A speech interface to genealogical data

Deryle Lonsdale

This talk presents an interface that has been developed to enable users to access genealogical information via speech. Whereas the more traditional approaches for accessing data are prevalent in today's commercial genealogical products (e.g. windows, icons, and point-and-click methods), hands-free access to information is becoming increasingly popular. There would seem to be a potential demand for speech-based access to genealogical information, particularly among enthusiasts who might require mobile access, or those who are uncomfortable with or unable to use manual methods. Yet software to perform this function for genealogical access does not appear to be widely available yet.

To be sure, there are significant issues that have precluded wide-scale deployment of speech applications in various domains. For example, any applications that require a large amount of linguistic processing, or that need to cover a wide vocabulary, are often costly and time-consuming to develop because of the required expertise. Similarly, it is difficult to develop applications versatile enough to adapt to a wide array of user voice characteristics (timbre, pitch, accent, etc.). Most problematic to the area of genealogical processing is the frequently indirect mapping between orthographic representation and pronunciation of proper names, both personal and geographical. All of these issues are relevant to the genealogical domain. Compounding this problem is the fact that many nationalities (and hence linguistic backgrounds) are often represented in a typical user's repository of genealogical information. On the other hand, some of the most successful applications are those that provide access to structured repositories of data. The standardization of genealogical information is well known, and the structure of GEDCOM files makes this type of data easy to access and process.

With these tradeoffs in mind, this paper mentions work done in designing, implementing and testing a speech application designed to enable a user to query information contained in GEDCOM files. The widely used OGI speech toolkit is adopted as the basic underlying system. Various characteristics of the system provide functionality that facilitates deployment of such a system. A survey of the system and the technical aspects of its use for this project are described in the paper.

The system includes a rapid-application development (RAD) kit which is useful in building initial prototypes of novel applications. The kit leverages a Tcl/TK canvas architecture which provides a palette of processing widgets; these widgets represent lower-level object-oriented functionality. On the canvas a developer plots out the overall flowchart of a system's processing, thus designing an interactive dialogue that places a user in contact with the genealogical data. Brief discussion is made of the possible dialogue scenarios that would seem to be most useful in querying a GEDCOM file, and a top-level view of the existing system with respect to its RAD canvas design is given.

In addition to the overall dialogue design, top-level functionality also involves specifying the linguistic properties of the interactive scenario. This includes specifying the syntactic form of acceptable queries by the use of a context-free phrase-structure grammar component, developing a lexicon of appropriate words used in the interactions, and supplying a phonetic transcription (i.e. a phonemicization) of each word. These components form the basic structure of the language model used during the recognition process, and are crucial to the success of the overall system. Mention is made in the talk of the various issues that confront the developer in working on the linguistic components for GEDCOM-related queries.

Middle-level functionality in the system must usually be supplied by data-specific interface routines; this is true for the genealogy application. Since the toolkit is built on the Toolkit Command Language (Tcl), which is specifically designed to provide the "glue" between various applications and data repositories, it was natural to develop Tcl data interface routines for GEDCOM access in this project. Thus, the top-level dialogue system is able to receive and formulate queries from the user, and then send those queries through the Tcl language to a library of Tcl routines that directly access the GEDCOM file for relevant information, which is then sent back to the dialogue level for response generation. Sample Tcl routines developed specifically for GEDCOM access are described, and their integration with the top-level dialogue component is illustrated.

Finally, the lowest level of the system involves the signal and speech processing component. This includes capturing and quantizing the sound signal, mapping it to a series of feature vectors, matching each of these vectors with a (set of) corresponding phoneme(s), and calculating the optimal sequence of phonemes from the set of all possible interpretations. Mention is made of salient properties of the neural network used for vector-to-phoneme mappings and of the hidden Markov model used in computing the optimal phoneme and word sequences. The scoring functions and their contributions to the overall success of the system are also explored, especially as they relate to genealogical data. Varying of dozens of parameters that affect this lower level of processing is possible within the toolkit, and the discussion addresses which settings most affect performance of the overall genealogical dialogue environment.

In addition, the toolkit includes synchronization with an articulatorily-correct animated agent, with which interaction can be carried out. The use of this "talking head" in the system as an interactive partner is demonstrated, and implications are mentioned.

Finally, performance of the system is discussed. Still a system prototype, the system nonetheless exhibits good functionality in a variety of respects; some of these have to do with the nature of genealogical data itself. Sample scenarios are presented, evaluations of accuracy and correctness are presented, and methods for improving the real-time response of the system are given. Scalability is a factor, and preliminary work done on assessing the system for this feature is presented. Current limitations of the system are discussed. Suggestions are made for future work, implications for related development, and how the basic components could be re-used in other toolkits, engines, and applications.