MICHIGAN STATE UNIVERSITY

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

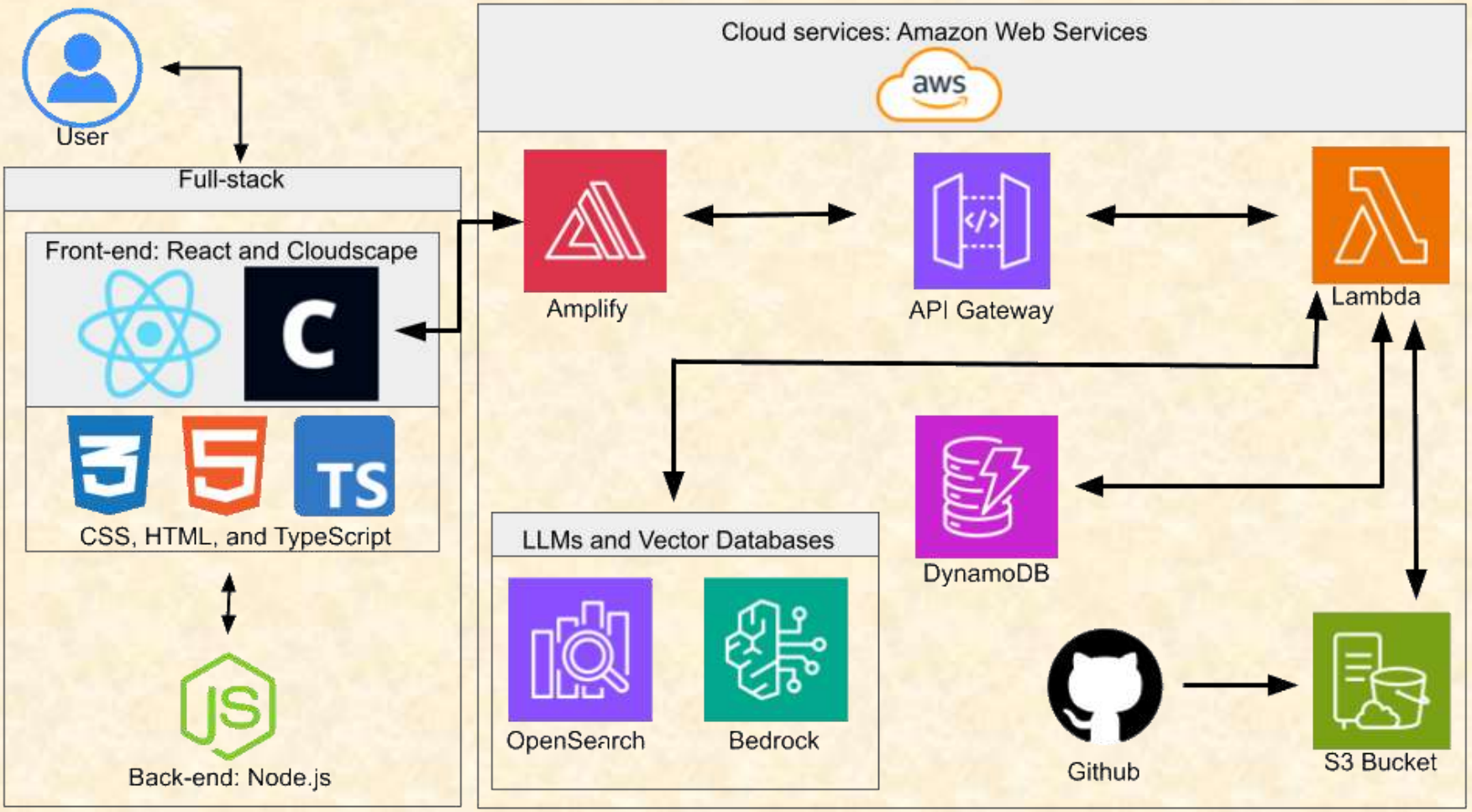
Technical Tasks Assigned

- 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

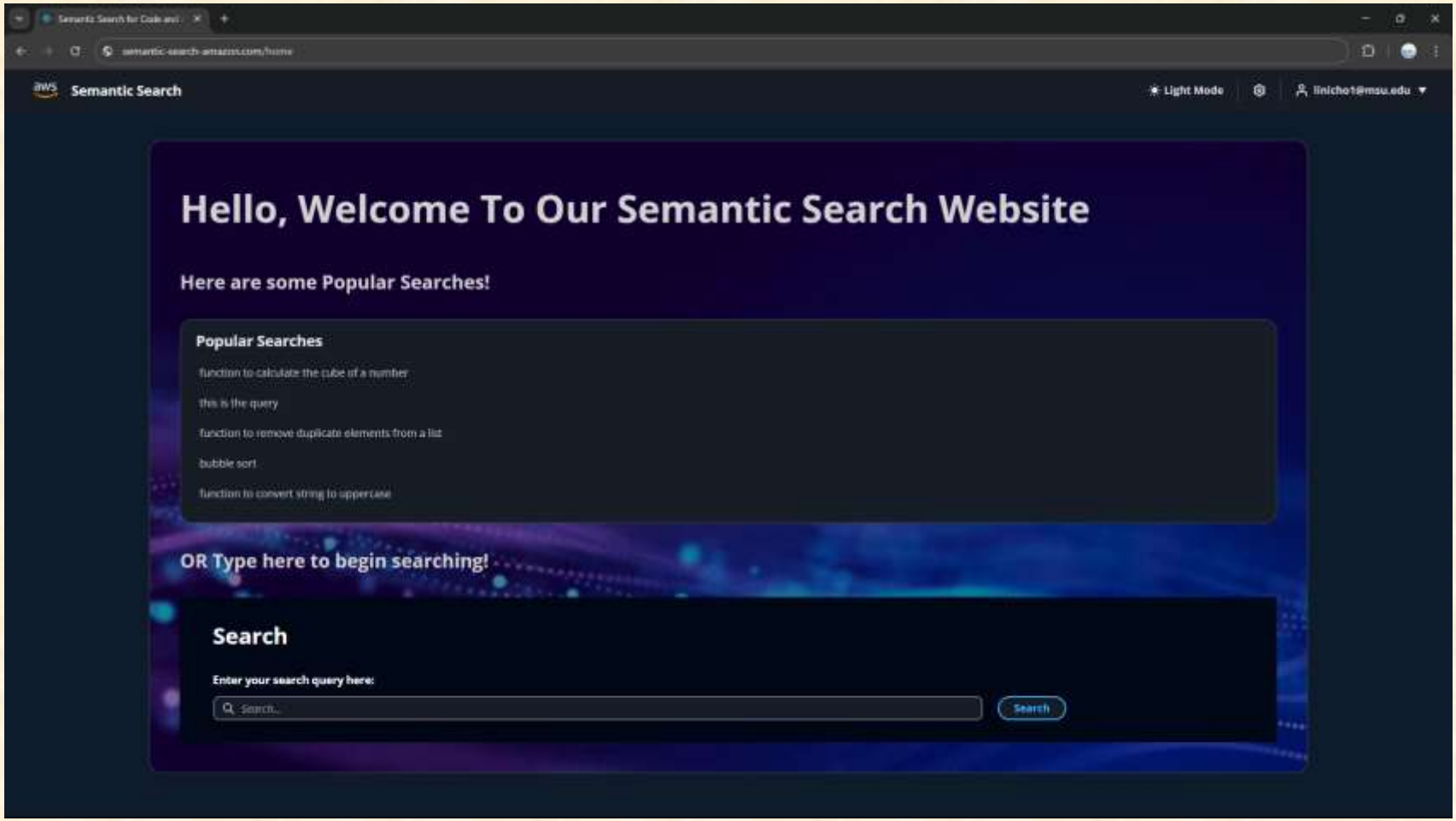
Technical Tasks Completed

- 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

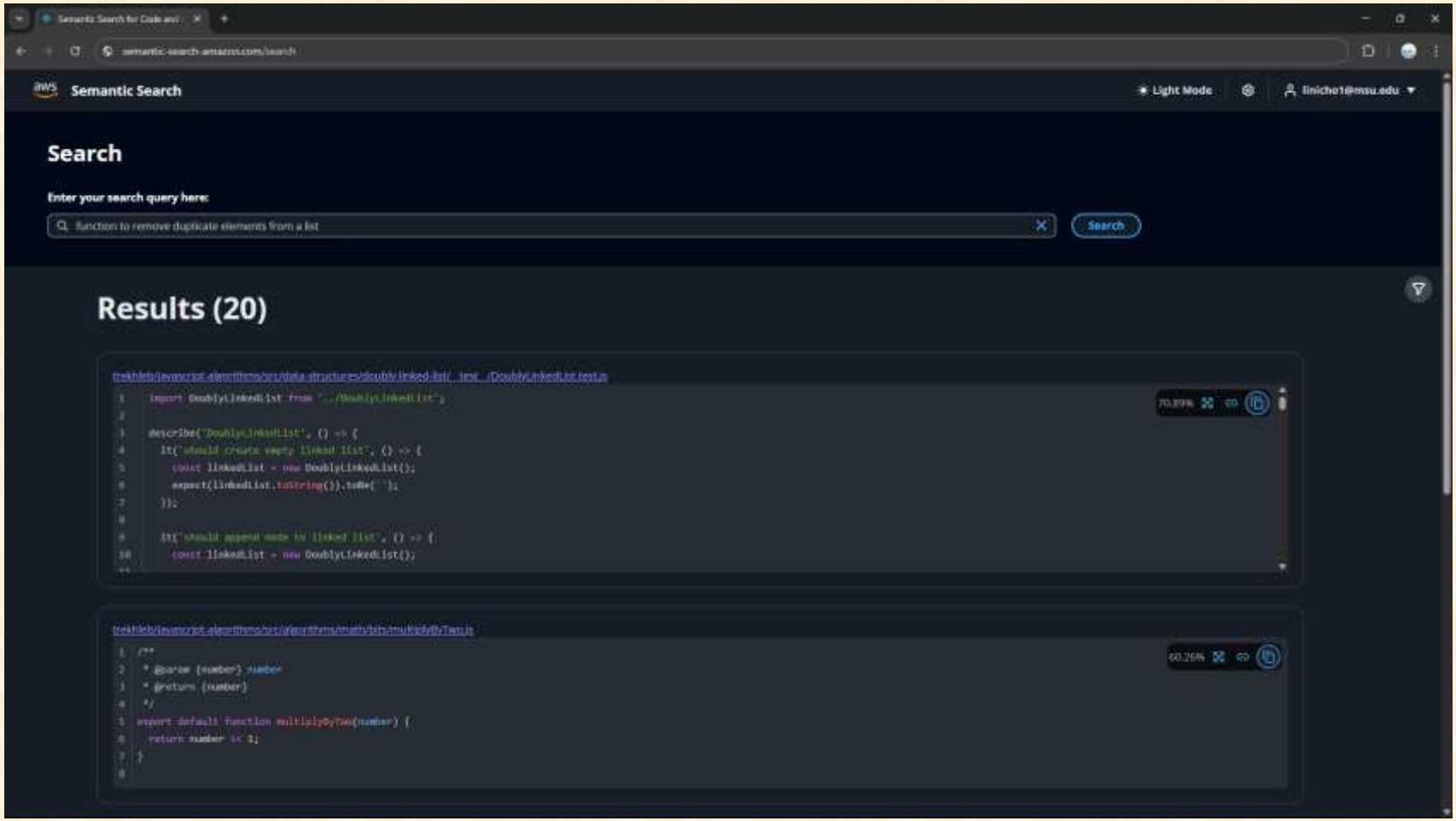
System Architecture



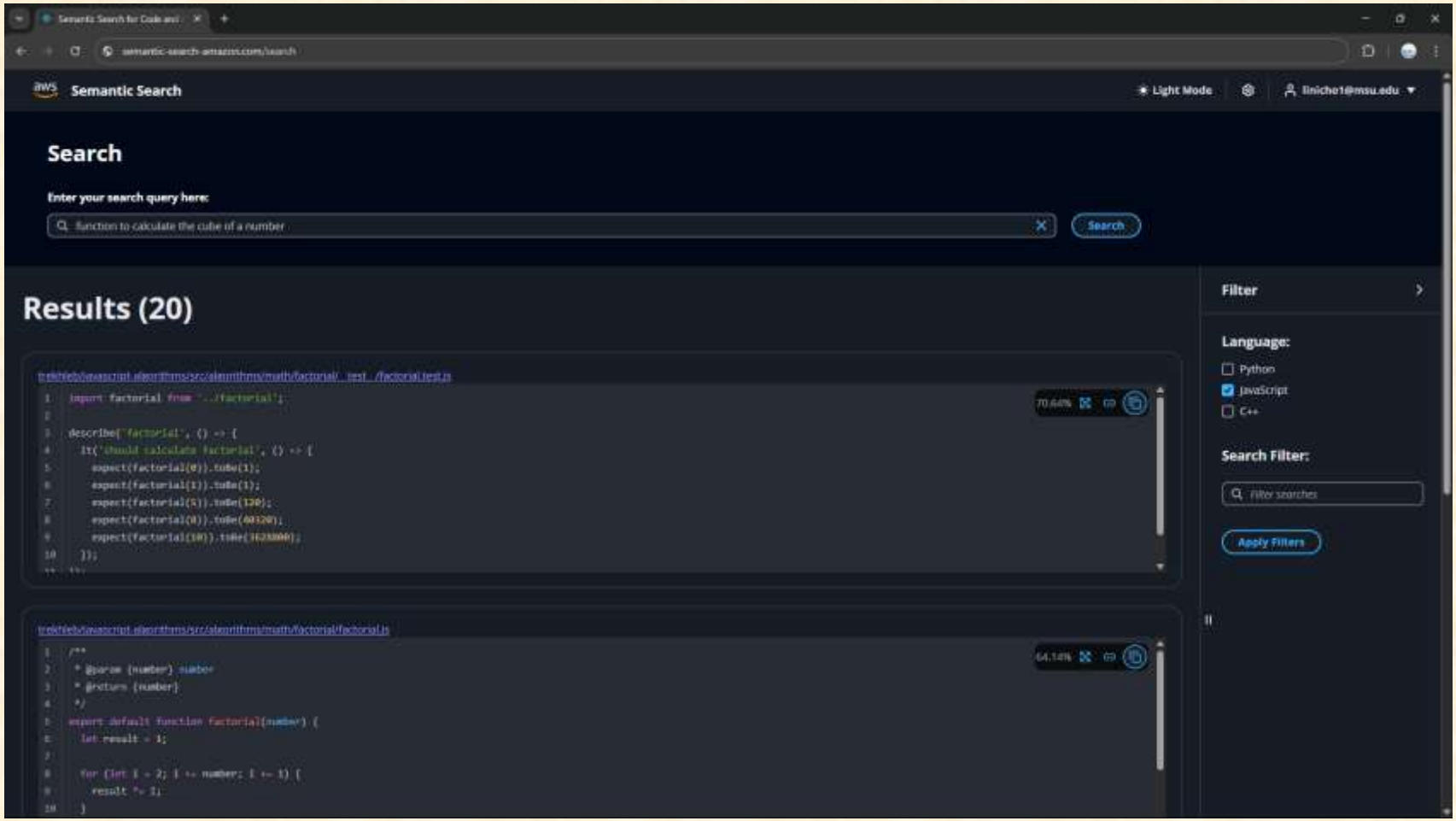
Home Page


Semantic Search
Light Mode
Hello, Welcome To Our Semantic Search Website
Here are some Popular Searches!
Popular Searches
function to calculate the cube of a number
this is the query
function to remove duplicate elements from a list
bubble sort
function to convert string to uppercase
OR Type here to begin searching!
Search
Enter your search query here:
Search...
Search

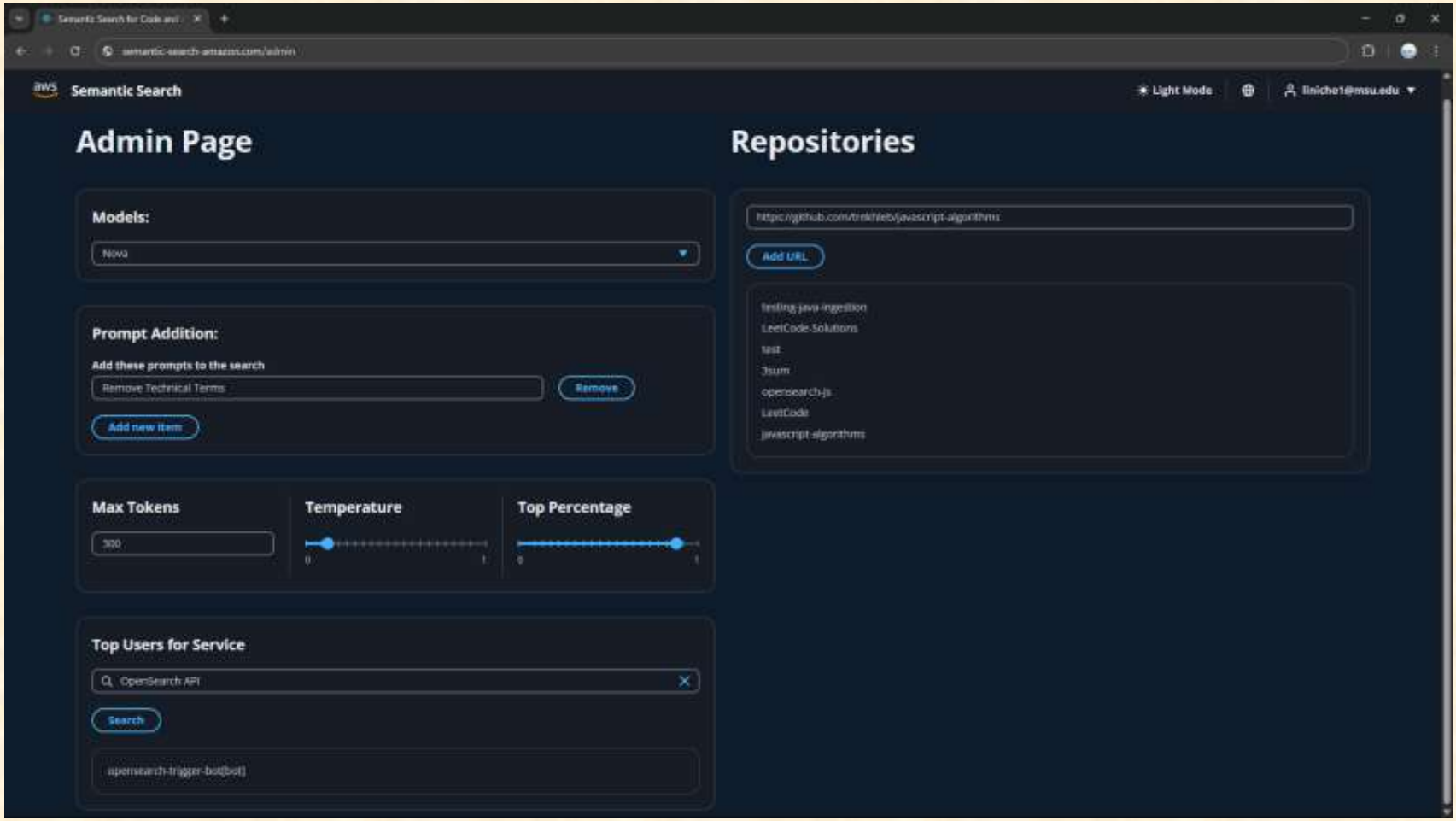
Results Page



Results with Filters Applied



Admin Page


Semantic Search
Light Mode
Admin Page
Repositories
Models:
Nova
Prompt Addition:
Add these prompts to the search
Remove Technical Terms
Remove
Add new item
Add URL
testing-java-ingestion
LeetCode-Solutions
test
3sum
opensearch-js
LeetCode
javascript-algorithms
Max Tokens
300
Temperature
Top Percentage
Top Users for Service
OpenSearch API
Search

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?