CSCI 3485 Final Project Proposal

Rafael Almeida

December 5, 2024

1 Problem Statement

Natural language search capabilities are becoming increasingly essential for modern web applications. However, implementing such features often requires relying on third-party services that are expensive, complex to integrate, and raise data privacy concerns. Developers, especially those working on smaller projects, need a more straightforward self-hosted solution to implement natural language search functionality.

2 Motivation

Current solutions in the market predominantly offer cloud-based services that require sending sensitive data to external servers. These solutions are often:

- Expensive and typically priced for enterprise-scale applications
- Complex to implement and maintain
- Limited in terms of data privacy and control
- Not optimized for smaller-scale applications

There is a clear need for an open source, self-hosted solution that prioritizes simplicity, privacy, and ease of implementation while maintaining robust search capabilities.

3 Main Task

The project will develop a Python package enabling developers to easily implement natural language search in their web applications. Key components include:

- Development of a streamlined PyTorch-based NLP model optimized for search functionality
- Creation of a clean, well-documented API for easy integration

- Implementation of efficient self-hosting capabilities
- Development of robust abstractions for model deployment and scaling
- Creation of comprehensive documentation and examples
- Building a demonstration application showcasing the package's capabilities

The final deliverable will be a complete, production-ready package that developers can easily integrate into their projects to add natural language search capabilities while maintaining complete control over their data and infrastructure.