

## Project Proposal Document

### Theme 1: Intelligent Browsing

#### Semantic Search on a Coursera Page

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

**“Solo Soul”** is one team member team consist of my self.

I, Anjali Jain the captain and member of my team named Solo Soul.  
My id is [anjali9@illinois.edu](mailto:anjali9@illinois.edu) .

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

**Topic :**

Semantic Search on Coursera Page

**Problem :**

Coursera, being online learning platform, faces challenges in helping students efficiently find and access the most relevant and personalized educational content from its weekly lectures slides or contents.

Traditional search methods, which primarily rely on keyword matching, often fall short in interpreting the Student's / User's true intent and the contextual meaning of their queries. This results in a less than optimal user experience, as learners might get overwhelmed with irrelevant search results, potentially hindering their educational journey.

As a student I face this challenge while going through my weekly lectures.

Aim to achieve and solve from this project :

My project is first to build intelligent assistant which will retrieve information based on :

1. Understanding User Intent
2. Contextual Relevance
3. Improved User Satisfaction

**Relation to Theme 1 Intelligent Browsing :**

Semantic search is a cornerstone of intelligent browsing as it facilitates a more natural and intuitive interaction between the user and the information system.

**Reason of relationship :**

**1. Efficient Information Retrieval:** Intelligent browsing aims to make information retrieval more efficient and effective. Semantic search contributes to this by reducing the time users spend looking for the right courses or content, thereby improving the overall browsing experience.

## **Relation to Text Information Systems :**

Semantic search is deeply intertwined with text information systems, as it involves processing and understanding textual data to extract meaning and provide relevant search results.

1. **Information Retrieval:** At its core, semantic search is an advanced form of information retrieval, a critical component of text information systems.
2. **Natural Language Processing (NLP):** Implementing semantic search requires the use of NLP techniques to understand the context and meaning of queries, which is a significant area in text information systems.
3. **User Interaction:** Text information systems focus on how users interact with textual data. Semantic search enhances this interaction by making the search process more intuitive and aligned with human thinking. It will help in providing better search outcomes.

In summary, implementing semantic search on the Coursera page addresses the critical issue of information overload and the need for personalized, context-aware search capabilities, ultimately contributing to an enhanced and more efficient learning experience for users. It will also save time and efforts for students in search which is currently they spent with Traditional Search method.

3. Briefly describe any datasets, algorithms or techniques you plan to use

Technique I will be using for my project “Semantic search on Coursera Page “ is Text Embeddings:

**BERT or Sentence-BERT:** These are more advanced models based on transformers, capable of understanding the context of whole sentences and paragraphs, which can be particularly useful for interpreting complex search queries.

And Similarity Measures will be :

**Cosine Similarity:** Once you have vector representations of both the search query and the course content, you can use cosine similarity to measure how close or relevant they are to each other.

4. How will you demonstrate that your approach will work as expected?

Upon successful completion of the project, the functionality of the web page will be significantly enhanced, offering users a responsive experience when it comes to searching for information and asking questions

5. Which programming language do you plan to use?

I will be using Python as my programming language.

6. Please justify that the workload of your topic is at least  $20 \times N$  hours,  $N$  being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Task	Hours	Comment
Learning Phase:  1. Semantic Search 2. Fundamentals of Text Embedding	6 Hours	Considering my Novice level I allocated approx. 2 hours for learning & understanding this topics with respect to project development.

Task	Hours	Comment
Setting Up the environment	4 Hours	Allocated 4 hours to install necessary and required tools ,libraries ,BERT etc.
Learning to work with BERT / Sentence - BERT	8 Hours	Running BERT/ Sentence-BERT for Text Embeddings.  Understanding the Output and Tuning Parameters  Best Practices and Troubleshooting
Implementing the Semantic Search	16 Hours	Integrating Text Embeddings in Search Testing & Debugging
Evaluation & Fine Tuning	6 Hours	Setting Up Evaluation Metrics Analyzing and Interpreting Results. Fine-Tuning the Model
Documentation and Reporting	6 Hours	Documenting the Development Process  Preparing a Final Report and Presentation

As a Novice I assigned 46 Hours of efforts required for this project in order to grasp the concepts, implement the algorithm, and conduct thorough

testing and evaluation. Also document and prepare report & presentation for same.