Project Progress Report: Information Retrieval using Large Language Models through Question Answering for Machine Learning and Artificial Intelligence Research Papers

( CS 410: Text Information Systems , Fall 2023)

(Team: Lexical Wizard)

|  |  |  |  |
| --- | --- | --- | --- |
| Team : Lexical Wizards | | | |
| # | Name | ID | Email |
| 1 | Amrit Kumar | amritk2 | amritk2@illinois.edu |
| 2. | Gattu, Sudha Mrudula | sudhamg2 | [sudhamg2@illinois.edu](mailto:sudhamg2@illinois.edu) |
| 3. | Madhavan, Siddharth | sm120 | [sm120@illinois.edu](mailto:sm120@illinois.edu) |
| 4. | Ishika Awachat, Ishika | awachat2 | awachat2@illinois.edu |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | Versions | | |
| # | Version | Date | | Author | Update Details |
|  |  |  | |  |  |
|  |  |  | |  |  |

Table of Contents

[Introduction: 3](#_Toc150765773)

[Web Scraping: 3](#_Toc150765774)

[Text Embedding 3](#_Toc150765775)

[Question Answering using Language Model: 3](#_Toc150765776)

[Interactive Chat interface 3](#_Toc150765777)

[Summary: 3](#_Toc150765778)

[Appendix 4](#_Toc150765779)

# Introduction:

The rapid proliferation of research papers in the fields of Machine Learning and Artificial Intelligence has led to a substantial increase in the volume of textual information. This surge poses a significant challenge for knowledge seekers who wish to access specific information within this vast corpus. To address this issue, we propose a comprehensive project titled "Information Retrieval using Large Language Models through Question Answering." This project aims to leverage the power of modern NLP models to extract, index, and retrieve information from research papers in the domain of Machine Learning and Artificial Intelligence.

The aim of this document is to share the progress on this project implementation. This project has following main modules,

* Web scraping:
* Text Chunk generation/ embedding
* Question Answering Module
* Interactive Chat interface

# Web Scraping:

<Implementation Details>

# Text Embedding

<Implementation Details>

# Question Answering using Language Model:

<Implementation Details>

# Interactive Chat interface

<Implementation Details>

# Summary:

# Appendix