



Beta Presentation

Semantic Search for Code and Architecture Assets

The Capstone Experience

Team Amazon

Zayd Abualfellet

Sampan Chaudhuri

Jerry Chen

Atharva Kirkole

Nicholas Li

Department of Computer Science and Engineering

Michigan State University

Spring 2025



*From Students...
...to Professionals*

Project Overview

- Code snippet search
 - Natural language query
 - Most relevant results
- Centralizes code database
- Filter by various attributes
- Find subject matter experts

Team Member's Technical Tasks

Technical Tasks Assigned

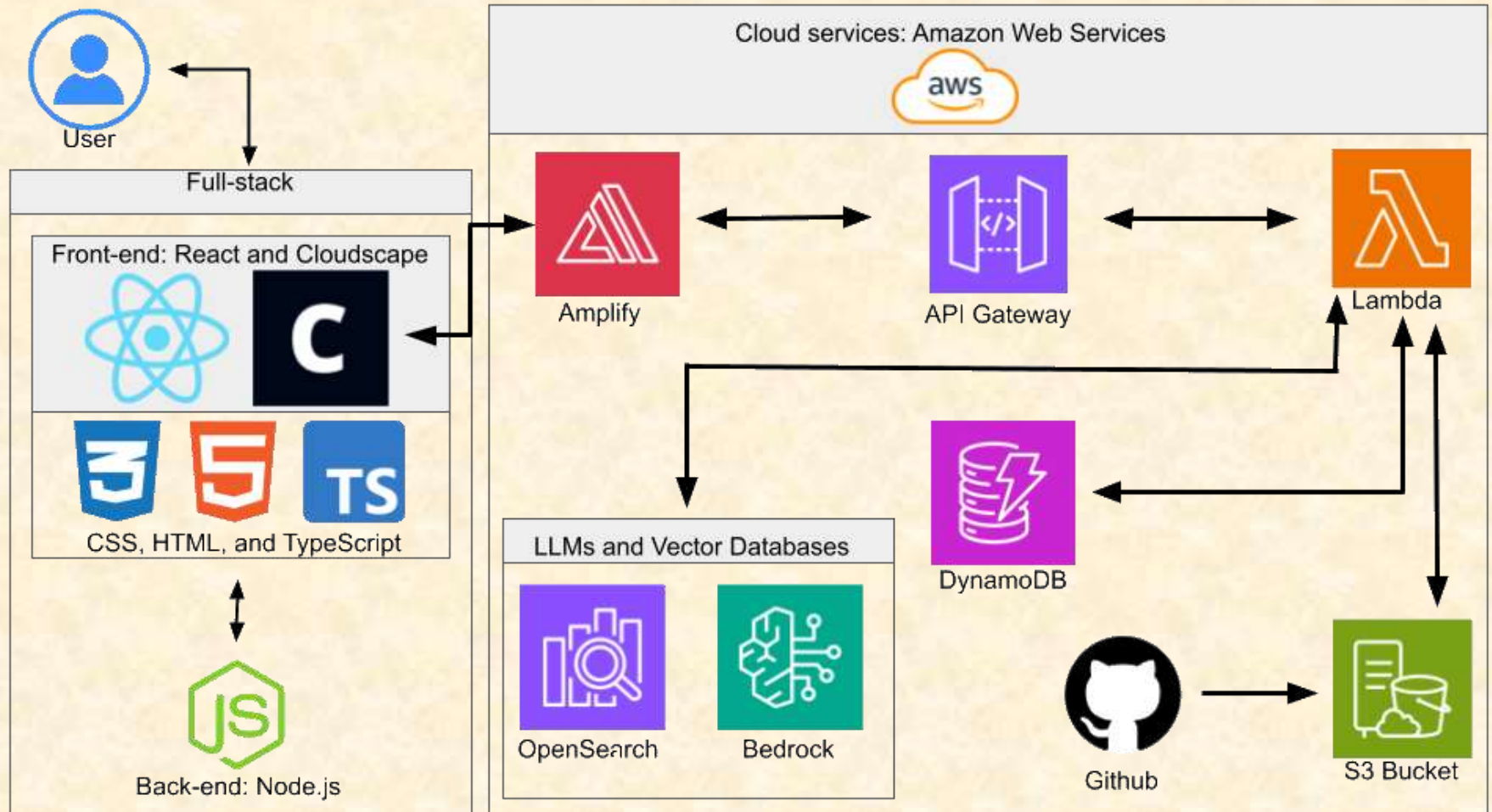
- Zayd Abualfellat
 - File parsing for Python, JS, C++
 - Extract commit history by file
 - Create vector embedding system for searches
 - Chunk vector embeddings to improve search accuracy
- Sampan Chaudhuri
 - Initialized CDK project to provision AWS resources
 - Created API endpoints and Lambda handlers for various events
 - Integrate filtering results from backend to frontend
 - Add repository ingestion process by URL extraction
- Jerry Chen
 - Class to store admin page settings
 - Prompt to generate function docstring
 - File parsing for C# and Java
- Atharva Kirkole
 - Create and configure OpenSearch domain for searching
 - Extract services by file
 - Support for various LLM models for docstring generation
 - System to retrieve SMEs working on specific services
- Nicholas Li
 - Prototype frontend to facilitate integration with backend
 - Revamp frontend using Cloudscape framework
 - Utilize NLP and DynamoDb to retrieve popular searches
 - Deploy website to Amplify and configure Cognito authentication

Technical Tasks Completed

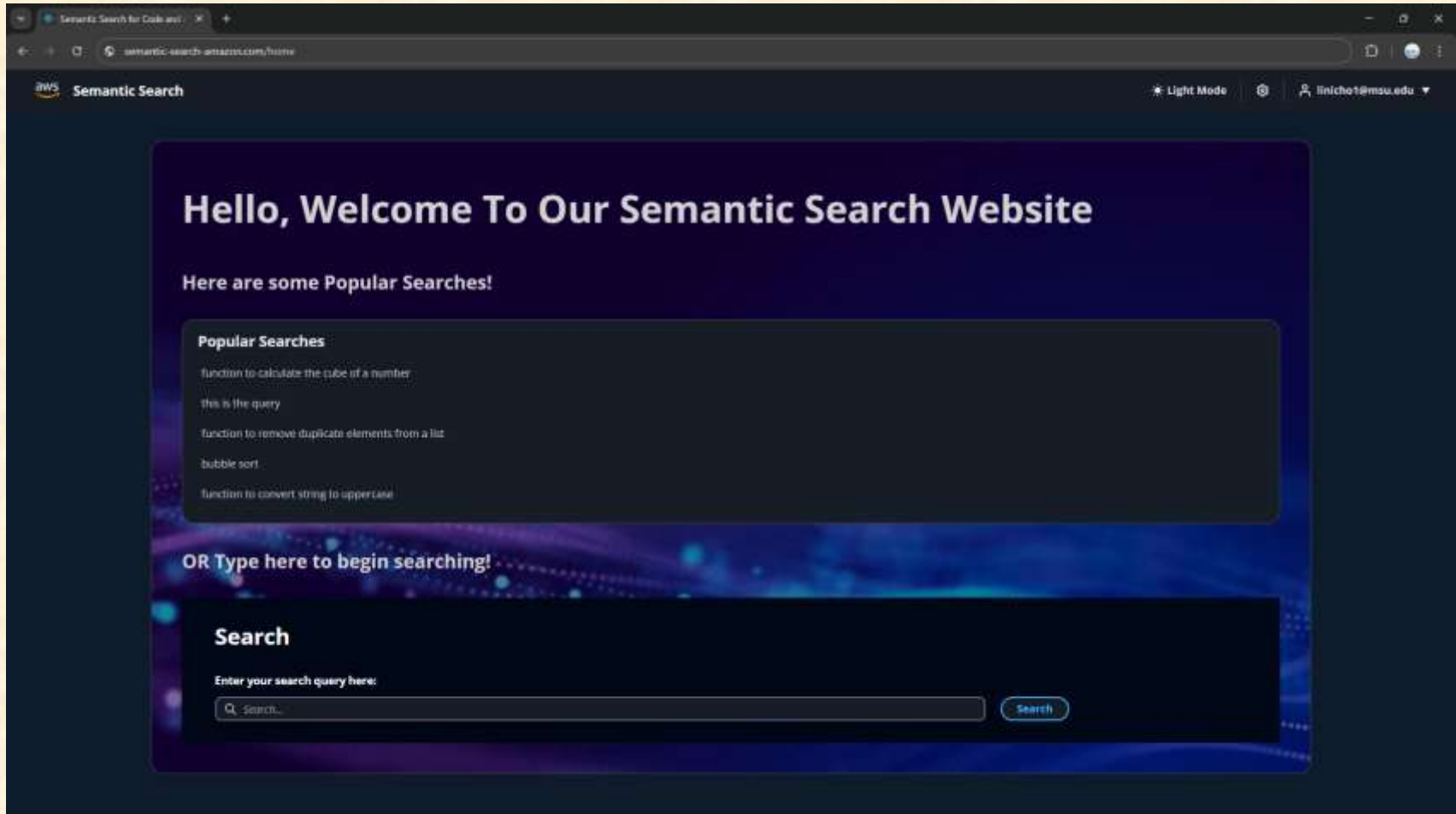
- Zayd Abualfellat
 - File parsing for Python, JS, C++
 - Extract commit history by file
 - Create vector embedding system for searches
 - Chunk vector embeddings to improve search accuracy
- Sampan Chaudhuri
 - Initialized CDK project to provision AWS resources
 - Created API endpoints and Lambda handlers for various events
 - Integrate filtering results from backend to frontend
 - Add repository ingestion process by URL extraction
- Jerry Chen
 - Class to store admin page settings
 - Prompt to generate function docstring
 - File parsing for C# and Java
- Atharva Kirkole
 - Create and configure OpenSearch domain for searching
 - Extract services by file
 - Support for various LLM models for docstring generation
 - System to retrieve SMEs working on specific services
- Nicholas Li
 - Prototype frontend to facilitate integration with backend
 - Revamp frontend using Cloudscape framework
 - Utilize NLP and DynamoDb to retrieve popular searches
 - Deploy website to Amplify and configure Cognito authentication



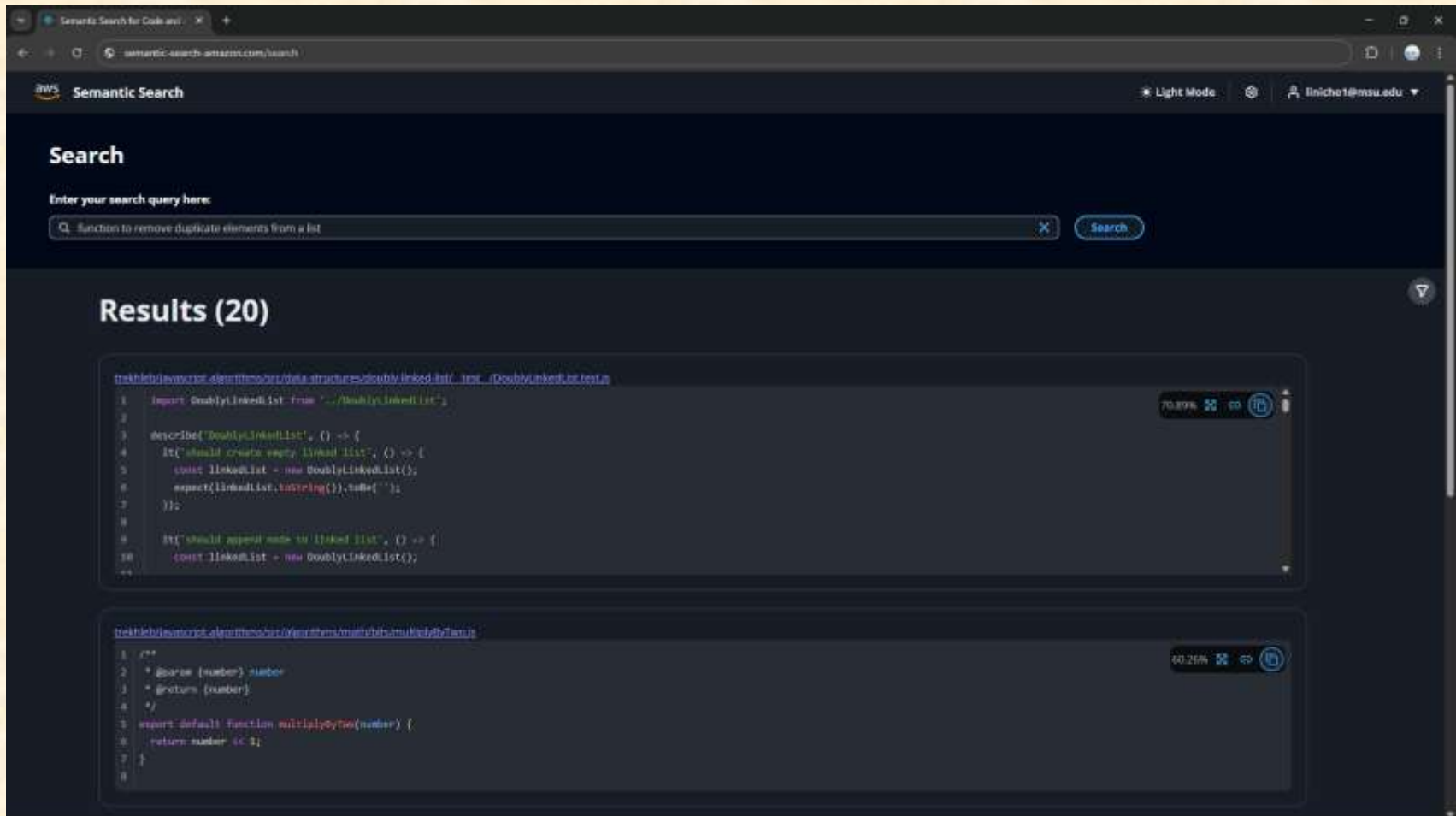
System Architecture



Home Page



Results Page



Results with Filters Applied

The screenshot displays the AWS Semantic Search web interface. At the top, the search bar contains the query "function to calculate the cube of a number". Below the search bar, the results section shows 20 results. The first result is a JavaScript file named "factorial_test.js" with a code snippet showing Jest tests for a factorial function. The second result is a JavaScript file named "factorial.js" with a code snippet showing the implementation of the factorial function. On the right side, a filter panel is visible, showing the "Language" filter set to "JavaScript" and a "Search Filter" input field.

Search

Enter your search query here:

function to calculate the cube of a number

Results (20)

```
1 import factorial from './factorial';
2
3 describe('factorial', () => {
4   it('should calculate factorial', () => {
5     expect(factorial(0)).toBe(1);
6     expect(factorial(1)).toBe(1);
7     expect(factorial(5)).toBe(120);
8     expect(factorial(8)).toBe(40320);
9     expect(factorial(10)).toBe(362880);
10  });
11 });
```

Filter

Language:

- ☐ Python
- ☒ JavaScript
- ☐ C++

Search Filter:

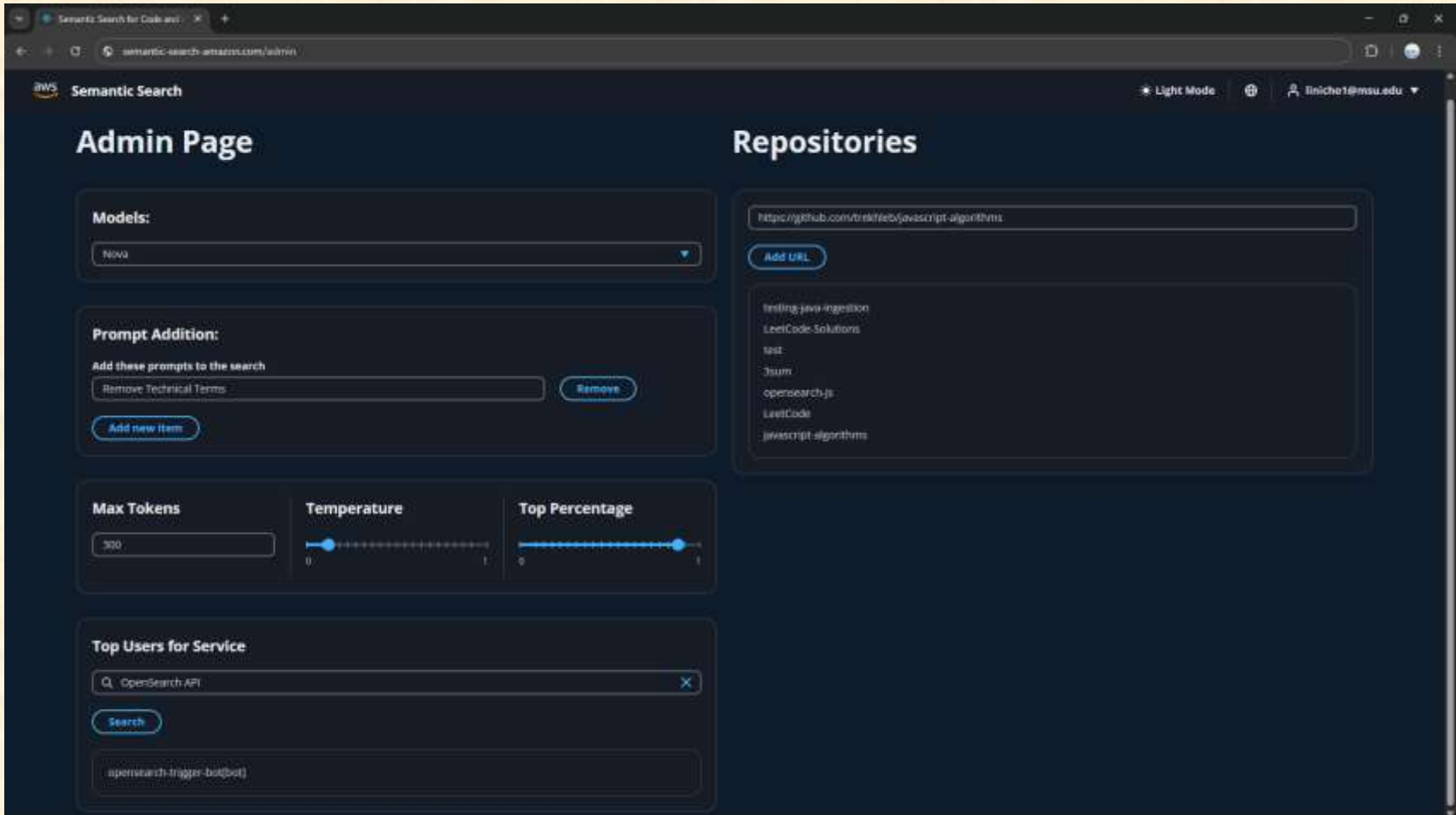
filter searches

Apply Filters

```
1 /**
2  * @param {number} number
3  * @return {number}
4  */
5 export default function factorial(number) {
6   let result = 1;
7
8   for (let i = 2; i <= number; i++) {
9     result *= i;
10  }
11 }
```



Admin Page



The screenshot shows the 'Admin Page' of the 'Semantic Search' application. The page is divided into two main sections: 'Admin Page' on the left and 'Repositories' on the right. The 'Admin Page' section includes a 'Models' dropdown menu with 'Nova' selected, a 'Prompt Addition' section with a text input 'Remove Technical Terms' and an 'Add new Item' button, a 'Max Tokens' input set to '300', a 'Temperature' slider set to 0.5, a 'Top Percentage' slider set to 0.8, and a 'Top Users for Service' section with a search input 'OpenSearch API' and a 'Search' button. The 'Repositories' section includes a text input for a URL 'https://github.com/trikiteb/javascript-algorithms' and an 'Add URL' button, followed by a list of repositories: 'trikiteb-javascript-algorithms', 'LeetCode-Solutions', 'test', 'jsm', 'opensearch-js', 'LeetCode', and 'javascript-algorithms'.

Semantic Search Light Mode lincho1@msu.edu

Admin Page

Models:

Nova

Prompt Addition:

Add these prompts to the search

Remove Technical Terms Remove

Add new Item

Max Tokens 300

Temperature 0.5

Top Percentage 0.8

Top Users for Service

OpenSearch API

Search

opensearch-trigger-bot(bot)

Repositories

https://github.com/trikiteb/javascript-algorithms

Add URL

- trikiteb-javascript-algorithms
- LeetCode-Solutions
- test
- jsm
- opensearch-js
- LeetCode
- javascript-algorithms



What's left to do?

- Features
- Stretch Goals
 - Supporting more languages
 - Customized login page
 - Retrieving user attributes from Cognito
- Other Tasks
 - Code documentation
 - Clean up
 - Security and minor bug fixes

Questions?

?

?

?

?

?

?

?

?

?

