



Trine University
Electrical and Computer Engineering

Trine Search Engine

Andrew Baughman, Joshua Braun, Conner Loveless, Joshua Middleton, Logan Rigdon

Advisor: Mahesh Khadka, PhD

Trine University

One University Avenue, Angola, Indiana 46703

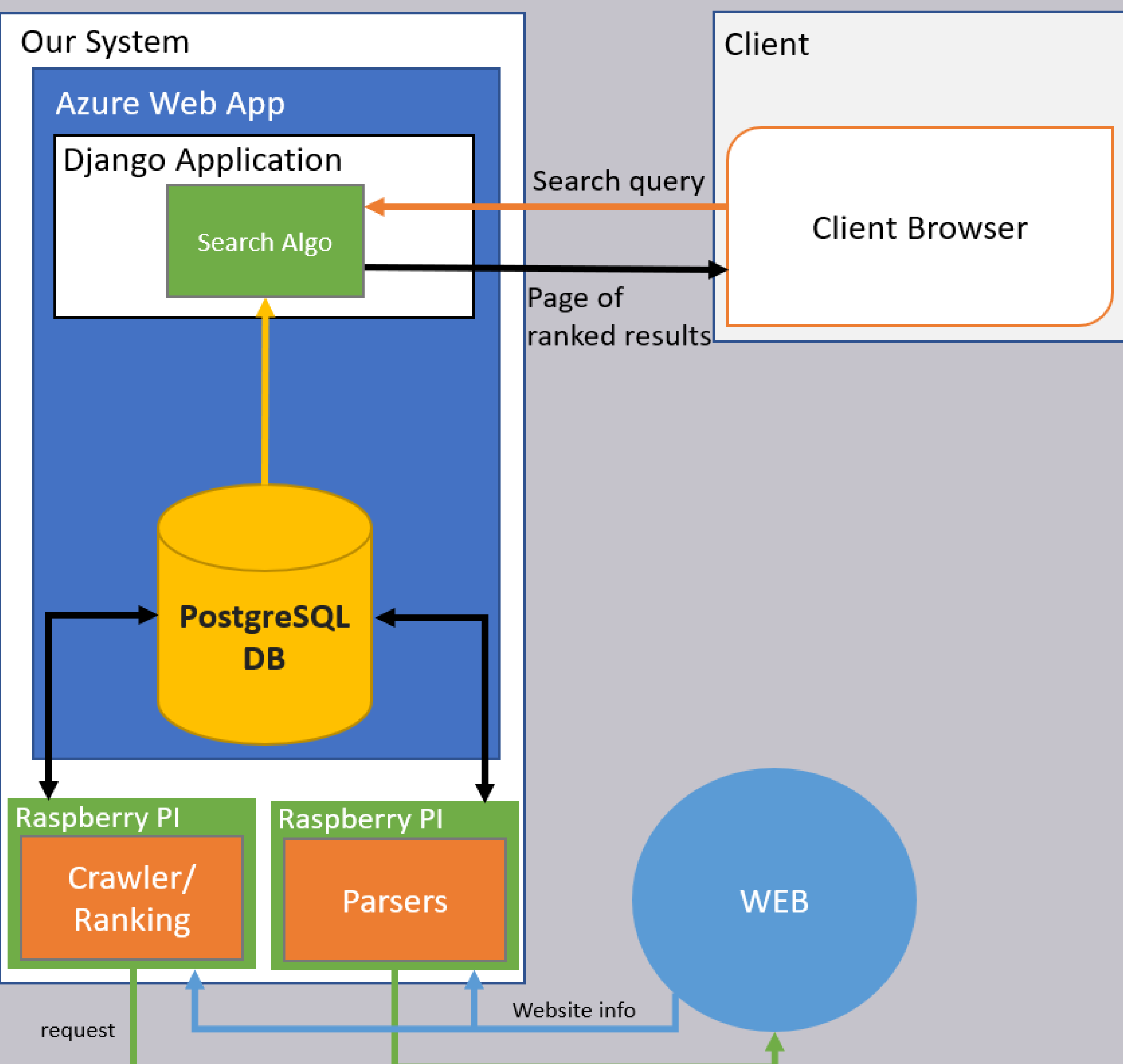


Introduction:

As students who have been at Trine for four years, our team has recognized that certain webpages for Trine are not easily found. Trine information such as the academic calendar, finals schedule, and TrineOnline should be readily available to anyone looking for them. To remedy this issue, this project aims to provide a search engine with the capacity to reliably find Trine resources. Our system crawls the Internet, finding new webpages. It then stores identifying information from those visited webpages. Another program goes through this database and finds further information from the associated webpages to store into other database tables. A query can be made on the search engine's website. Our search algorithm finds and sorts relevant webpages from the database to display to the end user.

System Design:

Our webservice uses the Django python framework to manage a PostgreSQL database and to serve a website to users. This webservice runs on an Azure cloud virtual machine and is accessible from any internet connected computer. The data for the database is collected by two python programs running on two different Raspberry Pi computers.



System Code

The code that makes queries possible can be divided into two categories: preprocessing and postprocessing. Preprocessing occurs on the Raspberry Pi systems that run the crawler and parser. As pages are collected from the internet, important data is stored. This data includes title, description, URL, keywords that appear on the page, and keywords that appear in the URL or title. This preprocessed data can be comprehensive without slowing the server since it is collected before any query is made. Postprocessing is the code that runs after the user makes the query. This is what most would call the search algorithm. The search algorithm makes use of the data collected in the preprocessing steps to return results to the user quickly.

Results and Discussion:

- Users can access the Trine Search website at: <https://trine-search.azurewebsites.net/>
- Results from a query are returned within 3 seconds on average when the query consists of 2 terms
- Returned results are relevant to the query term(s)
- Features such as voice search, image search (see Figure 2), Feelin' Lucky search, and theme customization are available to the end-user
- The web crawler and parser can run continuously, building the system's database of webpages

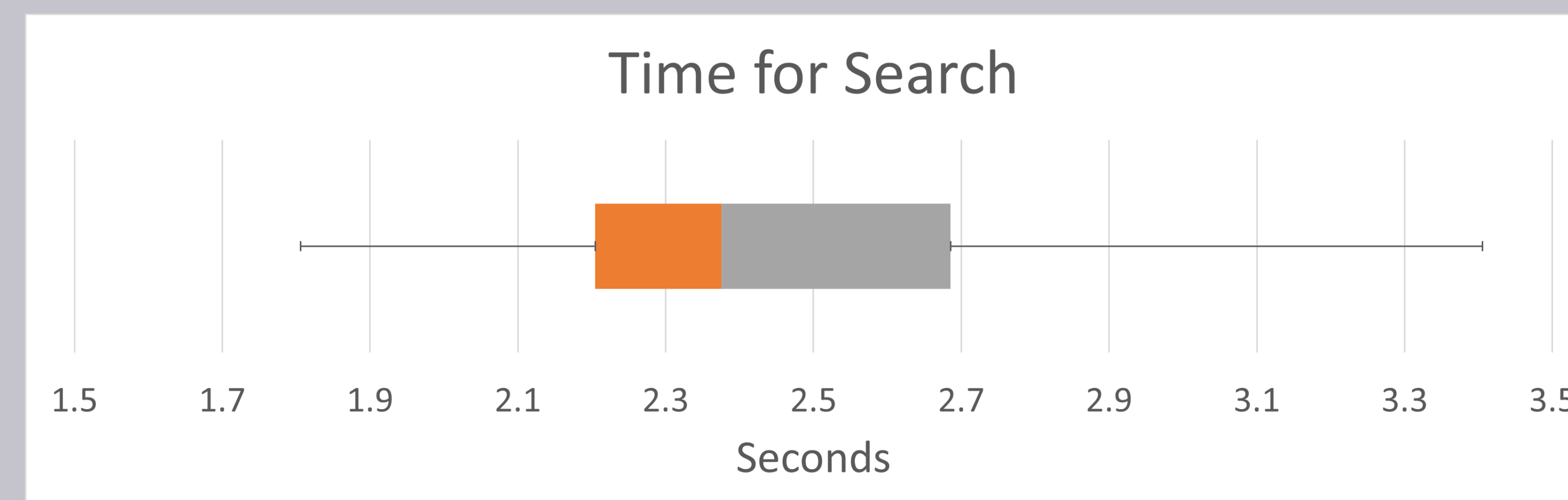


Figure 2: How long does a search take?

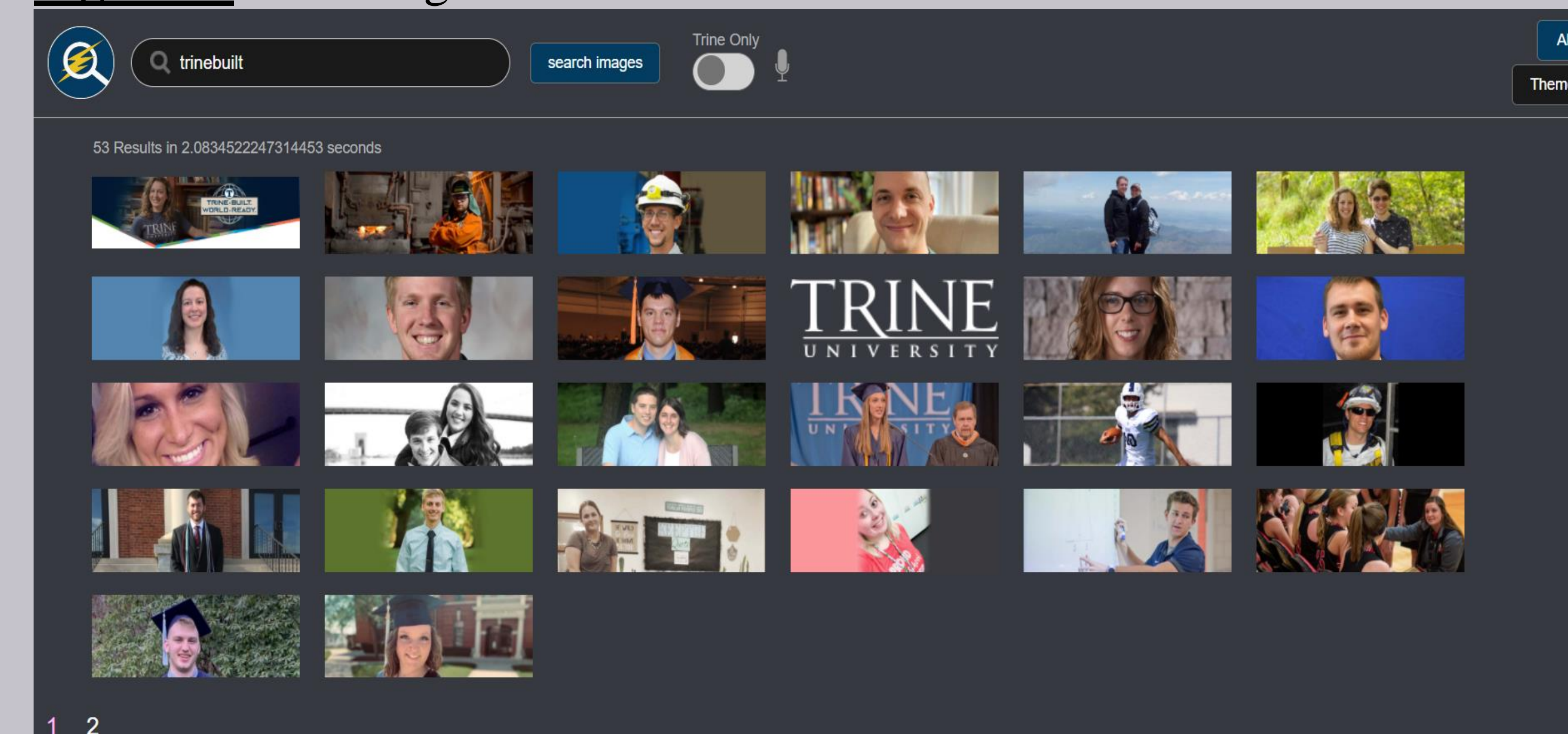


Figure 3: Images are displayed from an image search.

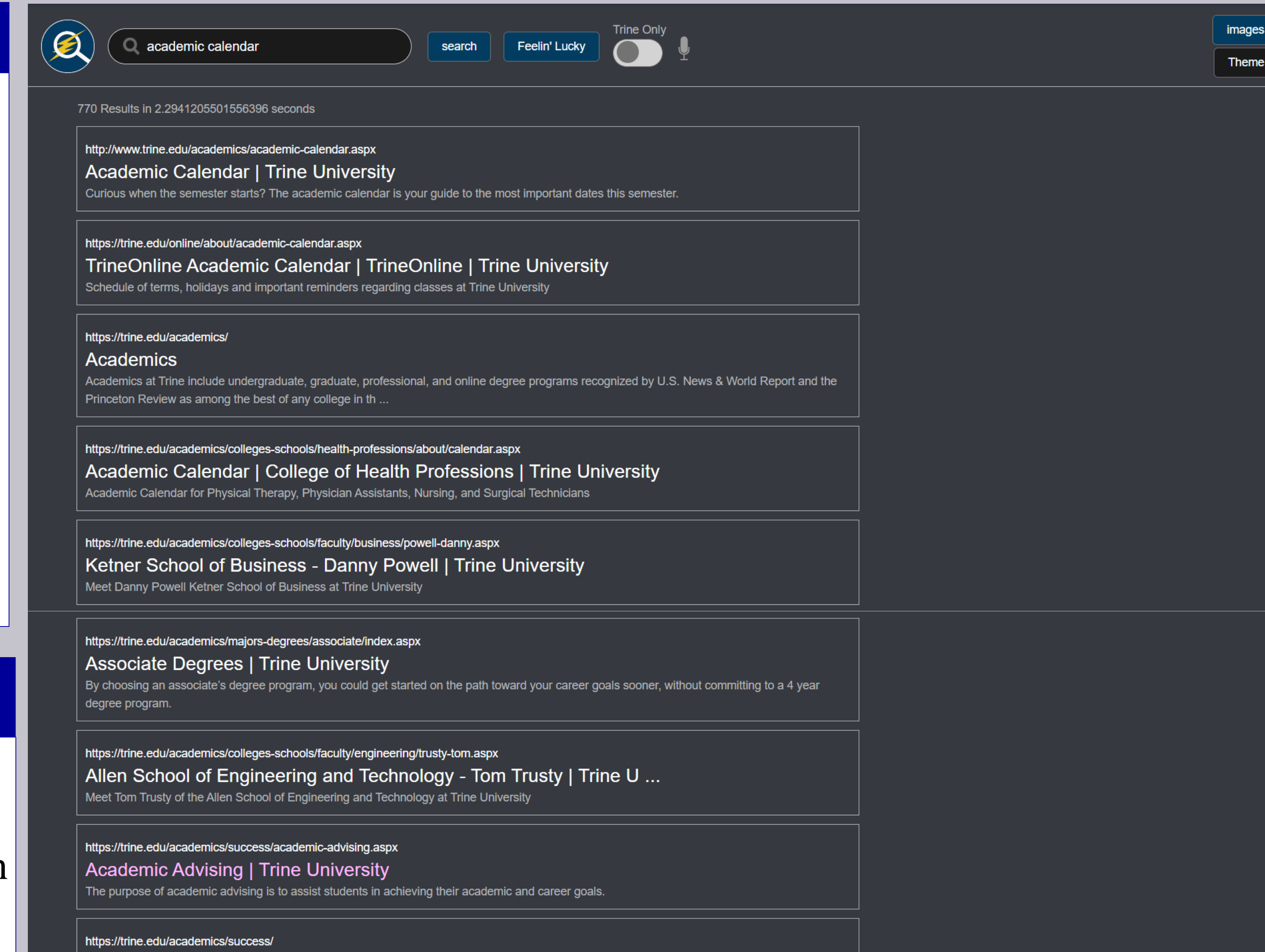


Figure 4: The results of a search. This features a search bar, voice search tool, Trine toggle, theme editor, and image search option.

Conclusion:

The result of this project is a search engine tailor-made for those interested in information associated with Trine University. In the future our engine could be used by staff, students, and perspective new students. The comprehensive documentation compiled in the duration of this project will make future implementations simple and effective.

Future Work:

The following are ideas to improve the Trine Search Engine project:

- Predictive text feature for the search field
- Translate page feature for the results page
- Further refinement of the search algorithm
- Quick info boxes (like Google's Knowledge Panels)
- User driven ranking system (vote on result quality)

Acknowledgements:

The Trine Search Engine senior design team would like to thank the following for their contributions, facilities, and resources:

- Trine University Electrical and Computer Engineering Department
- Dr. Mahesh Khadka

Figure 1: How the Trine Search Engine system is logically designed.