

## **Project Title:** Seeken

**Summary:** Seeken is a web application that helps people seek and save valuable information that is mentioned in a podcast. After watching a podcast, users can input the URL to Seeken to find the information on companies, books, or important people. Users can also save certain pieces of information to maintain their favorite entities from the podcasts they've watched by “favoriting” them.

**Description:** A problem that a lot of people face when listening to podcasts is when proper nouns are mentioned, we tend to want more information on it. Our application will help provide users with information on books, companies and famous people. If a famous person is mentioned a description of who they are is displayed. For companies or books that are mentioned, we will provide a description about them and a link that takes them to a page with more information. We are going to provide descriptions of these entities by web scraping for information on them and then summarize the information using AI. With this app, we will help all podcast watchers who are curious about things that are mentioned in the podcast get more information about whatever they are curious about.

**Good Creative Component:** The creative component that improves the functionality of our application is saving the information on the celebrities, books, and companies mentioned. We are going to introduce another page on the website separate from the original page that gives the information and links that has all the entities the user saves. This new page will be like a “favorite” function that keeps a log of all the entities you are interested in. Entities can be added and removed freely in this log and if the log becomes too big can also be searched directly. This component is crucial to the overall functionality of our application as it helps the user keep track or remember individuals or companies or books they were interested in.

**Usefulness.** This is a problem that we as a group have faced in our everyday lives. Podcasts are an audio-focused media and as such creators don't tend to put a lot of visuals or links to research/articles they mention. We believe that this website will allow people to interact more actively with podcasts and allow visual learners to see important things they may have missed by just listening. Since we want this to be as user friendly as possible we want to make the extension require no setup past installation and users to just have to click the icon to open the log. The main function is whenever an important name, company, website, etc. is mentioned that it'd automatically send an image/link to the log. There aren't really any similar apps to ours, there are some apps that summarize this for text but we want to take it to audio which is a significantly different domain.

**Realness.**

**Celebrity 100:**

<https://www.kaggle.com/datasets/slayomer/forbes-celebrity-100-since-2005>,

**5000 Companies:**

<https://www.kaggle.com/datasets/mysarahmadbhat/inc-5000-companies>,

**Inc 5000:**<https://www.kaggle.com/datasets/mysarahmadbhat/inc-5000-companies>

**Unicorns:**<https://www.kaggle.com/datasets/shubhamoujlayan/all-the-unicorns-in-the-world>

**Top2k:**<https://www.kaggle.com/datasets/joebeachcapital/top-2000-companies-globally>

**Unicorn Companies:**

[https://www.kaggle.com/datasets/mysarahmadbhat/unicorn-companies/data?select=Unicorn\\_Companies.csv](https://www.kaggle.com/datasets/mysarahmadbhat/unicorn-companies/data?select=Unicorn_Companies.csv),

**Famous Birthdays:**

<https://www.kaggle.com/datasets/mexwell/famous-birthdays>

All of the mentioned datasets are from kaggle. They are all csv's. The first dataset contains everyone who has been in the forbes 100 since 2005, their net worth, and some other basic information. It has a cardinality of 1547 and a degree of 3. The second dataset is a dataset of each company of the inc 5000 in 2019 and some basic information about what their company does. It of course has a cardinality of 5000 and a degree of 9. There were a few more datasets that we found but aren't sure if we are going to use. We linked them above just in case we choose to use them later down the road.

**Functionality:** Aside from being able to input a podcast URL and seeing the relevant information, one feature of our web application is that users can save different pieces of information, whether it be a company, name, or person. In a separate page, they can see all the various entities they have saved which allows for creating, deleting, updating, searching, and filtering through their custom database of entities.

**Work distribution:** All members will equally work on frontend and backend tasks so everyone can learn and experience both. All group members will work together on tasks for the users "Favorites" database and user sign in. In terms of the datasets mentioned earlier, Vivek & Ojas will be in charge of the "companies" dataset and Srivanth & Mahi

will be in charge of the “celebrities” dataset. If we end up being able to use four datasets ideally, each group member will be responsible for one dataset. This includes the whole process of setting up and connecting the database to the appropriate frontend points. For frontend, the home page will be led by Vivek & Mahi, while the “Favorites” page will be led by Srivanth & Ojas. For additional information gathering and presentation, web scraping will be led by Ojas & Mahi and AI summaries will be led by Vivek & Srivanth. Note that although the pairs are leading certain tasks for the application, they are not the only ones working on those categories and all members will try to contribute to each category.

UI mockup:

