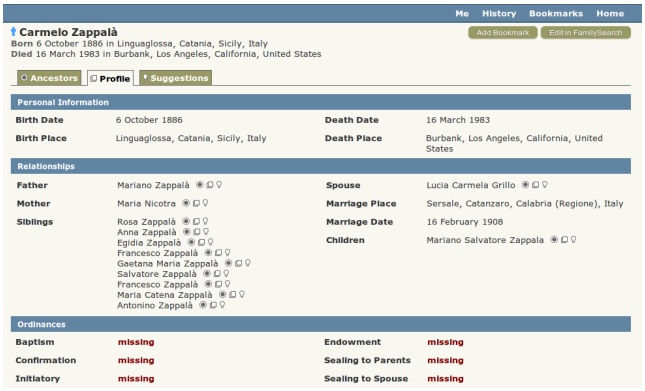

Figure 2: Profile page for Carmelo Zappalà

Figure 3: Suggestions page for Carmelo Zappalà

this manner, the interface permits easy navigation through parents, siblings, or children. A history is maintained for all individuals who have been the focus of the console.

In a similar manner, the "Suggestions" view of the person at the center of the ring chart can be accessed by clicking on the "Suggestions" tab, and a list of suggestions for any individual can be accessed by clicking the "Suggestions" icon for an individual.

These interface components work in harmony to provide a console that provides a user with a big picture of the overall challenge at hand, a grasp of the ancestors not yet discovered, and a comprehension of the specific puzzles yet to be solved for a particular individual. The ring chart of the "Ancestors" view helps the user to manage the sense of accomplishment and zero in on a particular individual of interest. The "Profile" view for an individual identifies specific missing information, each of which can be viewed as a quest to perform or puzzle to solve. Hence a context is established with a very short time investment, and at a very low cognitive investment on the part of the user. The "Suggestions" view is valuable when a user either doesn't want to be bothered at all with the big picture, or simply trusts the system to manage the overall quest by simply suggesting the most proximal ancestor in need of work. For users who prefer the "Suggestions" view as a method of navigating and contextualizing genealogy research, they arrive in their context as quickly as it takes them to select an individual from the list of suggestions and pull up their "Profile" view to identify specific missing information. A user is thus enabled to dive immediately into the actual work of family history research

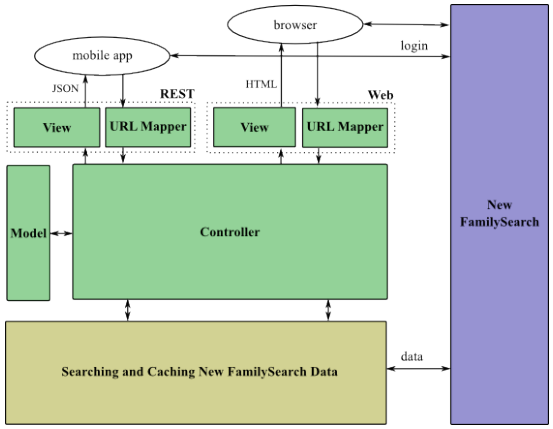

Figure 4: Architecture for Twenty Minute Genealogist

with a contextual latency on the order of tens of seconds. The "threshold of thrashing" experienced by all genealogists is thus effectively eliminated, particularly for the first 63 individuals within six generations of the user.

In its current incarnation, the Twenty Minute Genealogist accesses live data from FamilySearch, but does not provide the user with editing capabilities. Instead, a button enables the user to "Edit in FamilySearch," taking the user to a separate tab where they can use FamilySearch to edit and modify information in their family tree. The data in the Twenty Minute Genealogist application is updated to reflect changes made using New FamilySearch.

3. ARCHITECTURE

Figure 4 shows the architecture of the project. While we have presented the Web-based application for this paper, mobile applications for iOS and Android are currently under development. The bulk of the architecture comprises the server, which we describe below.

The server is built using a Django web framework, employing a standard model-view-controller web development framework, with both REST and web interfaces. The browser uses OAuth to login to FamilySearch whereas mobile applications use HTTP authentication.

One of the main innovations in the project is a module within the server that provides high-level searching and caching of data from the New FamilySearch service. This module takes as input the starting ancestor, the number of generations to search, whether to include children of each family, and a search criteria. The search module then accesses the New FamilySearch API to search the starting ancestor's pedigree, returning a list of all people in the tree, marked up with the information missing based on the search criteria. The current search criteria include:

- *Missing event.* This marks a person who is missing a complete birth or death event. Our web application currently uses the criteria that each person must have either a complete birth event or a complete death event. However, the server also supports the criteria that both events must be present, or can limit itself to just the birth or just the death event. To be complete, the event must include both a date and a place.

- *Missing family.* This marks a person who is missing a parent, spouse, or children. Our web application currently uses the criteria that each person must have both parents and a spouse in order to be complete.

- *Missing ordinances.* This marks a person who is missing any temple ordinances. Our web application currently uses the criteria that each person must include all temple ordinances in order to be complete. However, the server supports searching on individual ordinances or combinations of ordinances, so future versions will allow more targeted searches.

The server also supports a descendancy search and complete traversal of the family graph, but these are not yet used by the web application.

Most genealogy interfaces present the user with a set of names displayed in a pedigree chart or family group sheet, after which the user may request information concerning a specific individual. In contrast, our interface delivers a graphic representation that includes all individuals within five generations of the central person, as well as an immediately discernible summary of work completed versus work yet to be performed for all individuals in the chart. This requirement imposes a significant level of complexity on the search module. Rather than simply requesting names and then fetching detailed information on demand, the search module must perform detailed queries for every individual in the chart even before the chart is displayed to the user.

To accumulate this information, the search module first reads the pedigree of an ancestor to acquire a list of individuals. It then reads the detailed records of those intervals in batches of 10 in a single HTTP request, employing 10 threads at a time. Finally, the module reads the relationship data for the first spouse in each relationship, again using 10 threads at a time. This design provides the high throughput needed to read data from a remote server while also producing a timely visualization for the user.

To determine whether an individual needs additional work requires dealing with a few complexities. First, we must differentiate between deceased persons who lack relevant information, from living persons (who will, per force, lack death information). Following FamilySearch guidelines, we identify an individual without death information as deceased if their birth date was more than 110 years ago, or if their marriage was more than 95 years ago. Conversely, a living individual is one who lacks death information, was born less than 110 years ago or married less than 95 years ago. We therefore refrain from pointing out the lack of spouse, children, or temple ordinances for living persons.

Of course, it's entirely common for a deceased individual to have been born less than 110 years old, and to have never married or had children. At present our system marks this person as missing these family members, but future work will require a mechanism for users to indicate when a deceased person never married or had children.

4. USABILITY TESTING

We have conducted usability testing with approximately 100 beta testers to gain feedback on the interface design and concepts that we have employed. Users have consistently reacted with enthusiasm for the ring chart as a means of rapidly assessing the state of completion of ancestors within their next five generations. Because users can easily redraw the chart with any ancestor at the center, they can focus their attention on the ancestors that they are currently interested in. Users also responded positively to the command console metaphor as a mechanism for managing information at various levels of abstraction, from the high-level graphical view of the ring chart, to the detailed information-rich "Profile" view, to the work-related list of individuals in the "Suggestions" view.

Users have also identified a number of useful features that would enhance the web application. Because the application is so effective at pointing the user to missing information, users immediately want to know where they can go to find the information and how they can add it to their tree. On the one hand, this clearly demonstrates the success of our system, because users are quickly jumping past a major impediment to progress (navigation and context), overcoming the threshold of thrashing, and becoming excited about doing family history research. Unfortunately, they are then running directly into the next logical hurdle – finding relevant genealogical information for a particular ancestor. This is particularly poignant when a user is a relative beginner in genealogy research. Fortunately, links to existing services that solve this problem and provide research guidance (e.g., Ancestry, FamilySearch) fit very well here. Prototype versions of the system have been built that direct a user to products and systems that have strong expertise in these areas of genealogy research.

Though the application currently supports a limited form of bookmarking, we plan to include support for tagging and better integration of notes and tags across the application. We also are developing a "Descendants" tab that will graphically show the descendants of an individual, since many enthusiasts perform this style of research.

Finally, this initial release of the Twenty Minute Genealogist shows only "summary" information from FamilySearch. In reality, data in the New FamilySearch database is often extremely complex, including multiple marriages and potentially large levels of assertions and duplicate records for certain individuals. Future versions of our server and application must accommodate this level of complexity in order to be valuable to more sophisticated genealogists.

5. REFERENCES

[1] G.M. Draper, Y. Livnat, and R.F. Riesenfeld. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5):759 –776, September 2009.

[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 *9th Annual Workshop on Technology for Family History and Genealogical Research*, March 2009.

[4] 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.

[5] Gabe Zichermann. *Gamification by Design.* O'Reilly, August 2011.