

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.

Document

The Moon (Latin: Luna) is Earth's only natural satellite. Although not the largest natural satellite in the Solar System, among the satellites of major planets it is the largest relative to the size of the object it orbits (its primary). It is the second-densest satellite among those whose densities are known (after Jupiter's satellite Io).

Document Vector

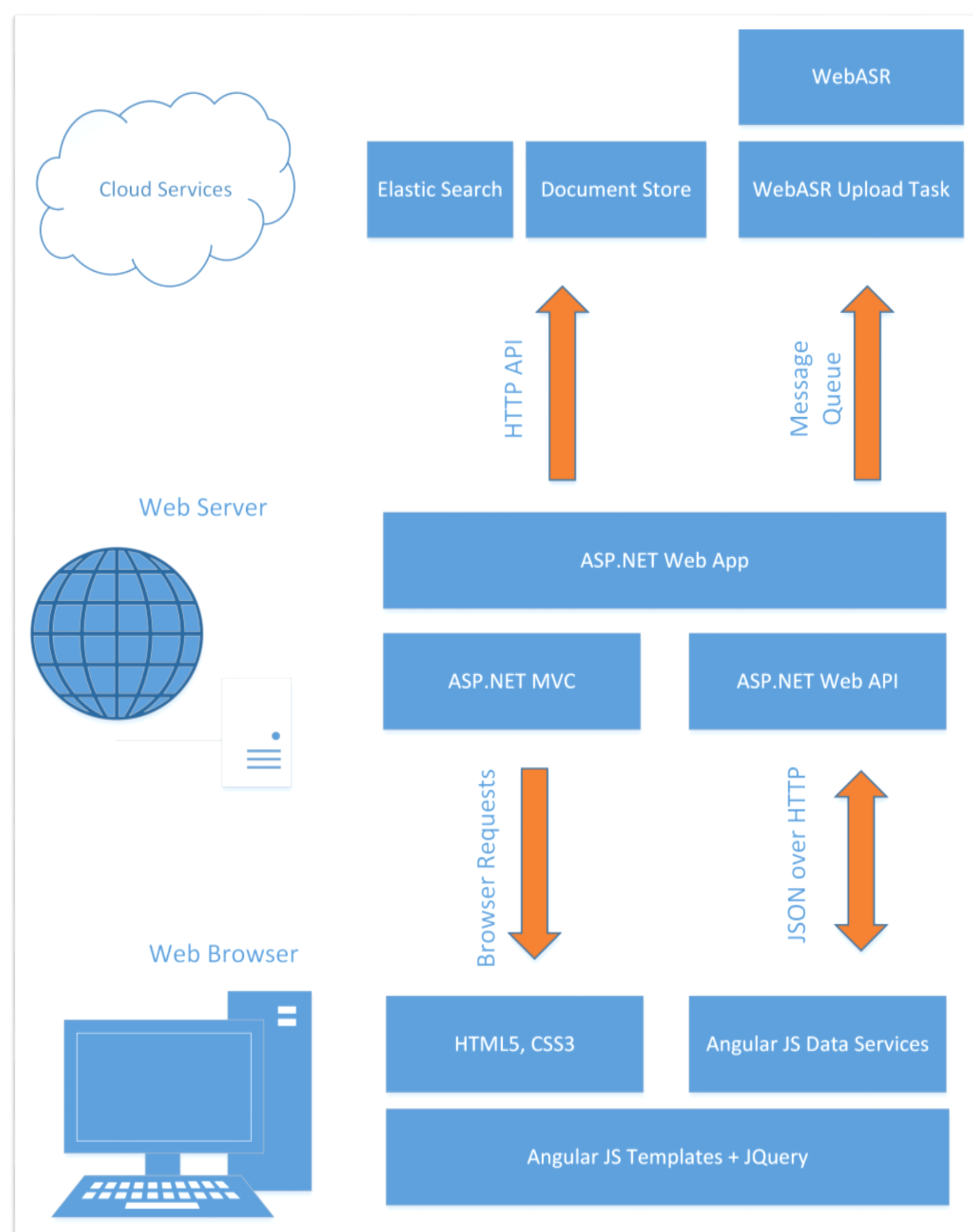
3
5
2
4
...
0
23
1
6

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.

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.



Browsing Oral History

General

Submit

11 results for General

Bill Hutchinson 07-07-2010

Summary

- ... with the place - I don't know I got transferred to **General** Montgomery's headquarters We were stationed...
- ... - Dr ? -but for rank he was given **General** and there was a Major Wilson who was Civil Affairs...
- ... - their names So I was batman/driver to two or three of these people We had **generals** and brigadiers...
- ... moved - a **general** and a brigadier! Horrocks was the name of the requisition at that time - **General**...

David Gilling 19-11-2012

Manual Transcription

- ... think my grandfather had the idea he could repair shoes [**general** talk about shoe and then acquiring...
- ... was a **general** dogsbody ,there were two ambulance drivers one called Albert Le Brun and Albert Woodward...
- ... ,it was better wages at the pit ,it was **generally** better wages at the colliery . V.A. It's just...

Automatic Transcription

- ... the **generation** that you're taking you back what i am not clear about ease those why she might have been calls...
- ... in and out through here through the **generations** presiding too take on crickey i'll never be man...
- ... on the surface of utah and he he was a joke what's **general** dogsbody really erm those two bands...
- ... and when vera **generation** with subtitles guerrilla force includes and work covertly on the ground i...

Edwards Audrey 19-10-2004

Automatic Transcription

- ... into the hall they **generate** the role then your destiny originally there was no gas lay ten...

Summary

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.