Publication Portal for Faculty Research at Pepperdine

Abstract

Pepperdine University is in need of a platform that allows students and faculty to easily access information about faculty research. In this project, information provided by the university was used to gather further information about faculty and their publications in order to create a database and dashboards to view the data. Starting with a list of all faculty at Pepperdine and using Semantic Scholar API and Google Sheets Apps Script, the faculty researchers were identified and lists of their publications were generated. These lists and Google Colab were used to generate a database and four different dashboards that display information about the faculty and their research.

Introduction

Currently students and faculty do not have access to a platform that allows them to easily search through professors to discover what professors are researching at Pepperdine University. The goal of this project was to build a database for faculty and students at Pepperdine to be able to search professors and find information about their research with ease. The need for such a database at Pepperdine is evident. To search for what specific professors are researching at Pepperdine, each professor would have to be searched individually on a search engine and look for their publications. With a centralized location for all of this information, a specific token could be searched to see the professors with the most publications on a topic or query a specific professor and have all of their publications listed.

Methods

1. Fetch Data

An excel file that contained a list of all of the faculty at Pepperdine and specific details about each professor was provided at the beginning of the project. From that list we searched a database called Semantic Scholar for further information. Semantic Scholar API is an open source project that provides access to a large database of authors and their publications (cite). Semantic Scholar is free with limited bandwidth and contained most of the faculty researchers at Pepperdine, therefore it was a good fit for the project. Semantic Scholar was able to determine which faculty had done research and provide their Author IDs. The Author IDs provided a means to query further information about each professor as well as information about each of the professor's publications.

2. Form the Database

Another call was made to Semantic Scholar for each author using their Author IDs. A CSV file was generated for each author that contains the Publication ID of all of their publications as well as their titles and other significant information. The data from the author CSV files was combined with that from the original author file given to create a database in Google Colab. Google Colab is an online notebook that allows anyone to produce code on their browser.(cite) This resource was particularly useful as Google Drive can be mounted to Google Colab, then is easily accessed and used while running code in the browser.

3. Create Dashboards

The Dashboards were designed in Google Colab using the python libraries Plotly and Jupyter Dash. Plotly allows users to create detailed and interactive diagrams that can be displayed inline in Google Colab.(cite) It may also be used in Python-built web applications that use Dash. The data collected was connected to Google Colab using Google Drive. Then, we imported Plotly and Jupyter Dash in order to create inline applications.

Results

Describe dashboards
1. Author Search
   a. The 'Author Search' dashboard consisted of a dropdown menu and a table that was updated with each selection of an author. In the dropdown, the user is able to either scroll through the list to find the professor they are looking for or type in their name. For the selected professor, the table lists for each of their publications, the title, year of publication, topic, and coauthors.
2. Number of Publications per Author:
   a. The 'Number of Publications per Author' dashboard displays a dropdown list that allows the the user to choose or search for an author. For the selected author, a table of citationCount, HIndex, and number of publications is generated while a histogram with number of papers per year is presented.
3. Rankings with Fields of Study and departments
   a. The 'Rankings with Fields of Study' and the 'Rankings with Departments' dashboards contains a dropdown list of either fields of study that our professors' works are published on or the department at Pepperdine they are working in. By selecting a field of study/department, a table of rankings is returned and populated with professors' name, department/field of study, H-Index, citation count, and paper count. It is an interactive table; by clicking the arrows near each label the users are able to sort data in decreasing or increasing order for each column.

Discussion
1. Despite being successful in creating dashboards that are user-friendly, there were a few setbacks in their generation. First, the search for faculty members by names does not always return the correct Pepperdine faculty because there are same names in the database. In addition, some papers may not be recorded in the database. In our example, Semantic Scholar has researchers with duplicate names, and it does not include all of the faculty researchers. If the researcher is a Pepperdine faculty member, they had to be screened individually to confirm their identity. One possible solution to this problem is to do a pre-query to faculty, allowing themselves to validate which profiles match them. Also, another possible aid to the project would be to utilize an API that is better fit to Pepperdine faculty.
2. This project culminated in the creation of four dashboards as inline Apps. However, Plotly and Jupyter Dash have the capability to assemble these dashboards to a single website where tables, graphs and other widgets can interact with one another. In addition, Plotly and Jupyter Dash would allow the users to build individual dash apps and place them in selected locations within existing websites. By implementing the dashboards on a website, this project would be even more user friendly.

Conclusion

The purpose of this project is to create a platform that allows faculty and students at Pepperdine may easily search for and locate information about faculty researchers. We utilized Semantic Scholar Academic Graph API to fetch our data. We then, using Google Apps Script, generated CSV files for each faculty researcher that included important details about their

publications. Finally, the dashboards were created in Google Colab using Plotly and Jupyter Dash. Four dashboards were developed to present aggregated data of over 400 faculty researchers and 15,000 publications.

References

1. "Google Colab - What is Google Colab?," *Google colab - what is google colab?* [Online]. Available: https://www.tutorialspoint.com/google_colab/what_is_google_colab.htm. [Accessed: 23-Apr-2022].

2. Plotly, "Dash documentation & user guide," *Plotly*. [Online]. Available: https://dash.plotly.com/. [Accessed: 23-Apr-2022].

3. Google, *Google apps script overview | google developers*. [Online]. Available: https://developers.google.com/apps-script/overview. [Accessed: 23-Apr-2022].