USC VITERBI SCHOOL OF ENGINEERING

DeciQuest

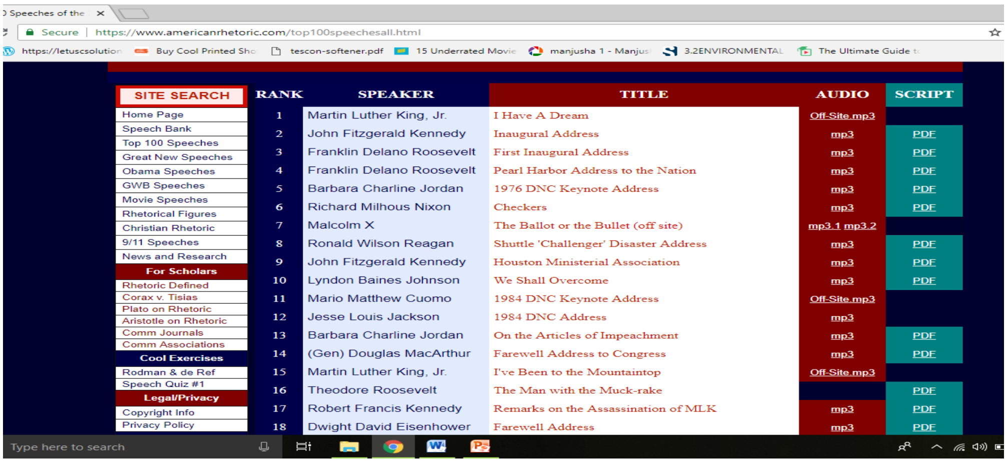
Mansi Ganatra (4669963813)
Radhika Rao Annamraju (2800000280)

1 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. It also allows users to download the audio files and their PDF transcripts.

2 DATA SOURCE

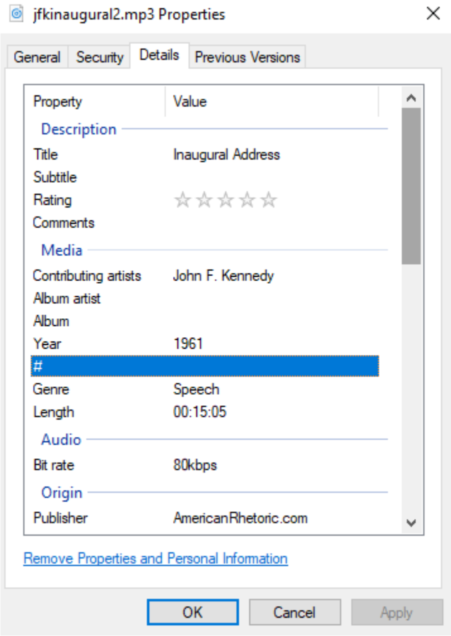
We have obtained the audio files from the online audio repository:
https://www.americanrhetoric.com/top100speechesall.html



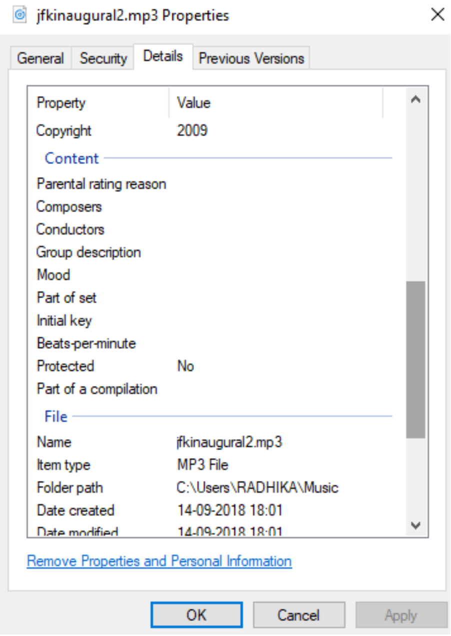
RANK	SPEAKER	TITLE	AUDIO	SCRIPT
1	Martin Luther King, Jr.	I Have A Dream	Off-Site mp3	
2	John Fitzgerald Kennedy	Inaugural Address	mp3	PDF
3	Franklin Delano Roosevelt	First Inaugural Address	mp3	PDF
4	Franklin Delano Roosevelt	Pearl Harbor Address to the Nation	mp3	PDF
5	Barbara Charline Jordan	1976 DNC Keynote Address	mp3	PDF
6	Richard Milhous Nixon	Checkers	mp3	PDF
7	Malcolm X	The Ballot or the Bullet (off site)	mp3.1 mp3.2	
8	Ronald Wilson Reagan	Shuttle 'Challenger' Disaster Address	mp3	PDF
9	John Fitzgerald Kennedy	Houston Ministerial Association	mp3	PDF
10	Lyndon Baines Johnson	We Shall Overcome	mp3	PDF
11	Mario Matthew Cuomo	1984 DNC Keynote Address	Off-Site mp3	
12	Jesse Louis Jackson	1984 DNC Address	mp3	
13	Barbara Charline Jordan	On the Articles of Impeachment	mp3	PDF
14	(Gen) Douglas MacArthur	Farewell Address to Congress	mp3	PDF
15	Martin Luther King, Jr.	I've Been to the Mountaintop	Off-Site mp3	
16	Theodore Roosevelt	The Man with the Muck-rake		PDF
17	Robert Francis Kennedy	Remarks on the Assassination of MLK	mp3	PDF
18	Dwight David Eisenhower	Farewell Address	mp3	PDF

3 Details of Metadata

The primary metadata that we have extracted is the name of the artist, original date, file size, duration, genre, copyrights, title, type of the audio file etc. We also parsed the PDF transcripts of the speeches to obtain the actual contents.


jfkinaugural2.mp3 Properties
General
Security
Details
Previous Versions
Remove Properties and Personal Information
OK
Cancel
Apply

Property	Value
Description	
Title	Inaugural Address
Subtitle	
Rating	☆☆☆☆☆
Comments	
Media	
Contributing artists	John F. Kennedy
Album artist	
Album	
Year	1961
#	
Genre	Speech
Length	00:15:05
Audio	
Bit rate	80kbps
Origin	
Publisher	AmericanRhetoric.com


jfkinaugural2.mp3 Properties
General
Security
Details
Previous Versions
Remove Properties and Personal Information
OK
Cancel
Apply

Property	Value
Copyright	2009
Content	
Parental rating reason	
Composers	
Conductors	
Group description	
Mood	
Part of set	
Initial key	
Beats-per-minute	
Protected	No
Part of a compilation	
File	
Name	jfkinaugural2.mp3
Item type	MP3 File
Folder path	C:\Users\RADHIKA\Music
Date created	14-09-2018 18:01
Date modified	14-09-2018 18:01

4 Project Implementation

4.1 Technologies

Python, JavaScript, Jquery, PyTika, HTML, CSS, Booststrap

4.2 Metadata Extraction

We have used PyTika - Apache Tika library for python to extract the metadata from audio files. Apache Tika also handles extraction of content from PDF - hence it was an ideal choice.

4.3 Keywords Index

We have also created a reverse index for keywords-documents to handle the search functionality of the application.

4.4 Upload to Firebase

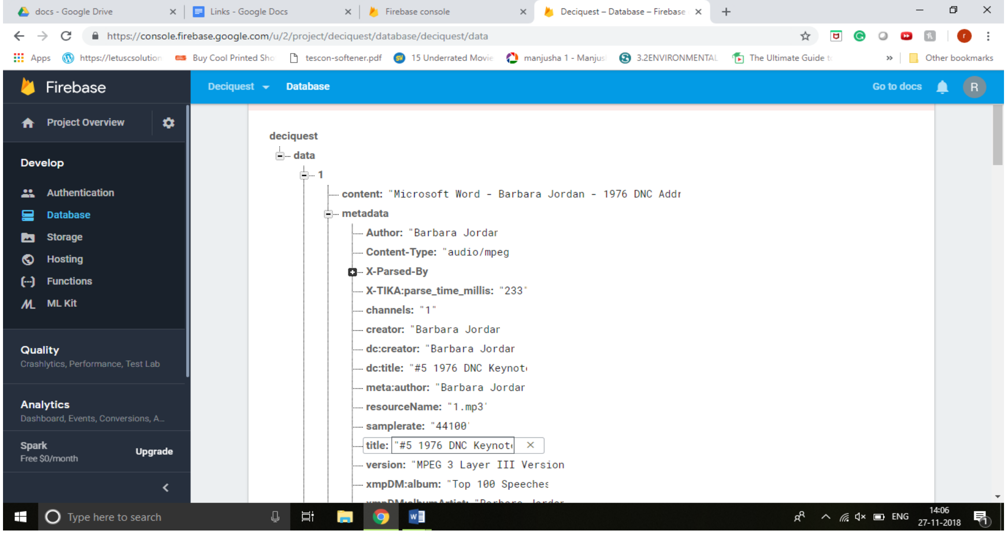
We have written python scripts to upload the metadata and keywords to Firebase.

4.5 Codebase

Our code is present at: https://github.com/mansiganatra/DeciQuest

5 Metadata Storage

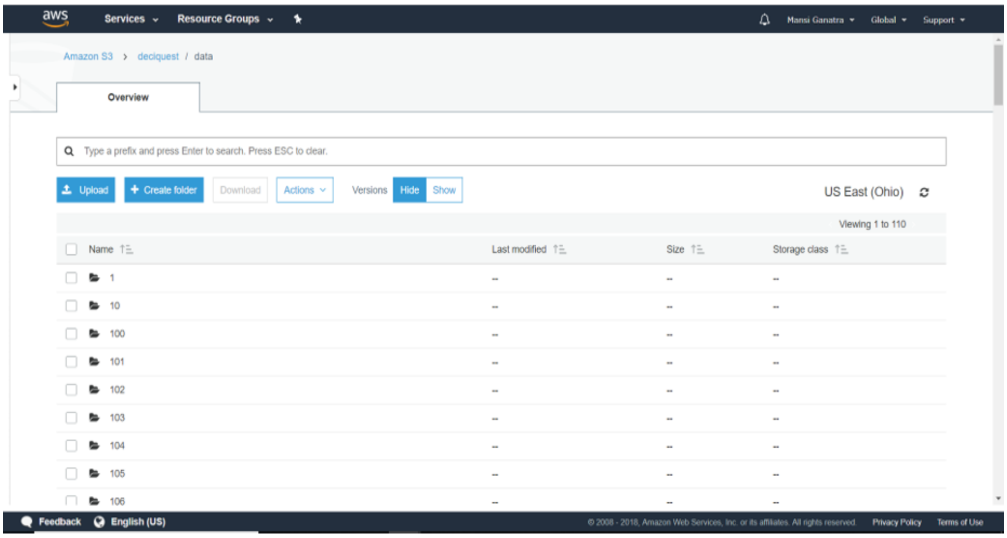
The extracted metadata and Keywords are stores in Firebase at: https://deciquest.firebaseio.com/



```
deciquest
  data
    1
      content: "Microsoft Word - Barbara Jordan - 1976 DNC Addr
      metadata
        Author: "Barbara Jordar
        Content-Type: "audio/mpeg
        X-Parsed-By
        X-TIKA:parse_time_millis: "233"
        channels: "1"
        creator: "Barbara Jordar
        dc:creator: "Barbara Jordar
        dc:title: "#5 1976 DNC Keynot
        meta:author: "Barbara Jordar
        resourceName: "1.mp3'
        samplerate: "44100'
        title: "#5 1976 DNC Keynot
        version: "MPEG 3 Layer III Version
        xmpDM:album: "Top 100 Speeches
```

6 Data Storage

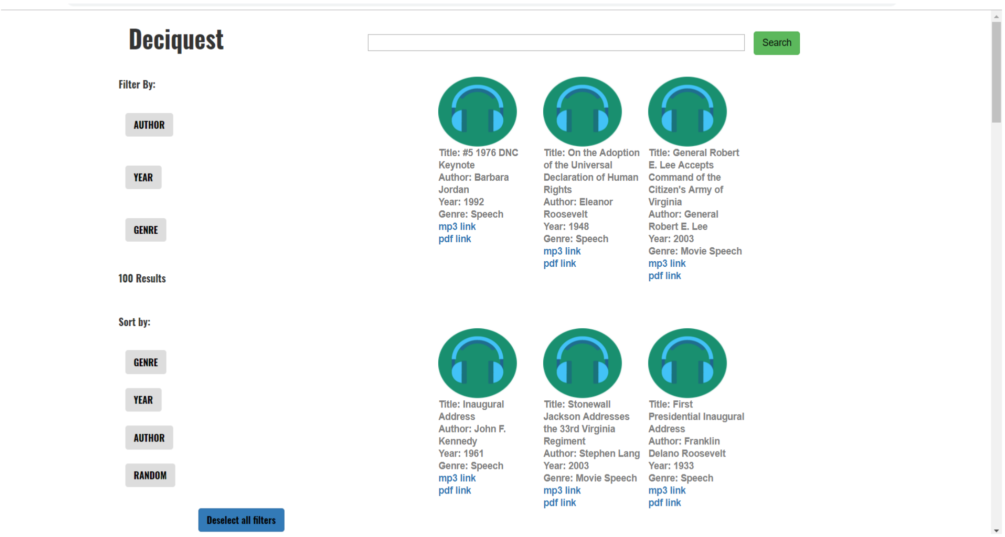
The audio files and the PDF transcripts are stored in Amazon S3 at: http://deciquest.s3.us-east-2.amazonaws.com/



Name	Last modified	Size	Storage class
1	--	--	--
10	--	--	--
100	--	--	--
101	--	--	--
102	--	--	--
103	--	--	--
104	--	--	--
105	--	--	--
106	--	--	--

7 User Interface

The user interface for the application has been developed using Javascript, Jquery, HTML, CSS and Bootstrap. The UI allows the user to filter information using facets of Author, Year and Genre. The user can also sort the data using the same categories. The search bar enables the user to perform keyword search.



Title: #5 1976 DNC Keynote
Author: Barbara Jordan
Year: 1992
Genre: Speech
mp3 link
pdf link

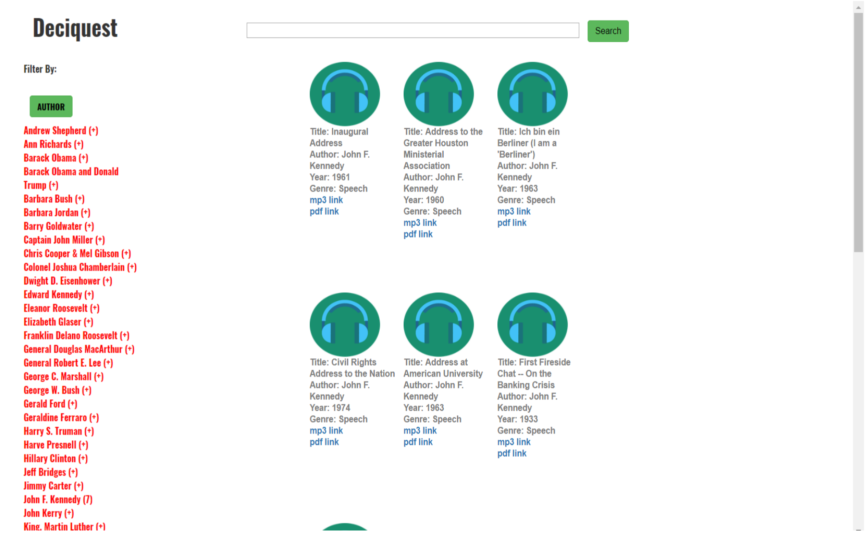
Title: On the Adoption of the Universal Declaration of Human Rights
Author: Eleanor Roosevelt
Year: 1948
Genre: Speech
mp3 link
pdf link

Title: General Robert E. Lee Accepts Command of the Citizen's Army of Virginia
Author: General Robert E. Lee
Year: 2003
Genre: Movie Speech
mp3 link
pdf link

Title: Inaugural Address
Author: John F. Kennedy
Year: 1961
Genre: Speech
mp3 link
pdf link

Title: Stonewall Jackson Addresses the 33rd Virginia Regiment
Author: Stephen Lang
Year: 2003
Genre: Movie Speech
mp3 link
pdf link

Title: First Presidential Inaugural Address
Author: Franklin Delano Roosevelt
Year: 1933
Genre: Speech
mp3 link
pdf link



Andrew Shepherd (+)
Ann Richards (+)
Barack Obama (+)
Barack Obama and Donald Trump (+)
Barbara Bush (+)
Barbara Jordan (+)
Barry Goldwater (+)
Captain John Miller (+)
Chris Cooper & Mel Gibson (+)
Colonel Joshua Chamberlain (+)
Dwight D. Eisenhower (+)
Edward Kennedy (+)
Eleanor Roosevelt (+)
Elizabeth Glaser (+)
Franklin Delano Roosevelt (+)
General Douglas MacArthur (+)
General Robert E. Lee (+)
George C. Marshall (+)
George W. Bush (+)
Gerald Ford (+)
Geraldine Ferraro (+)
Harry S. Truman (+)
Harve Presnell (+)
Hillary Clinton (+)
Jeff Bridges (+)
Jimmy Carter (+)
John F. Kennedy (7)
John Kerry (+)
King, Martin Luther (+)

Title: Inaugural Address
Author: John F. Kennedy
Year: 1961
Genre: Speech
mp3 link
pdf link

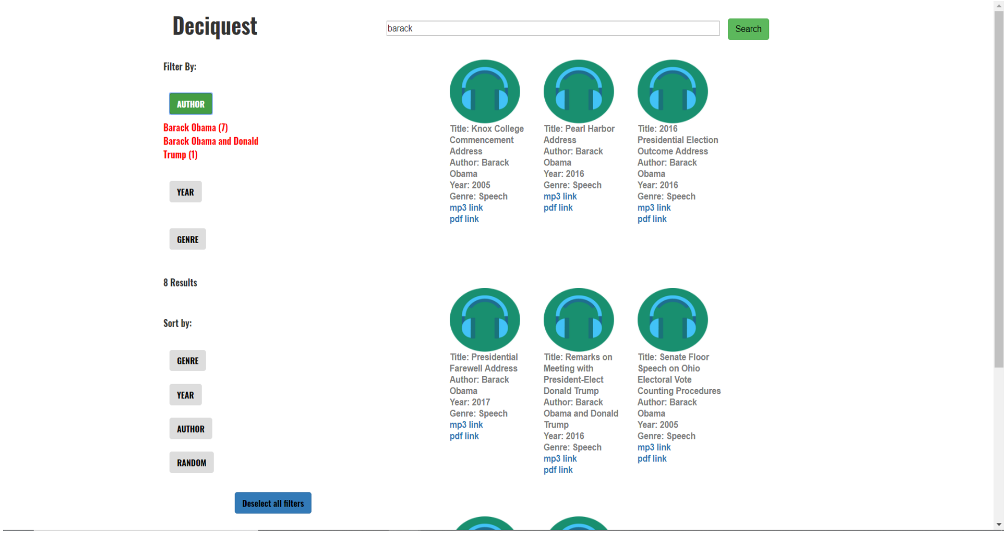
Title: Address to the Greater Houston Ministerial Association
Author: John F. Kennedy
Year: 1960
Genre: Speech
mp3 link
pdf link

Title: Ich bin ein Berliner (I am a 'Berliner')
Author: John F. Kennedy
Year: 1963
Genre: Speech
mp3 link
pdf link

Title: Civil Rights Address to the Nation
Author: John F. Kennedy
Year: 1974
Genre: Speech
mp3 link
pdf link

Title: Address at American University
Author: John F. Kennedy
Year: 1963
Genre: Speech
mp3 link
pdf link

Title: First Fireside Chat -- On the Banking Crisis
Author: John F. Kennedy
Year: 1933
Genre: Speech
mp3 link
pdf link



Barack Obama (7)
Barack Obama and Donald Trump (1)

8 Results

Title: Knox College Commencement Address
Author: Barack Obama
Year: 2005
Genre: Speech
mp3 link
pdf link

Title: Pearl Harbor Address
Author: Barack Obama
Year: 2016
Genre: Speech
mp3 link
pdf link

Title: 2016 Presidential Election Outcome Address
Author: Barack Obama
Year: 2016
Genre: Speech
mp3 link
pdf link

Title: Presidential Farewell Address
Author: Barack Obama
Year: 2017
Genre: Speech
mp3 link
pdf link

Title: Remarks on Meeting with President-Elect Donald Trump
Author: Barack Obama and Donald Trump
Year: 2016
Genre: Speech
mp3 link
pdf link

Title: Senate Floor Speech on Ohio Electoral Vote Counting Procedures
Author: Barack Obama
Year: 2005
Genre: Speech
mp3 link
pdf link

8 Individual Contribution

Radhika Rao Annamraju:

- Data collection from different sources
- Design and development of facets
- Faceted Search from UI
- Keyword Search from UI

Mansi Ganatra:

- Extract metadata and pdf content
- Generate keywords index
- Upload metadata to Firebase
- Upload data to S3