

AI4AV: Building and Testing Machine Learning Methods for Metadata Generation in Audiovisual Collections

Tanya Clement¹, Aaron Choate², Maria Esteva³, Weijia Xu³, Peter Cui, Eugene Castillo, Kewen Wang, Ruizhu Huang

¹Department of English, ²University of Texas Libraries, ³Texas Advanced Computing Center

Introduction

Audiovisual (AV) materials play a fundamental role as historical and scientific records and their number is increasing exponentially in every field. Yet, professionals in libraries and archives tasked with preserving and making these materials accessible at scale often lack the resources to do so. This project aims to:

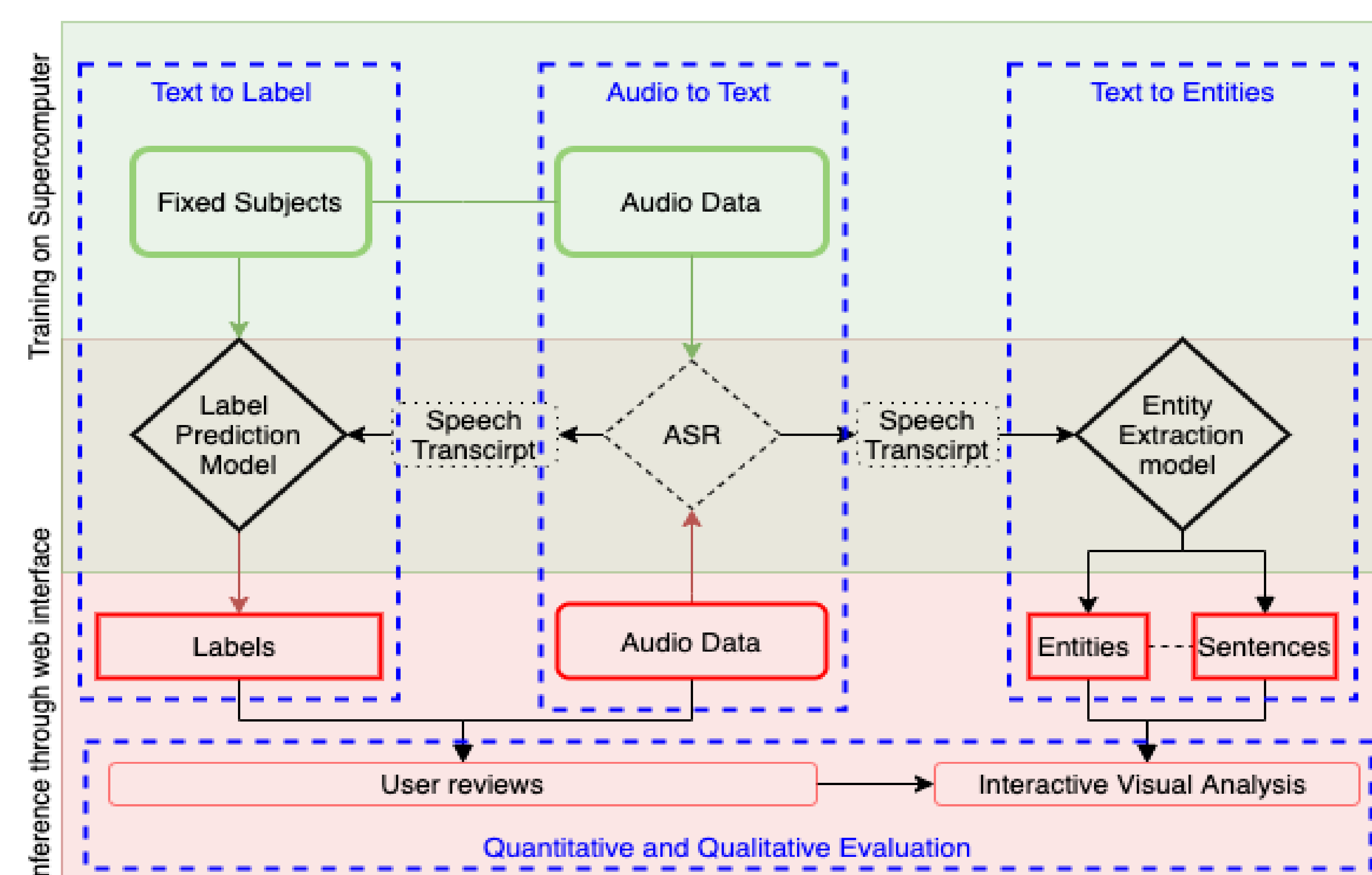
- Develop new ways to procure access to large-scale audio collections.
- Explore Machine Learning (ML) methods to accelerate, improve audio processing, and describe audio data.
- Operationalize and embed ethical values and professional best practices in the ML system.

Research Goal

1. What kinds of technical understanding and human and computer interfaces best facilitate the use of machine learning in LAMs?
2. Can we design workflows for storage and processing incorporating centralized technologies that produce useful results and adhere to LAM standards and values?
3. What are the perceived local and global issues for sustaining these workflows in terms of cost, time, dependencies, and usefulness?

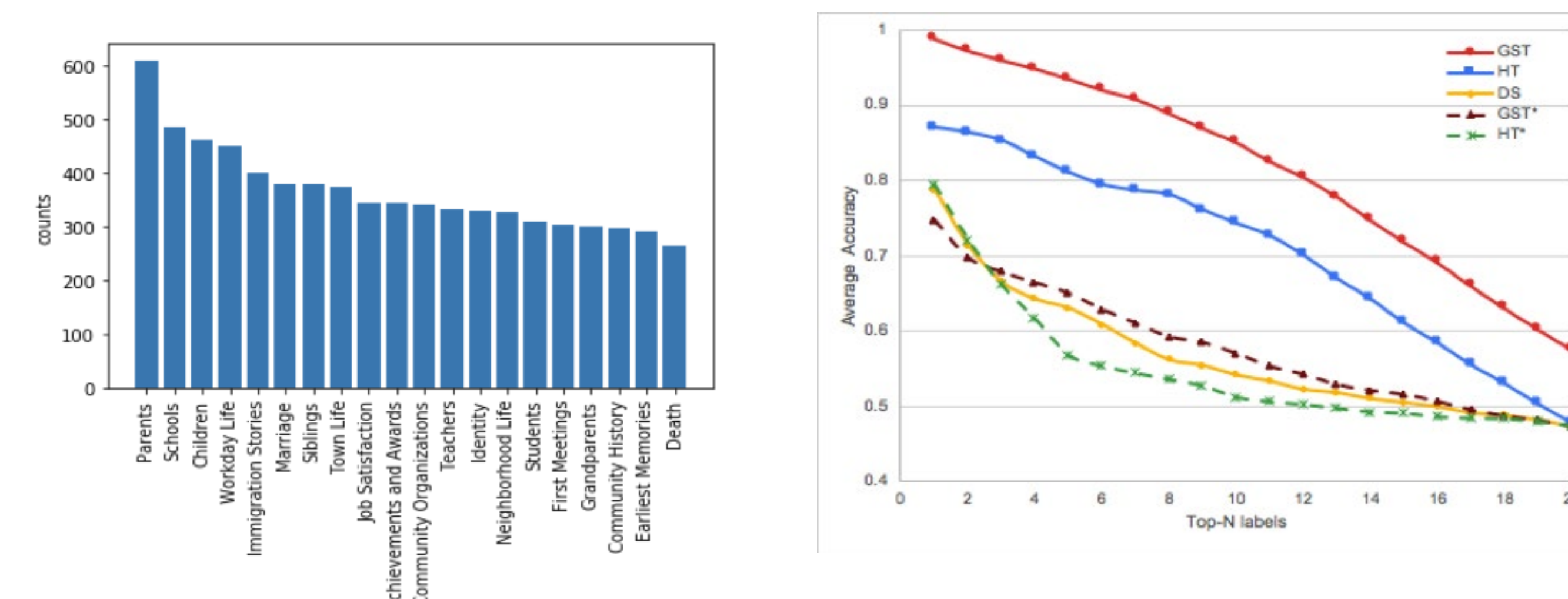
Methods

Project workflow in three connected phases: 1) Audio to Text Transcription, 2) Predicting Labels with Supervised Machine Learning (SML) methods, and 3) Extracting Entities with Unsupervised Machine Learning (UML) methods.

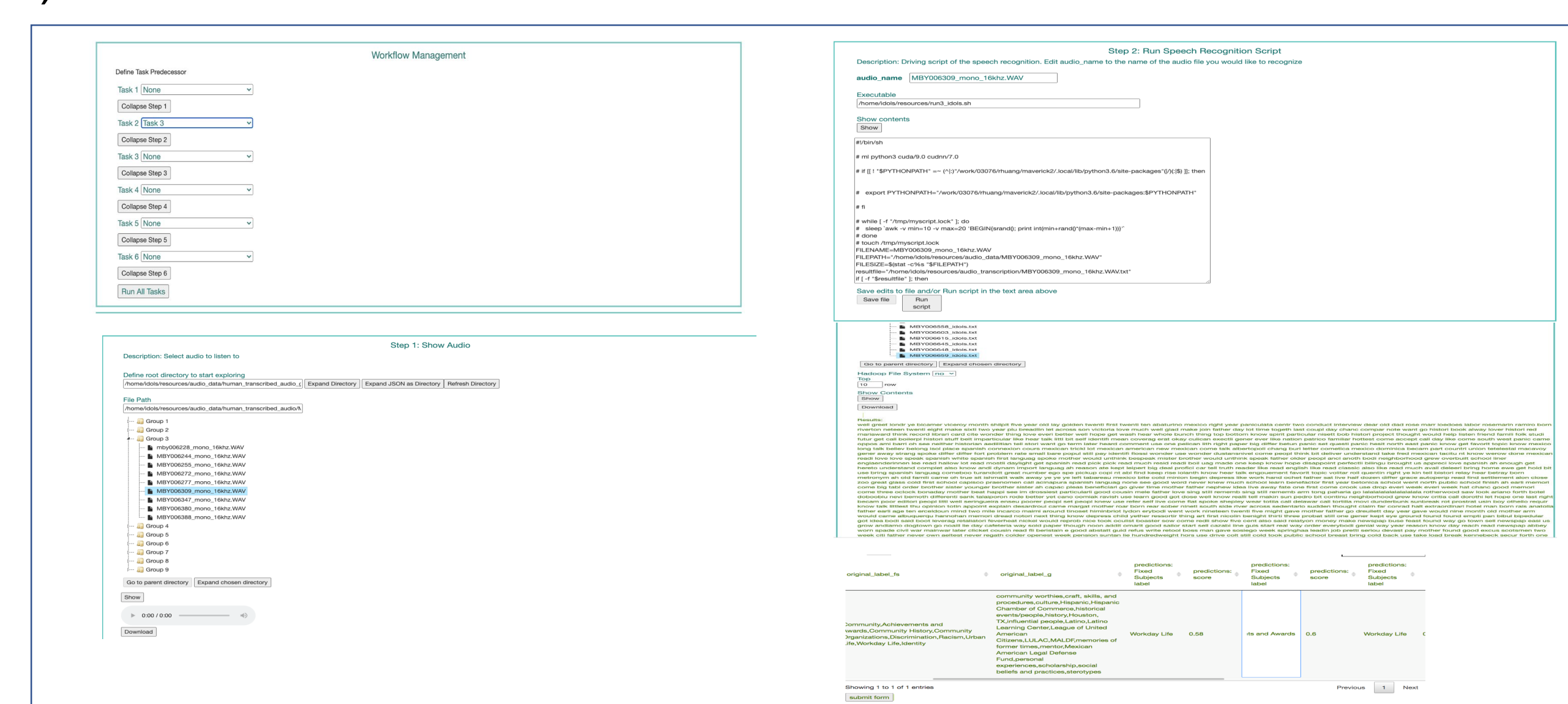


Figures and Results

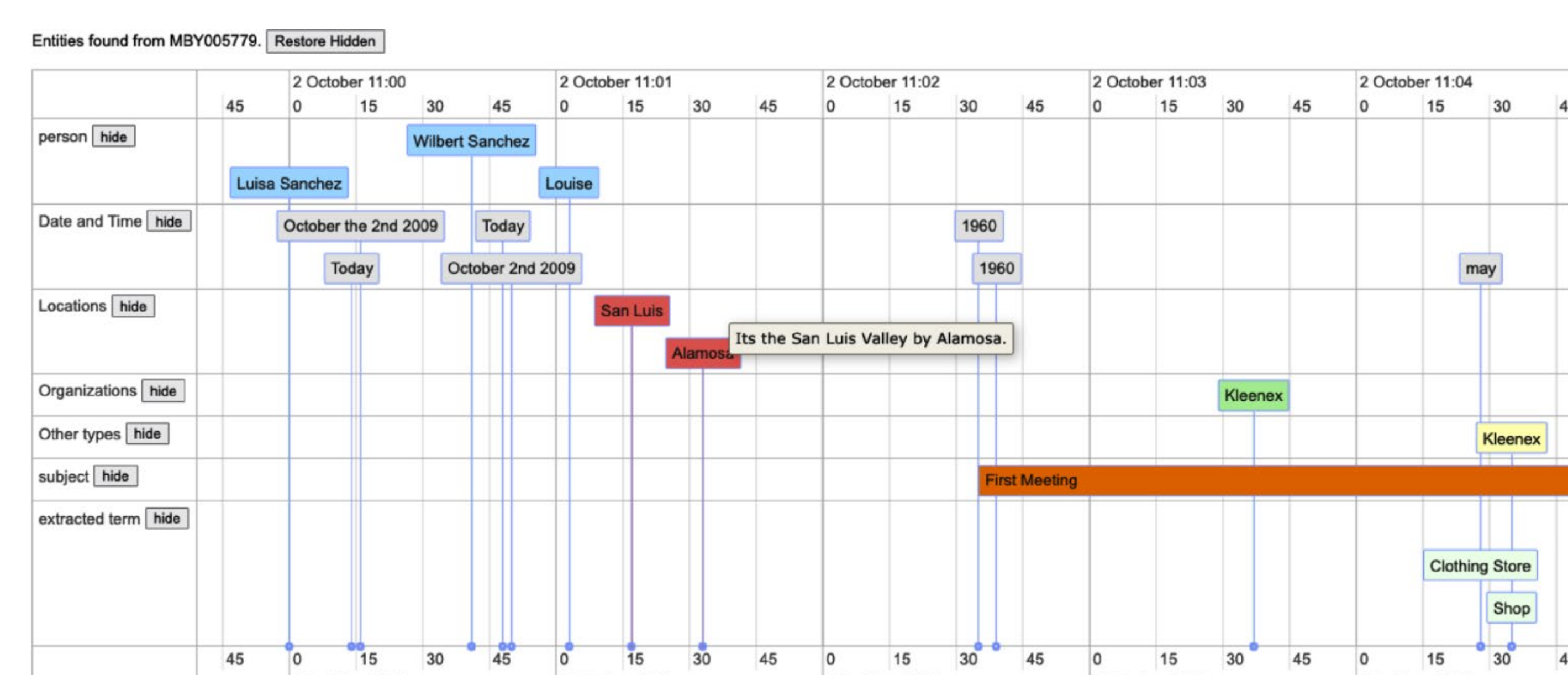
Founded in 2003, StoryCorps deploys trained facilitators—"citizen archivists"—to record and archive short oral histories. Historias focuses on capturing the "diverse stories and life experiences" of the U.S. Latino community (StoryCorps).



Predicted subject headings and keywords for a subset of 815 hour-long oral histories from the StoryCorps Historias collection (left). Comparison of multiple ML methods in prediction of the occurrence of top-20 subject headings assigned by facilitators to describe the Historias recording (right).



A web interface allows, transcribing, processing and classifying audio data files using Machine Learning on supercomputers.



An interactive visualization shows an oral history timeline, the extracted entities organized by type, and the predicted subject headings. A mouse-over reveals the extracted entities in context with the narrative.

Conclusion

This research began with the intention of modelling ML to library descriptive practices. Using existing subject headings to predict labels to describe individual audio files -as it would have been done by librarians- was the first methodological approach. The team also investigated entity extraction to provide more nuanced descriptions. Both methodologies complement each other to represent the content of the audio materials.

A web interface allows librarians to conduct complex process in supercomputers. The interactive visualizations facilitate exploration of the ML results for both librarians and users of audio materials.

Acknowledgments

We thank StoryCorps and the Benson Latin American Collection for providing the data and metadata used in this study. We acknowledge the computational resources provided by TACC. The project was funded by UT Austin Good Systems, and by the National Science Foundation (Award# 1726816),

Publications and Presentations

Tanya Clement, Maria Esteva, and Weijia Xu (2021) Labeling This Thing Called Life: Using Machine Learning to Automatically Generate Metadata for Oral History Collections (under review)

Xu, Weijia, Maria Esteva, Peter Cui, Eugene Castillo, Kewen Wang, Hanna-Robbins Hopkins, Tanya Clement, Aaron Choate, and Ruizhu Huang. "A study of spoken audio processing using machine learning for libraries, archives and museums (LAM)." In 2020 IEEE International Conference on Big Data (Big Data), pp. 1939-1948. IEEE, 2020.

Maria Esteva, Tanya Clement, Weijia Xu, Choate Aaron, Hannah Robbins Hopkins (2020) AI4AV (Artificial Intelligence for Audiovisual): Design and Evaluation of a Shared System for Libraries, Archives and Museums (LAMs) Digital humanity Conference 2020