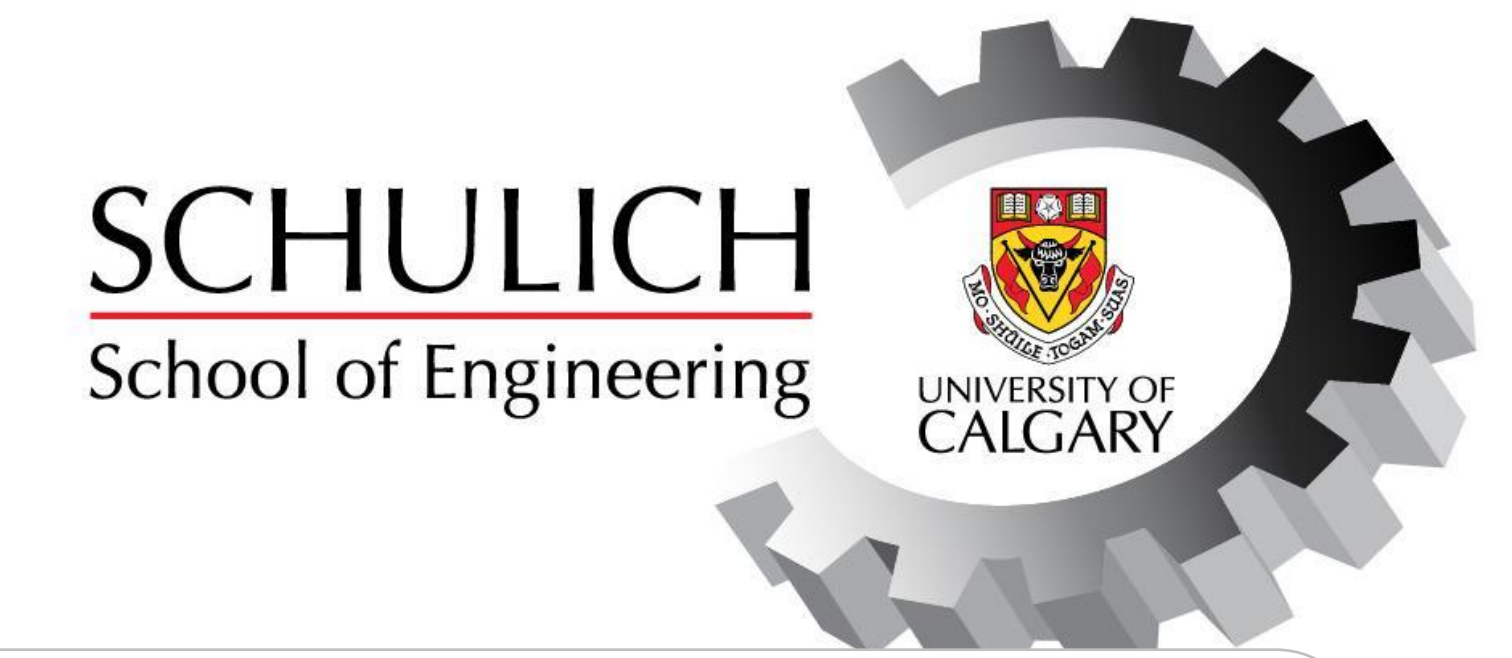


# International Funding and Researcher Matching V2.0

Ibrahim Asad, John Cedric Acierto, Sanchit Kumar, Raiyan Kabir, Aninda Shome, Ivan Lou Tompong  
Schulich School of Engineering, University of Calgary



## Abstract

- Researchers are always in need and looking for opportunities that may provide them with funding to further their research. This can often be a complicated and time consuming process that can often result in wasting the researcher's time.
- As such, a program to automate the process of delivering Funding Opportunities to Researchers was created. The program obtains keywords from FOs and Researchers using Natural Language Processing and attempts to match the best FOs to researchers using Machine Learning techniques.
- To create the resulting program, the BeautifulSoup and KeyBERT libraries were used in Python to extract content from webpages and obtain keywords from the content. The matching process is then performed by a Support Vector Machine model.

## Introduction

- Finding funding is an important role that any researcher needs to carry out in order to be able to conduct their research.
- However, due to their busy schedules, researchers often have little to no time to properly search for the various Funding Opportunities that exist. This process becomes harder as researchers try to find Opportunities that are truly relevant to their field of study and as researchers try to find Opportunities that are offered by organizations outside of the US and Canada.
- IFARM V1.0 exists to facilitate this process. Researchers can use a Programmable Search Engine that has been created to specifically return results related to Funding Opportunities. This is however a manual process.
- Our object is to automate the process of matching researchers with Funding Opportunities that they could truly benefit from and to bring more Funding to the University of Calgary overall.

## Results

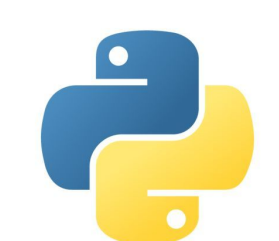
In terms of Keyword Extraction from Funding Opportunity Links and Researcher Profiles, it was found that KeyBERT, along with its cosine similarity as well as the synonyms creator algorithm that was created, would reliably output keywords relevant to the webpage being extracted. On average, 85% of keywords extracted would be directly relevant to the webpage and useful to summarize its contents. The performance of this algorithm was also evaluated and it was found that the algorithm could extract through 400 web pages in a span of 10 minutes.

On top of this, the Keyword Matching algorithm was also thoroughly tested to determine its accuracy. A number of researcher profiles were matched against and it was found that the number of relevant Funding Opportunities that were matched with a researcher ranged between 0 to 8 relevant opportunities out of 10. On average, 3-4 relevant funding opportunities would be output to the researcher. The performance of this algorithm was also tested and it was found to produce the result for one researcher in around 7 seconds.

Lastly, a server and a website were successfully setup to facilitate the usage of the application.

## Discussion

- In our results, the number of relevant funding opportunities matched with a researcher ranged between 0 to 8 opportunities out of 10. Out of all the researcher profiles tested, only one had 8 relevant opportunities matched, a few had 5-7 matched, and the rest either had 0-4 relevant opportunities matched with them.
- From these results, we can tell that our application is highly dependent on the contents of the funding opportunity database itself. In our results, the application is able to properly match relevant funding opportunities with researchers if the funding opportunities are present in the database. However, if none of the opportunities in the database are relevant to the researchers profile, then the researchers will be matched with irrelevant links.
- Since our application matches funding opportunities with researchers using keywords, the researcher profiles are also important to get good results. If a researcher's profile lacks proper information, the application is bound to extract keywords that are irrelevant and not useful to the matching of researchers and funding opportunities, possibly skewing the results.
- Outside of improving the database and researcher profiles, here are some possible methods that can be used to improve the results in further testing:
  - Instead of using keywords for the matching of researchers and funding opportunities, key phrases can be used.
  - Different machine learning models or techniques such as ensemble learning can be used to further improve the matching results



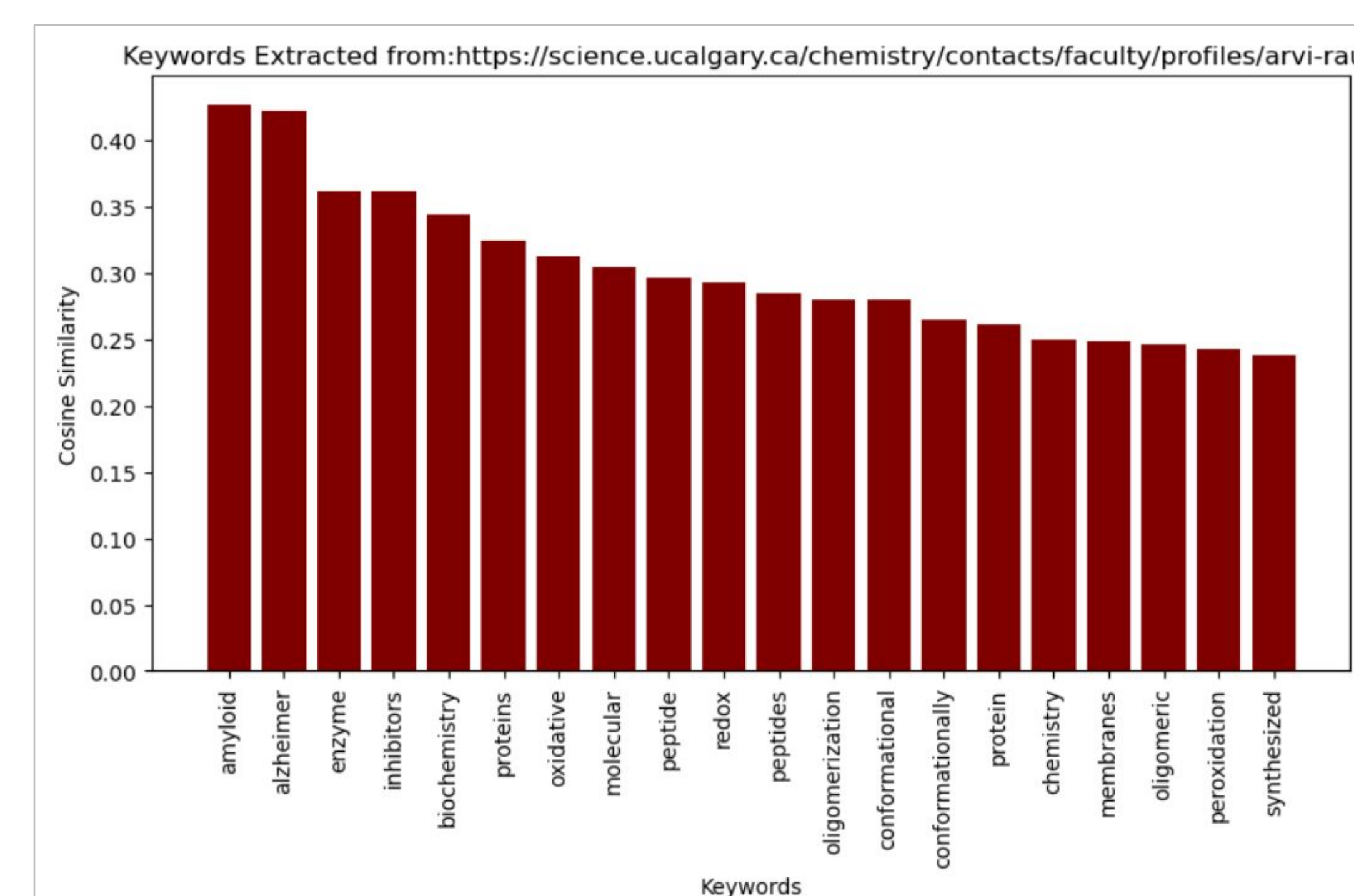
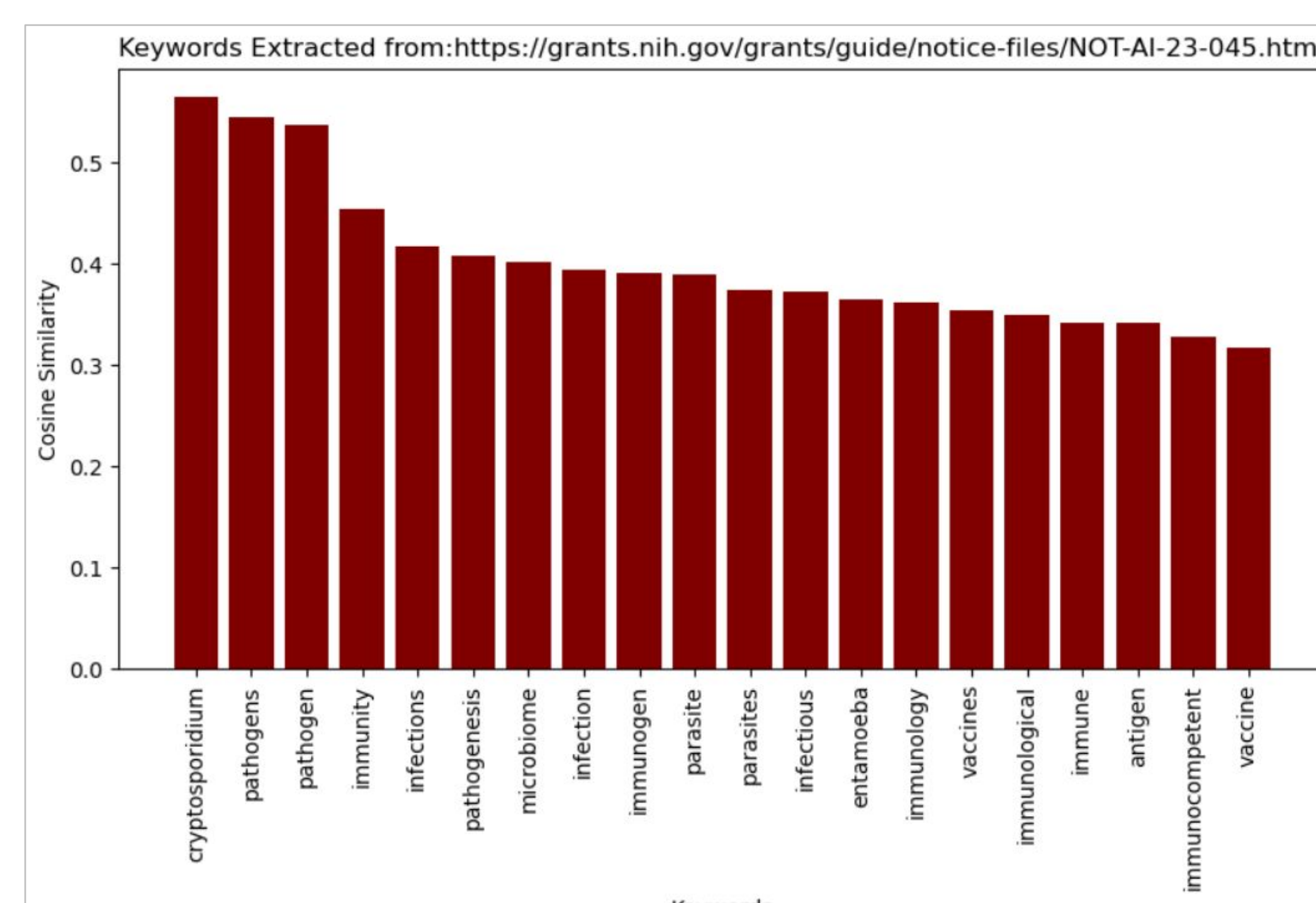
Python



Flask



Apache CloudStack



## Methods and Materials

To perform the functionality of matching Funding Opportunities to Researchers, we were provided with Researcher Profiles that contained the research researchers would like to perform. As a result, it was decided that we would obtain keywords from these profiles and try to match them to keywords obtained from Funding Opportunities. As a result, the overall project was separated into 3 parts:

1. Extract keywords, and find their synonyms, from researcher profiles and funding opportunity pages.
2. Match the best Funding Opportunities to each researcher using a machine learning model. The machine learning model of choice was an Support Vector Machine (SVM)
3. Create and Host a Frontend application that can be used by the sponsor to add researchers and funding opportunities to the database and to have the matching be performed.

