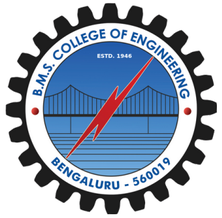
Hewlett Packard Enterprise



Open Source Finder

Student Team:

Akshat Pandey (1BM20IS015)
Arpan Bhusal (1BM20IS198)
Gagandeep N K (1BM20IS038)
Kshama Bhatt (1BM20IS202)
Rahul T G (1BM20IS112)
Rohit DB (1BM20IS122)

HPE Mentor:

Mr. Arun Ramachandra
Mr. Murali Krishna

Faculty Mentor:

Dr. Sindhu K
Dr. Nalina V

Agenda

- Abstract
- Problem Statement
- Objectives
- Initial Architecture
- APIs - Libraries.io
- Implementation of Backend API using Flask
- Fuzzy Logic
- Demo
- Learnings
- Challenges
- Future Scope
- References

Abstract

- Open Source Finder is an open-source application that allows users to search for open-source components for their projects.
- It provides a user-friendly interface to discover and explore various open-source repositories based on specific criteria such as license and platform.
- With Open Source Finder, developers can easily find and integrate open-source components into their projects, saving time and effort.

Hewlett Packard Enterprise

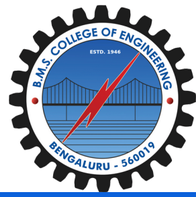
B.M.S. COLLEGE OF ENGINEERING ESTD. 1946 BENGALURU - 560019

Problem Statement

- Selecting an open-source software based on specific needs can be challenging due to the lack of efficient and reliable methods to identify open-source software that aligns with the user's requirements. As a result, users often face difficulties in finding the most suitable open-source solution that meets their specific needs and preferences.

- To address this issue the Open Source Finder software is developed. This software aims to provide a systematic and efficient approach for users to discover and select open-source software based on user needs.

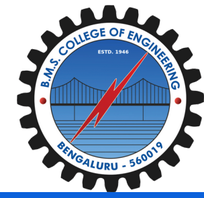
Hewlett Packard Enterprise



Objectives

- Develop a software tool that enables users to search for open-source components based on specified criteria, covering both internet resources and a local open-source components database.
- Allow users to save the retrieved documents/resources as potential candidates for further evaluation, providing a mechanism to store the results.
- Extract relevant information from the saved documents, including the download repository and source code repository , as well as details related to number of forks,language used, latest updated date and licensing.
- Provide a summarized view of the components, presenting key information in a concise manner to facilitate decision-making and comparison among the candidates.

Hewlett Packard Enterprise



Initial architecture

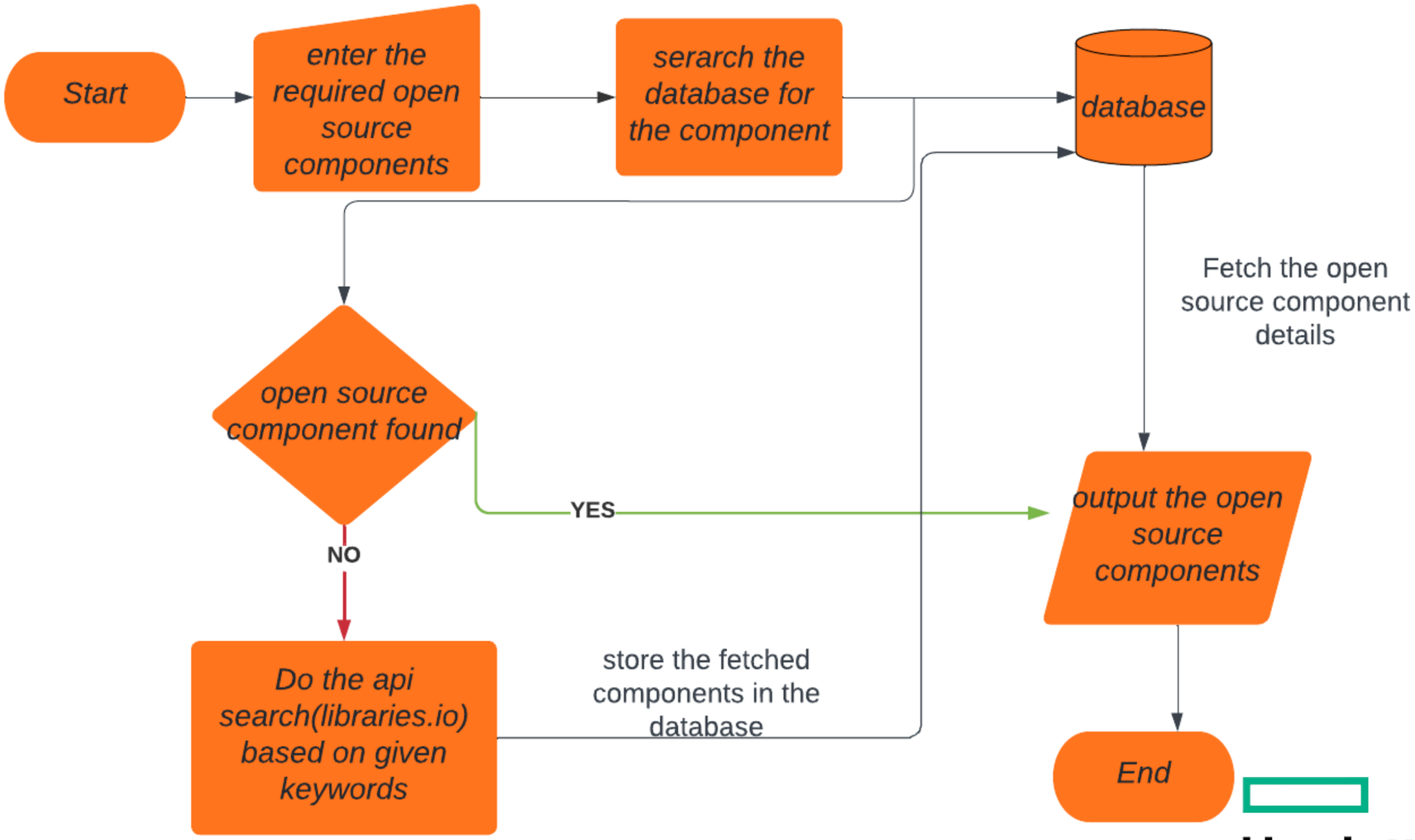


fig 1: Architecture

Hewlett Packard Enterprise

API's Libraries.io

The Libraries.io API allows to search for open source components and retrieve information about them. Here is a breakdown of how to use the Libraries.io API:-

- Registration: Create an account and obtain an API key from Libraries.io. The API key will be used to authenticate requests.
- Search for components: Use the search endpoint to find open source components based on various parameters such as keywords,licenses, languages and platform.
- Retrieve required information: Once the required component is obtained, make a request to the API endpoint for retrieving various components information. This may include details like the component's name, description, version, license, repository URL and maintainers.

Hewlett Packard Enterprise

B.M.S. COLLEGE OF ENGINEERING ESTD. 1946 BENGALURU - 560019

Implementation of backend API using Flask

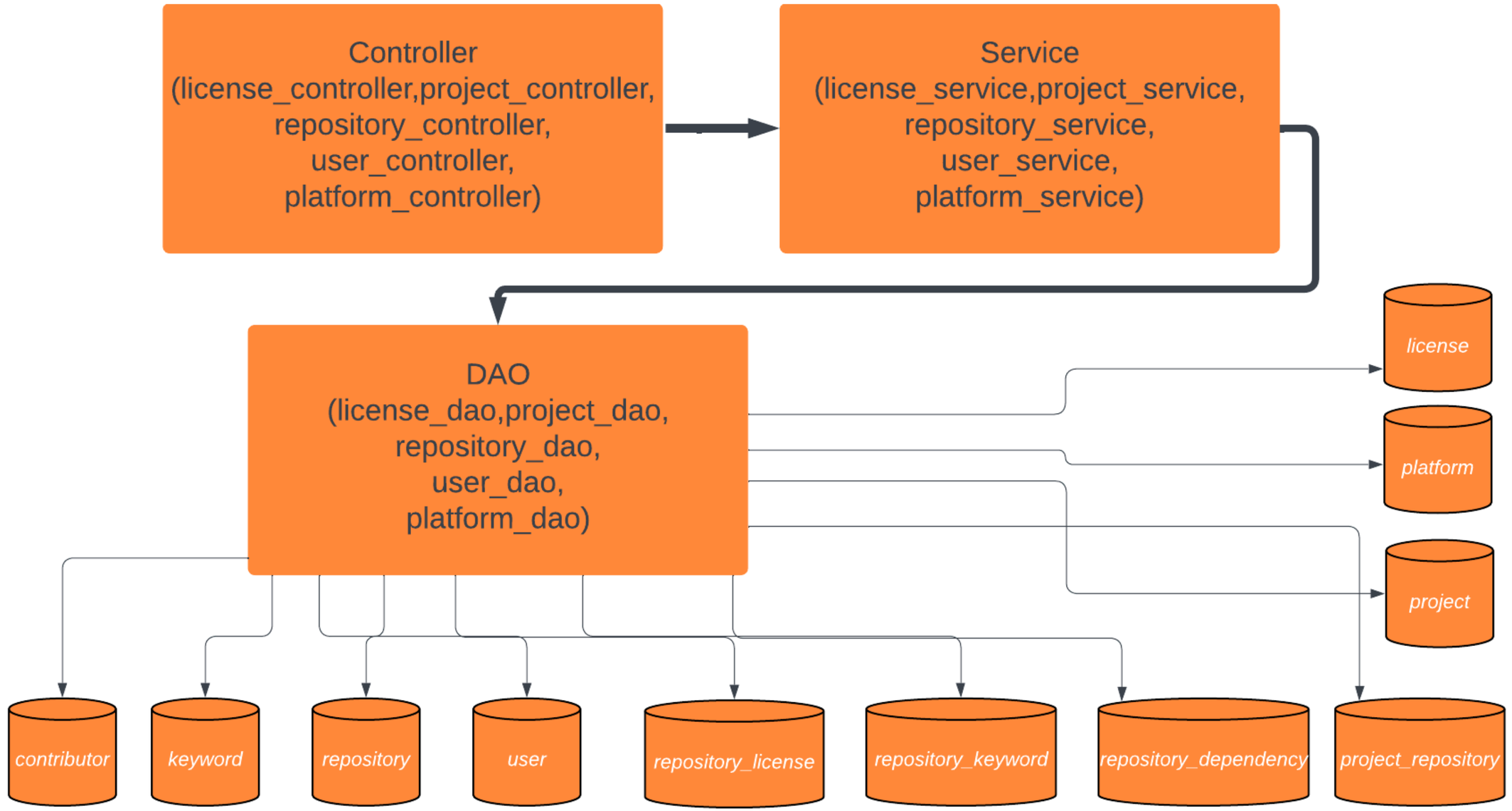


fig2 Backend architecture

Fuzzy Logic

Fuzzy logic is a technique used in information retrieval to account for approximate matches and handle typographical errors or variations in search queries.

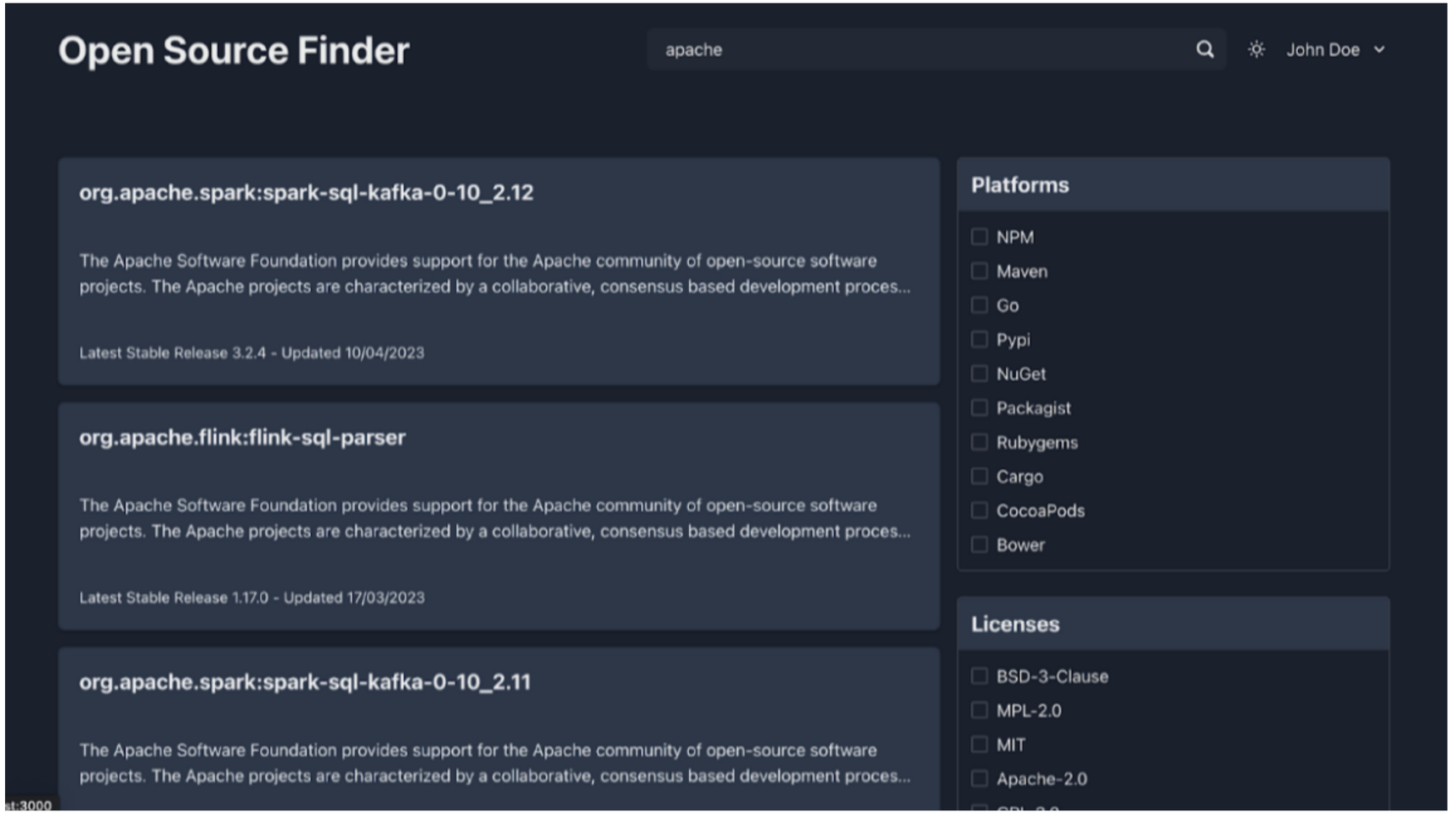
- Fuzzy Query: Fuzzy Query feature allows to search for terms that are similar to a specified term, taking into account variations like misspellings user can search for the components even if the keyword is incorrect.

- Edit Distance: Fuzzy matching in Elasticsearch is based on the concept of edit distance. Elasticsearch uses the Levenshtein distance algorithm to calculate the edit distance.

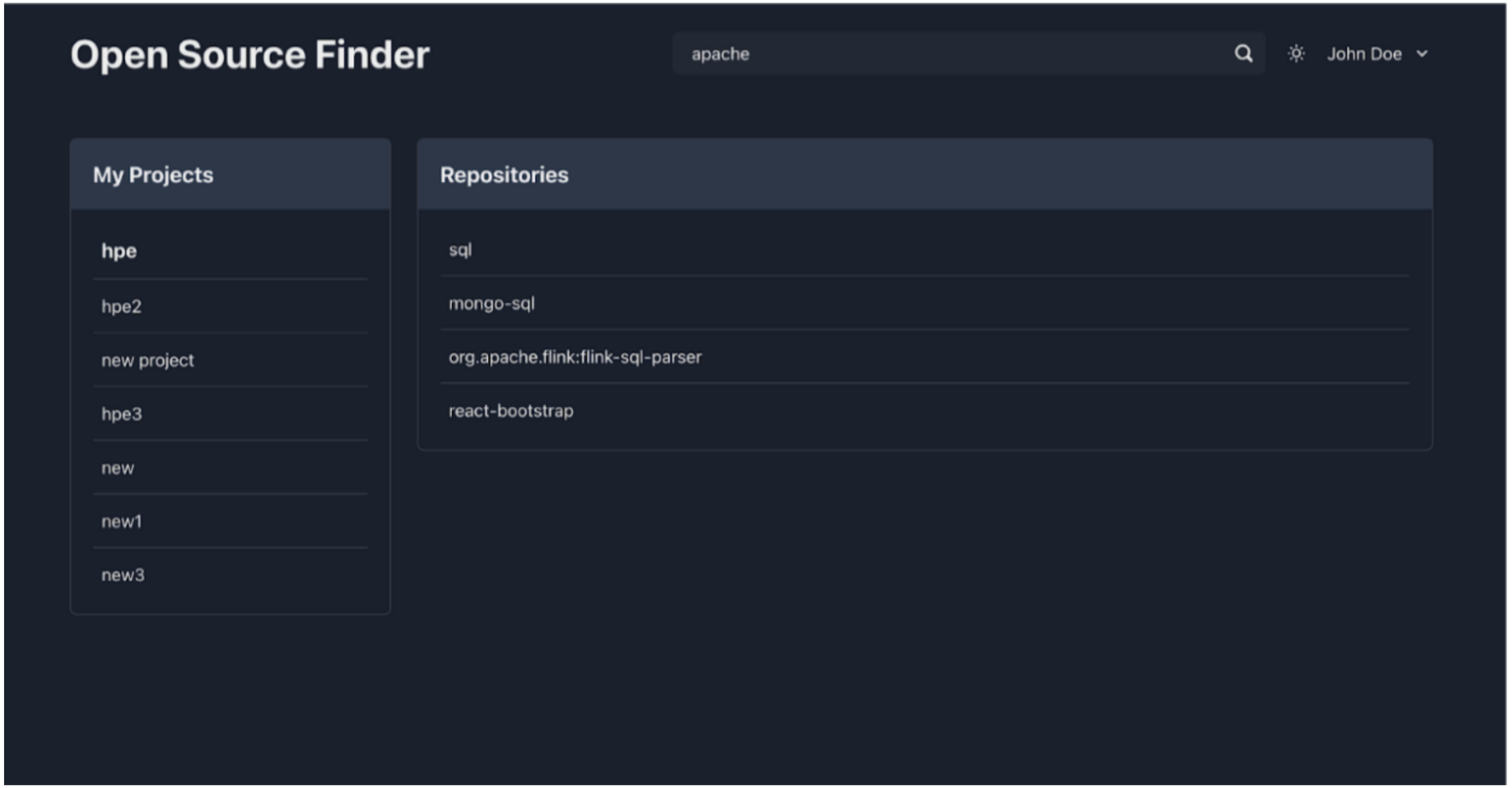
Hewlett Packard Enterprise

B.M.S. COLLEGE OF ENGINEERING ESTD. 1946 BENGALURU - 560019

Demo

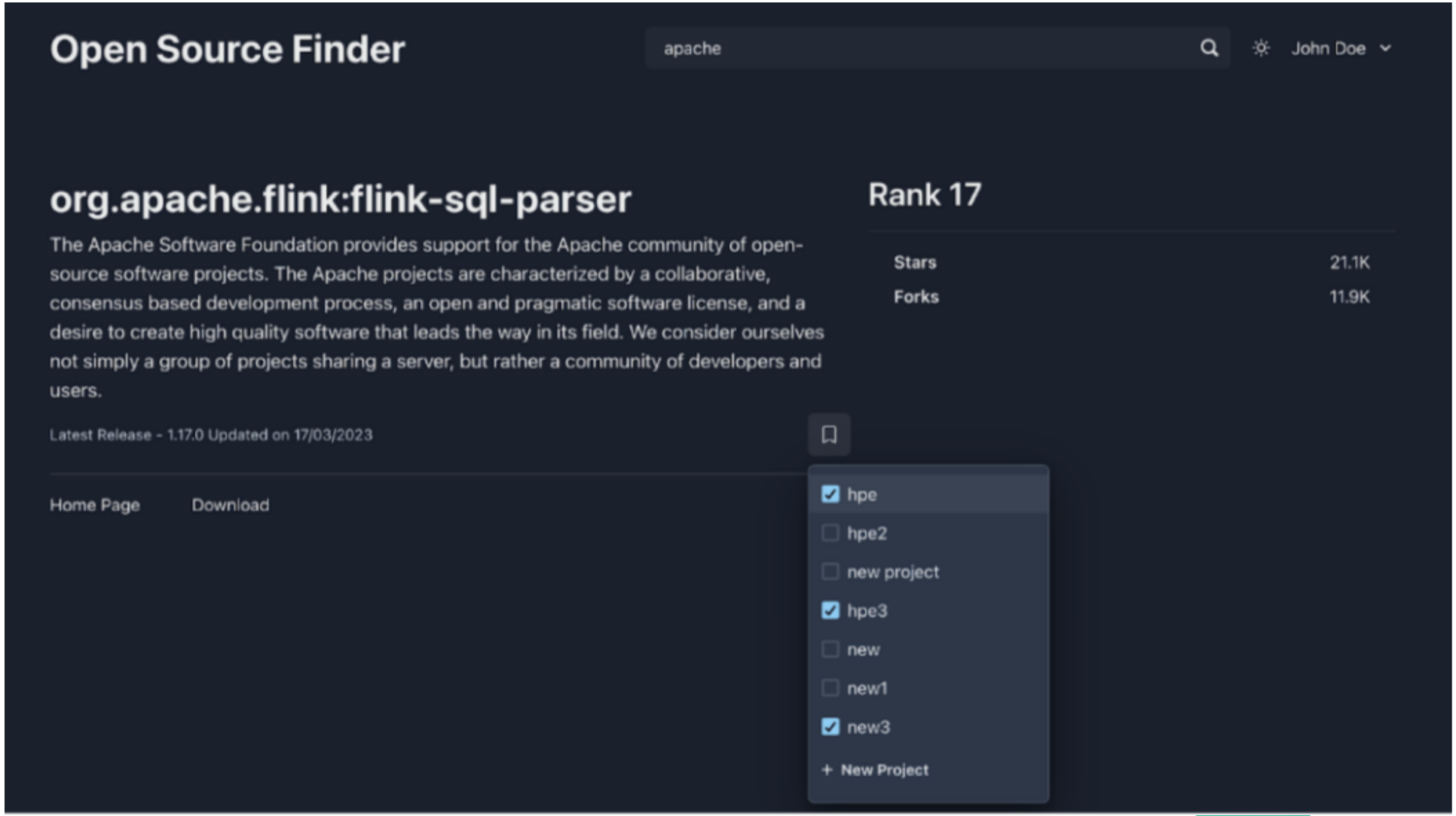


Hewlett Packard Enterprise



Hewlett Packard Enterprise

B.M.S. COLLEGE OF ENGINEERING
ESTD. 1946
BENGALURU - 560019



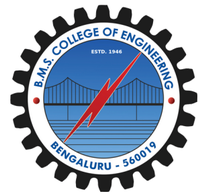
Hewlett Packard Enterprise

B.M.S. COLLEGE OF ENGINEERING ESTD. 1946 BENGALURU - 560019

Learnings

- Introduction to fuzzy logic .
- Good understanding of Model view controller(MVC) architecture.
- Introduction to openai and langchain to train a large language model for building a generative AI.
- Introduction to redis cache

Hewlett Packard Enterprise



Challenges

- Extending the search to support more english like statements
- Dependency on libraries.io api
- Validation

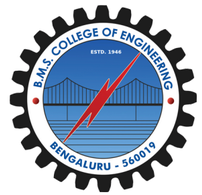
Hewlett Packard Enterprise

B.M.S. COLLEGE OF ENGINEERING
ESTD. 1946
BENGALURU - 560019

Future Scope

- Build a generative AI chatbot by training a large language model with data extracted from the database.
- Integrate a user-friendly form that enables users to manually input and add open-source components to the system.
- Implement a validation mechanism to ensure the integrity and quality of the fetched open-source components.
- Reduce Dependency on libraries.io api , by exploring other api's such as PyPI (Python Package Index), npm (Node Package Manager), Maven (Java packages), or RubyGems (Ruby packages),

Hewlett Packard Enterprise



Resources and Links

Resources:-

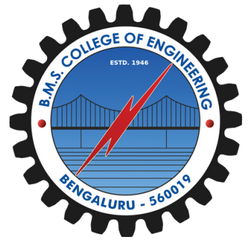
- https://libraries.io
- https://platform.openai.com/docs/api-reference
- https://pypi.org/project/fuzzywuzzy/

Links:-

- https://github.com/Akshat9254/hpe_open_source_finder

Any Questions?

Hewlett Packard
Enterprise



Thank You