# INF 551

**DECIQUEST**

**Theme: A Cloud-based Metadata Extraction & Search Webpage**.

**• Project idea**

**1. A description of the project idea:**

Deciquest is a web application that allows the user to search over a collection of audio files such as all the iconic speeches from the past as well as the present times, podcasts, interviews etc. using the metadata associated with it.

**2. Source of the data:**

We plan to obtain the audio files from online audio repositories such as:

[https://archive.org/](https://urldefense.proofpoint.com/v2/url?u=https-3A__archive.org_&d=DwMFaQ&c=clK7kQUTWtAVEOVIgvi0NU5BOUHhpN0H8p7CSfnc_gI&r=iyM05g593cAt7DXcJdMzIg&m=8X28TWVthfOVxKLVUUX8t4HFvmFZUmuSmpcG7IBFh_k&s=iPsDAlzcSDFJNt39xXz5UHZRObJE242a3tbsdYhvWy8&e=)

[https://www.americanrhetoric.com/top100speechesall.html](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.americanrhetoric.com_top100speechesall.html&d=DwMFaQ&c=clK7kQUTWtAVEOVIgvi0NU5BOUHhpN0H8p7CSfnc_gI&r=iyM05g593cAt7DXcJdMzIg&m=8X28TWVthfOVxKLVUUX8t4HFvmFZUmuSmpcG7IBFh_k&s=jZi4jnV-9QDBQo13Opl79Zc2HKJAkCBVCx8VdUbiO4k&e=)

[http://www.freeinfosociety.com/media\_index.php?type=3&cat=8](https://urldefense.proofpoint.com/v2/url?u=http-3A__www.freeinfosociety.com_media-5Findex.php-3Ftype-3D3-26cat-3D8&d=DwMFaQ&c=clK7kQUTWtAVEOVIgvi0NU5BOUHhpN0H8p7CSfnc_gI&r=iyM05g593cAt7DXcJdMzIg&m=8X28TWVthfOVxKLVUUX8t4HFvmFZUmuSmpcG7IBFh_k&s=t2U6fp2-Dv7yirkqdH3cqwLT2ta71-u6Rtp7FXBBgtg&e=)

**3. Details of the metadata:**

The primary metadata that we want to extract is the name of the artist, original date, file size, duration, genre, copyrights, title, type of the audio file etc. However, we will keep adding to this list of metadata over the duration of this project.

**4. Storing the metadata:**

For the purpose of this project, we intend to use firebase as our database. We will connect this database to our webpage.

**5. Extraction and uploading of extracted metadata to the cloud database:**

We will be using Apache Tika, Pytaglib libraries for the extraction of metadata. In order to upload the extracted metadata to Firebase, we will be writing python scripts.

**6. The user interface:**

We will be providing a web application as the user interface which will be created using HTML, CSS, and JavaScript etc.

**7. Programming languages and software libraries:**

Python, Java, JavaScript, HTML, CSS, Apache Tika library, PyTagLib library, Firebase etc.

**• Group formation:**

**1. Group size:**

Our team comprises of two members.

**2. Group members:**

**i. Name: Mansi Ganatra**

**USC ID: 4669963813**

**Individual responsibilities include**: Programming to extract metadata from source and upload the metadata to Firebase, connecting Firebase to the web browser, design and development of the key word based search etc.

**ii. Name: Radhika Rao Annamraju**

**USC ID- 2800000280**

**Individual responsibilities include**: Data collection from different sources, programming for metadata extraction from the source, design and development of facets, metadata-based search, indexing for the web browser etc.

**• Milestones:**

1. Submission of initial proposal: 9th September

2. Data collection from different sources: 2 weeks (9th September to 23rd September)

3. Programming to extract metadata: 3 weeks (23rd September to 14th October)

4. Programming to upload metadata to the Firebase: 2 weeks (14th October to 28th October)

5. Designing and creating a webpage: 2 weeks (28th October to 4th November)

6. Connecting web app to firebase: 1 week (4th November to 11th November)

7. Testing the developed web application: 1 week (11th November to 18th November)

8. Modification and improvements to the web-page (if needed): 1 week (18th to 25th November)