Browsing Oral History
Making Oral History Accessible on the Web

George Powell | Phil Green | Oscar Torralba

Abstract

Oral History is the audio recording of people’s experiences, lives and memories. It is often in the form of interviews with older people, for example talking about wartime events. At present, there is no centralised digital oral history archive, and the existing archives are either manually indexed or not indexed at all. It is impractical and time-consuming to find information from oral history archives, as they are distributed as lengthy and unannotated digital audio files which can be hours long.

Oral history collections today are underutilised. In this project we use cutting edge speech recognition and information retrieval techniques in conjunction with modern web development practices to build a web application that makes browsing, searching and listening to oral history collections easy. The tool opens the door to providing an open-source centralised digital archive for multiple oral history collections.

Architecture

The three distinct physical systems which interact to solve our requirements. The Web Browser will render HTML, CSS and run JavaScript on the client’s computer. Our Web Server serves this data to the browser and acts as a proxy to the Cloud Services. The Cloud Services are software tools that run on a cloud server, and provide specific functionality to the Web Server when required. This architecture enables fine-grained scalability of resources and high availability. The layers all communicate with each other over HTTP.

Information Retrieval

Interviews are automatically transcribed and transformed into Document Vectors. Each component within a document vector represents the number of occurrences of a particular word within that document. For example, the word “satellite” appears 5 times in the above document, so the corresponding component will be 5. This representation allows them to be searched using information retrieval techniques.

Results

We successfully developed a working web application which drastically improves the accessibility and utility of the Brodsworth Hall oral history collection. We demonstrated that combining WebASR’s automatic transcriptions with text processing and indexing techniques can be an efficient and practical method for achieving information retrieval over oral history spoken document collections.

The tool has gained genuine interest from English Heritage and has the potential for wider adoption by organisations such as the Oral History Society and National Library of Scotland.