UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

Department of Computing (B.Eng.) in Software Engineering

5COSC021C2: Software Development Group Project

Design and Documentation

Project LumosLore: Improving Modern Education: Seizing Challenges in Rapid Learning and Deep Understanding"

Team: SE-22

Mentor: Mr. Nishantha Janith

Student Name	IIT ID
K.Samahithan	20210863
R.H.Abilaash	20221589
R.Ramanapirian	20210714
W.Janiru	20220218
S.Shahana	20221349

Declaration

We hereby declare that the content of this report and all related contents are original work and have not been previously submitted or are currently being submitted for any academic programs.

Full Name: Samahithan Krishnaanantham

IIT ID / UOW ID : 20210863/ w1953875

Signature : 

Full Name : Ramanapirian Rajakumaran

IIT ID / UOW ID : 20210714 / w1957427

Signature : 

Full Name: Janiru Ransith Wickramage

IIT ID / UOW ID: 20220218 / w1998814

Signature : 

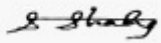
Full Name: Abilaash Radly Harison

IIT ID / UOW ID: 20221589/w1999509

Signature : 

Full Name : Shahana sivasubramanium

IIT ID / UOW ID : 20221349/w1956112

Signature : 

Abstract

The collaborative efforts along with expertise of our group in researching the neediness of knowledge evaluation. Our system helps to evaluate users' knowledge in specific text documents which users need to evaluate their knowledge. Our system goes through the text document which users feeded and extract the keywords to simply generate questions for analyse users knowledge in the respective document. By answering, the user gets feedback and summarization of users' own knowledge on the document. Through rigorous research, dataset analysis, and ethical considerations, we navigated the landscape of educational technology. The system provides an innovative method to track via user-oriented approach to learning evaluation and feedback output.

Acknowledgement

We would like to express our sincere gratitude to everyone who helped us to bring this project alive. During this whole process, the incredible team members have been genuinely crucial with their hard work and dedication.

We would like to thank our module leader Mr.Banu Athuraliya and mentor Mr. Nishantha Janith for their continued support and encouragement,which let us go in the right direction.

We extend our gratitude to IIT for their tremendous assistance and data resources were utilized in the development of this project.

Also we want to thank our families and friends for their continued motivations and patience during chaotic times we spent to bring this project together.

Despite everyone's support and encouragement, this project probably wouldn't have been achieved. We are grateful for the people who were with us this entire journey.

Table of Contents

List of Figures

List of Table

Abbreviations Table

Table 1: Abbreviation Table

Abbreviation	Explanation
AI	Artificial Intelligence
APA	American Psychiatric Association
API	Application Programming Interface
ASD	Autism Spectrum Disorder
BCS	British Computing Society
BERT	Bidirectional Encoder Representation from Transformers
CDC	Centres for Diseases Control Prevention
DL	Deep Learning
DSM	Diagnostic and Statical Manual of Mental Disorder
FAST	Families And School Together
GE	General Education
GLIDE	Guided Language to Image Diffusion for Generalization and Editing
GPT	Generative Pre-trained Transformer
HTTP	Hypertext Transfer Protocol
IEP	Individualized Educational Plan
JS	JavaScript
KPMG	Klynveld Peat Marwick Goerdeler
LR	Literature Review
ML	Machine Learning
NASS	National Autism Spectrum Disorder Surveillance System
NGO	Non-Governmental Organizations
NLP	Natural Language Processing
NLTK	Natural Language Took Kit
OOADM	Object Oriented Analysis Design Method

Chapter 1: Introduction

1.1 Chapter Overview

As technology becomes more and more integrated into education, this chapter explores a novel approach that aims to improve student involvement and comprehension. The opening establishes the scene by recognizing the move in education toward digital resources and highlighting the flexibility and accessibility that technology offers. Nonetheless, it acknowledges a recurring issue: pupils' difficulty using digital resources effectively, which frequently results in superficial comprehension. Our group's solution to this is a dynamic online application that aims to completely transform the process of teaching and learning.

The main objective of this project is to create and deploy innovative software that improves text document understanding in digital learning environments. The software analyzes, extracts, and prioritizes keywords and essential concepts from various document formats using natural language processing (NLP) approaches.

1.2 Problem background

Students struggle with a massive amount of lecture materials, little interaction from the instructor, and the difficulty of becoming independent in online classes. Finding important ideas quickly prevents in-depth comprehension and creates a delusion of mastery. Digital distractions make attention span constraints worse, which leads to problems with information retention. s. To achieve its goal of preparing students for future challenges, modern education must overcome three significant challenges. For clarification the background is divided into major problems as below:

1.2.1. Overwhelming Volume of Material:

Students find it very difficult to keep up with the large volume of material presented in lecture slides and modules. It can be overwhelming to students to read through every detail because of the enormous volume. As a result of having to prioritize, this overload hinders students' ability to learn and frequently leads to a cursory knowledge of the material (Xia, Y, et al 2022). Students who find it difficult to sort through this abundance of information run the risk of understanding the material only superficially. They might concentrate on superficial facts

rather than developing a deep comprehension, which could result in knowledge gaps that could obstruct their scholastic advancement. This problem goes beyond the classroom, impacting both the general standard of education and students' capacity to use newly acquired knowledge in practical contexts.

1.2.2. Lack of Focus in Online Learning or Lectures

When taking classes online, students find it challenging to stay focused when a professor isn't there in person. This lack of direction may cause students to get disengaged from the course material, which may make it more difficult for them to understand difficult topics and become proficient.

Students have a greater responsibility to manage their learning when instructors do not actively participate in the classroom. This independence could make it more likely that there will be miscommunications and comprehension problems(Bradbury,2016). Students who take greater ownership of their education are less likely to overlook important lessons and have difficulty successfully integrating material, which will eventually affect the standard of instruction as a whole

3. Difficulty in Identifying Key Concepts:

Problem: The difficulty of identifying important terms and concepts in a large amount of data makes it difficult to have a thorough understanding. It could be difficult for students to extract the fundamental concepts that underpin understanding, which would make it more difficult for them to make coherent connections between ideas. There is a chance that this identification challenge will lead to a disjointed comprehension of the material. Even when students identify general issues, it becomes difficult to identify concrete solutions. This difficulty raises the possibility of misunderstandings and application errors in addition to impeding the development of a thorough understanding. As a result, learning may be hampered as students struggle to close the knowledge gap between theory and real-world problem-solving.

Limited Time Availability:

Students' time to study lecture slides is limited by the rigorous nature of a competitive atmosphere and involvement in extracurricular activities. Students are put in a risky scenario because of the small amount of time allotted for reviewing lecture slides. Stress and exhaustion might result from trying to strike a balance between extracurricular activities and academic obligations. Students may find themselves unable to study subjects in-depth as time becomes a limited resource, which could jeopardize their academic performance and depth of comprehension. Beyond personal stress, the effects also impair the standard of education as a whole and students' capacity to fully comprehend and apply knowledge in a variety of circumstances.

Problem boundary

Project LumosLore explores the difficulties students have in the conventional educational system. It includes difficulties that students face when learning quickly, such as finding the right keywords and concepts in study resources like lecture slides and notes and generating precise questions. The project focuses on students who face these obstacles at all educational levels, from elementary school children to postgraduates, in an effort to improve their educational experiences and make them perform well in exams

Examples in problem

1. One issue facing undergraduate medical students is that they often lack the time to thoroughly review lecture slides and understand important and minute terminology after returning from clinical.

2. After fieldwork, engineering students find it extremely difficult to read LEC slides.

3. Students studying computer science frequently encounter difficulties following coding assignments or hands-on training. They struggle to go over in-depth class slides and comprehend complex programming topics.

4. Students of Language and Literature after taking part in language immersion exercises or creative writing workshops, literature and language students find it difficult to review in-depth literature notes and linguistic topics.

5. Elementary School Pupils It can be difficult for elementary school students to review and comprehend class materials, particularly basic language and math concepts, especially after participating in hands-on activities or outdoor play.

So overall students find it challenging to focus on reviewing foundational themes due to their restricted time and attention span, which may impede their basic comprehension and assessment preparedness.

Therefore, it would be advantageous if a program used automated processes to improve text document understanding. By utilizing sophisticated algorithms and natural language processing, it examines diverse document types, retrieves essential data, and creates inquiries to gauge comprehension by the user. By pointing out areas of strength and weakness in their knowledge, the feedback system gives users more authority and encourages deeper understanding. With its coverage of Word, PDF, and plain text, the application promises to boost information retention and absorption while providing a transformative learning experience. The suggested software presents itself as a comprehensive solution for a variety of learners by combining document analysis, question development, and feedback methods.

1.3 Problem Statement

The problems of deep knowledge and quick learning in education are addressed by Project LumosLore. It addresses problems like determining important keywords and creating targeted study questions from study materials. Through examining issues in conventional education, the project seeks to identify useful tools, methods, and strategies to improve student's learning results while maximising test-taking and progress monitoring.

1.4 Proposed Solution

The program that is being envisioned seeks to improve text document comprehension using automated procedures. It functions by closely examining textual content, identifying relevant themes and keywords, and then creating questions based on this research. By giving users this information and allowing them to access how effectively they comprehend the key points in the written material, the tool aims to improve retention of knowledge and create a more dynamic and effective learning. From a range of document formats, the software detects and retrieves essential data using cutting-edge algorithms and natural language processing

techniques. These omitted segments serve as a base for questions designed to assess how effectively the user understands the document's major concepts and features (DeepLearning.AI,2023).

Furthermore, the program does more than just ask questions. It includes an integrated survey system that helps users to evaluate their own level of understanding on the document. This approach provides users a better grasp over their deficiencies in knowledge and areas of strength, which leads to a great comprehension of the document's content.

All things considered, this program has the power to completely transform learning paradigms through the use of automated analysis, question creation, and feedback systems. Through its use, a more engaging and effective learning experience is promised in terms of improved information assimilation and retention.

A. Text Document Analysis:

The program takes in many types of documents, including Word, PDF, and plain text. It uses a powerful text analysis engine to thoroughly scan and understand the information contained in these papers.

B. Keyword and Important Word Extraction:

The text is tokenized by the system, which uses sophisticated Natural Language Processing (NLP) algorithms to divide it up into understandable components. It then separates regular words from important terms using entity recognition and part-of-speech tagging. In order to identify key terms, the software also examines word frequency and significance(GeeksforGeeks,2021).

C. Question Generation:

The software generates questions automatically based on the retrieved keywords and relevant words. These questions are intended to assess the user's comprehension of the main ideas and important details in the document.

D. Feedback and Scoring:

The program provides thorough performance feedback to the user after they respond to questions. This feedback identifies areas for improvement, assesses the reader's general comprehension of the material, and may even offer a score or rating.

E. Progress Tracking and Feedback Providing:

A progress dashboard allows users to monitor their progress over time. To fill in any knowledge gaps that are found, this dashboard may offer recommendations for more research as well as visual aids like charts and data.

1.5 Aim

We aim to change digital learning by creating an innovative software solution that dramatically improves text document comprehension. The software intends to automate the study of digital texts by extracting essential concepts and producing targeted queries by incorporating advanced natural language processing (NLP) techniques. The ultimate goal is to develop an adaptive and user-friendly learning tool that fosters a dynamic and engaging learning environment while also enabling people to absorb text content more efficiently. Through customization of questions based on user skill levels, provision of comprehensive feedback, and provision of progress monitoring tools, the project seeks to close the gap in the efficient use of digital learning resources. This goal is in line with the larger vision of enabling a paradigm shift in the field of education by guaranteeing that students from a variety of backgrounds may fully utilize digital resources for improved understanding and learning results.

1.6 Project Scope

Basically, LumusLore enables learners to learn independently via personalized evaluation. LumusLore allows users to submit a variety of documents, it creates customised questions right inside the text. Learners respond directly on the platform, and they get thorough feedback that identifies their areas of strength, points out information that is missing, and offers evaluative advice. LumusLore helps people on their learning journey toward a greater understanding and a sense of accomplishment by praising their accomplishments and providing clear metrics to track their progress.

1.6.1 In-scope

- By allowing users to upload several documents at once in standard file formats like Word, PDF, and TXT. LumusLore simplifies the assessment process. This helpful feature ensures an adaptable and inclusive user experience.
- The cutting-edge in-document question-generating technology from LumusLore improves self-assessment by automatically extracting important keywords from uploaded documents and creating appropriate questions that are in line with the particular content.
- Ability to promote deep learning in its setting, which enhances self-evaluation. Users can answer questions that are created within the platform without interruption or by switching between apps.
- Self-evaluation is elevated by LumusLore to individualized feedback that goes beyond simple right or incorrect. It explores further, highlighting advantages and disadvantages as well as information gaps. With the help of strong NLP models
- Users can maintain track of their learning progress with LumusLore's Progress Tracking function. By using insightful measurements and performance data visualization, learners can recognize their areas of strength, celebrate their accomplishments, and maintain motivation.

1.6.2 Out-scope

- As LumusLore emphasizes a stand-alone platform that is user-accessible, it is not currently designed to integrate with Learning Management Systems (LMS) or other third-party educational platforms. This intricate feature necessitates a large amount of development work and might not be necessary for the initial release.
- To reach a wider audience, LumusLore now concentrates on providing its services in English. The multilingual expansion calls for more NLP model development and localization work, which demands more resource spending. After the essential English features are established and the demand from users for additional languages is evaluated, this feature can be taken into consideration in later phases.
- At this time, there is no plan to develop a mobile app for LumusLore because the main focus is on creating an intuitive web platform. Platform compatibility and user interface design demand different resources and considerations during the creation of mobile

apps. To improve accessibility and reach a larger audience, LumusLore may investigate mobile app development in the future based on consumer input and industry trends.

- A chatbot for direct inquiries from users currently not included in the system. A chatbot system with advanced natural language understanding might be added in later stages to allow for real-time communication. Although not included in the first version, contextual learning, follow-up questions, and user-guiding features might be investigated in the future.

1.7 Rich Picture Diagram

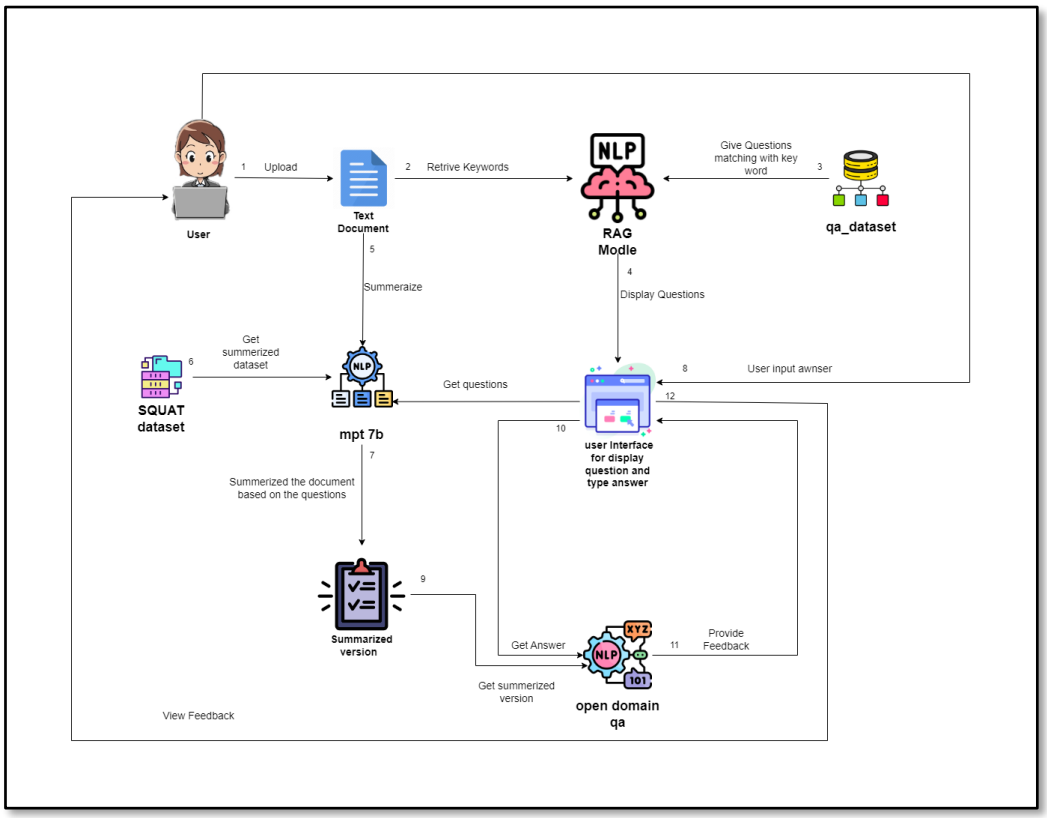


Figure 1 Rich Picture Diagram

1.8 Resource requirements

1.8.1 Hardware requirements

Table 2 Hardware Requirement

Devices	**Specifications**
Apple MacBook Air 13” (Primary development system): MacOS offers superior control over the Unix-based system environment and greater performance with development tools.	**MacO Sonoma 14.1.1, Apple M2 processor, 8GB RAM, 512GB SSD**
Windows Laptop (Secondary development system) - To store big datasets and execute Windows-compatible software as needed.	**Windows 11, HP intel i7 10th gen processor, 8GB RAM, 512GB SSD, 1TB HDD**
SLT FIBRE 100mbps / SLT 4G Router(Internet connection)	**Prolink PRN3005L**

1.8.2 Software requirements

The following languages, IDEs, additional software, APIs, and libraries were found to be necessary for the project's effective completion during the preliminary investigation.

Table 3 Software Requirements Languages, IDEs

Languages	
Python	For the backend of ML and NLP.
HTML, CSS, JavaScript, React	for web Application development.
IDEs and Other Software	
PyCharm	For Python programming using OOP.
Figma	For User Interface Designing.

Star UML	For diagrams creation
Draw.IO	For creating UML diagrams.
Google Doc	To create the reports and additional documentation.
Adobe Photoshop	To create and update wireframes and pics.
Google Drive	For cloud-based data and document management.
Jira	For project management.
Git	For managing versions.

Table 4 Software Requirements APIs and Frameworks

APIs and Frameworks	
Natural Language Processing (NLP)	To analyze text inputs to get keywords
Machine Learning (ML)	For develop models for question generation and feedback generation
Educational APIs	To provide additional assessment contents
Authentication and Authorization APIs	To provide secure user access
Cloud Services	To store data
Content Management Systems (CMS)	Construct question repositories and handle educational materials

1.9 Business model canvas

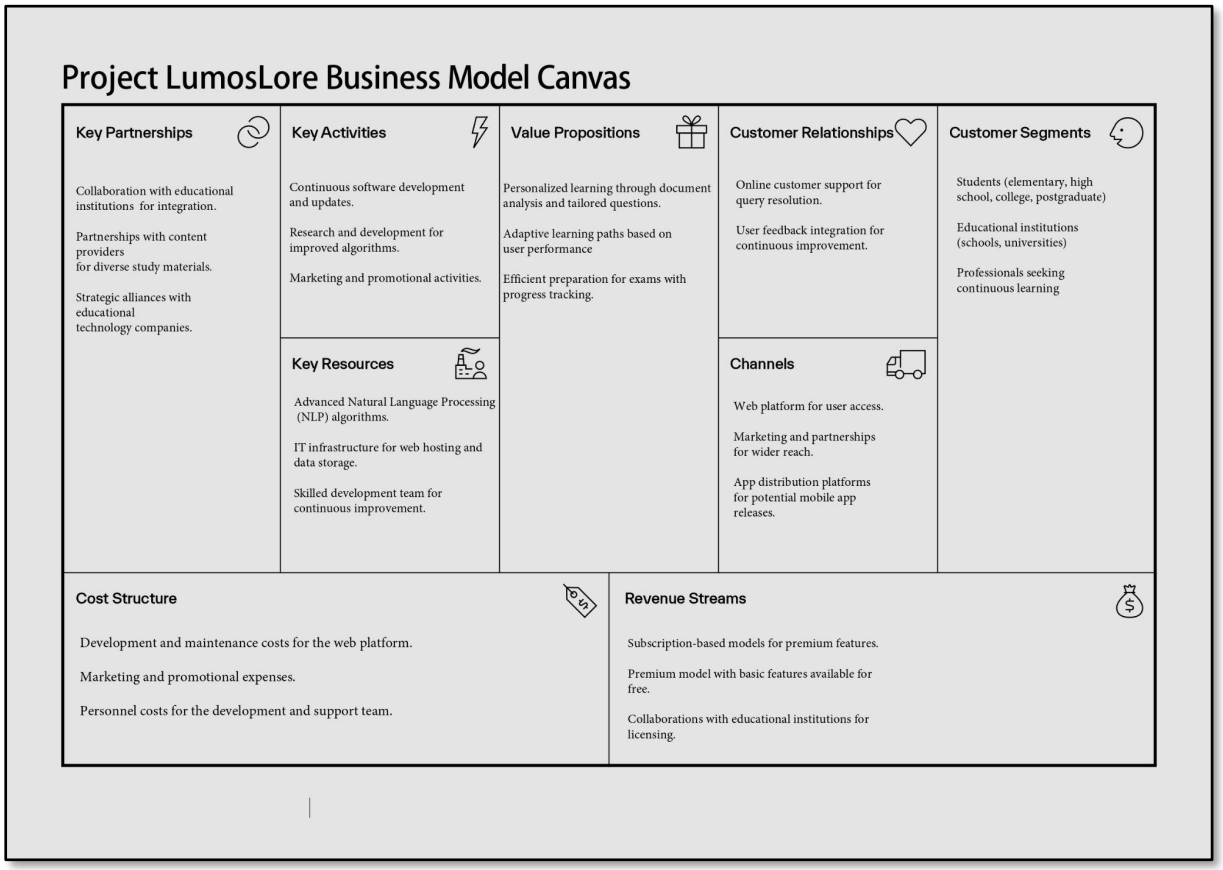


Figure 2 Business Model Canvas

1.10 Chapter Summary

This chapter provides an overview of the technologies and approaches that can be used to assess students' comprehension of the study material. The emphasis is on resolving issues that these students frequently face, such as having too many study materials and not enough time to be ready for exams. The principal objective is to improve their understanding of academic resources. The chapter provides crucial background information for the educational endeavor while diving deeply into a thorough analysis of the problem domain. It describes the precise aims, objectives, and extent of the project and clarifies the necessary resources needed for its effective execution. The next chapter will examine previous research and methods on the subject in great detail, concentrating on the methods chosen for the Lumoslore e-learning project.

Chapter 2: Existing Work

2.1 Chapter Introduction

In this chapter, we explore the existing landscape of solutions within the domain of educational technology, with a focus on systems and products that share similarities with projects LumosLore. Understanding the strengths and weaknesses of competitors' offerings is crucial for refining LumosLore's features and positioning in the market. This chapter briefly introduces the significance of studying existing work and lays the groundwork for a comprehensive review of competitor solutions.

2.2 Existing Work

2.2.1 Similar Products and Technologies

This section provides a detailed evaluation of the proposed solution, comparing its strengths and weaknesses with existing products. The comparison is presented through a product feature chart, which outlines how our solution fares against other products in the market that offer similar features.

2.2.2 Competitor Solutions and Features Comparison

In assessing the competitive landscape, we note that Exam Time distinguishes itself by enabling student participation in answering questions, thus promoting interactive learning. StudyBlue and Quizlet, meanwhile, focus on question generation, providing users with a diverse array of practice questions that support a self-tailored study experience. Edmodo offers a comprehensive educational platform, though it doesn't specifically target the features highlighted by the other competitors. This comparative analysis reveals that while there are robust solutions in the market, there remains an opportunity for LumosLore to introduce enhanced functionalities that address the gaps left by these products. Detailed feature comparisons will aid in understanding the market better and will be instrumental when showcasing LumosLore in competitive scenarios.

2.2.3 Technology Approaches Algorithm Reviews

Algorithms for machine learning and natural language processing (NLP) constitute the core of competing platforms' technology infrastructures. Among the key techniques used in textual processing and meaningful content extraction are tokenization, named entity recognition, and part-of-speech tagging. Competitors also use algorithms for classification and clustering for content summarization in an attempt to improve user learning experiences. These techniques are part of LumosLore's methodology, but it goes one step further with sophisticated algorithms—possibly incorporating in-house AI models—that enhance the engagement and relevance of the material. LumosLore's documentation establishes an adequate basis for current ed-tech tendencies as well as presents the initiative's inventive edge by exploring furthermore into these technological components further.

2.2.4 Benchmarking, and Evolutionary Matrics

In order to assess a tool's efficacy and effectiveness on learning, benchmarking is a crucial phase in the assessment of educational technology. Rivals are evaluated using a range of measures, including user involvement, material delivery accuracy, and learning outcome improvement. Clear standards are crucial for LumosLore, and the project will use both cutting-edge and industry-standard measures to assess its systems. These indicators could include the quality of generated feedback, the accuracy of question-answer matching, and user retention rates. By recording these criteria, LumosLore will demonstrate its dedication to empirical evaluation and ongoing development while also guaranteeing adherence to high standards.

2.3 Tools and Implementation Plan

The following techniques must be used to create LumosLore that pulls keywords from certain content, creates questions based on that information, gives students a chance to respond, and compares the responses to provide feedback.

1. Data collection

2.Preprocessing

3. Keyword extraction.

4. User Interface Design

5. Answer evaluation

2.3.1 Data collection

Through cutting-edge capabilities including question creation, question answering, text generation, text categorization, text summarization, and token classification, Lumosmore, an all-inclusive educational platform, seeks to improve learning experiences. Lumosmore makes use of cutting-edge datasets and research to enable these features. For various question-answering tasks, the platform takes into account datasets such as SQuAD, NewsQA, HotpotQA, and Natural Questions. Datasets like ARC.MCScript, OpenBookQA, and others are useful for assessing machine understanding and reasoning abilities in educational contexts. Rich resources can also be found in custom datasets from learning platforms like Kaggle's question-answer dataset

question generate

https://paperswithcode.com/datasets?task=question-generation

question answering

https://paperswithcode.com/datasets?task=question-answering&page=1

text generation

https://paperswithcode.com/datasets?task=text-generation

text classification

https://paperswithcode.com/datasets?task=text-classification&page=1

text summarization

https://paperswithcode.com/datasets?task=text-summarization&page=1

token classification

https://paperswithcode.com/datasets?task=token-classification

tools for data collection

1. **SQuAD (Stanford Question Answering Dataset):**
    - SQuAD
2. **NewsQA:**
    - NewsQA:
3. **HotpotQA:**
    - HotpotQA:

4. **Natural Questions:**
    - Natural Questions

OpenBookQA:

OpenBookQA:

1. **ARC (AI2 Reasoning Challenge):**
    - ARC:
2. **MCScript:**
    - MCScript:

3.Custom Educational Platforms:

Kaggle

https://www.kaggle.com/datasets/rtatman/questionanswer-dataset

2.3.2 Preprocessing

When study material is supplied, a comprehensive pipeline for preparation is set up to ensure that the text data is consistent and suitable for additional analysis. Lowercasing, or changing all text to lowercase, is the first step toward establishing uniformity. Tokenization is then used to break up the study material into manageable chunks, making analysis simpler by arranging the content. The removal of stop words further refines the text by eliminating common but uninformative keywords and focusing more on words that add content. Special letters and punctuation are removed to distinguish important words, producing a cleaner and more relevant

dataset. The next step is to apply lemmatization or stemming to maintain linguistic accuracy and computational efficiency throughout the upload process.

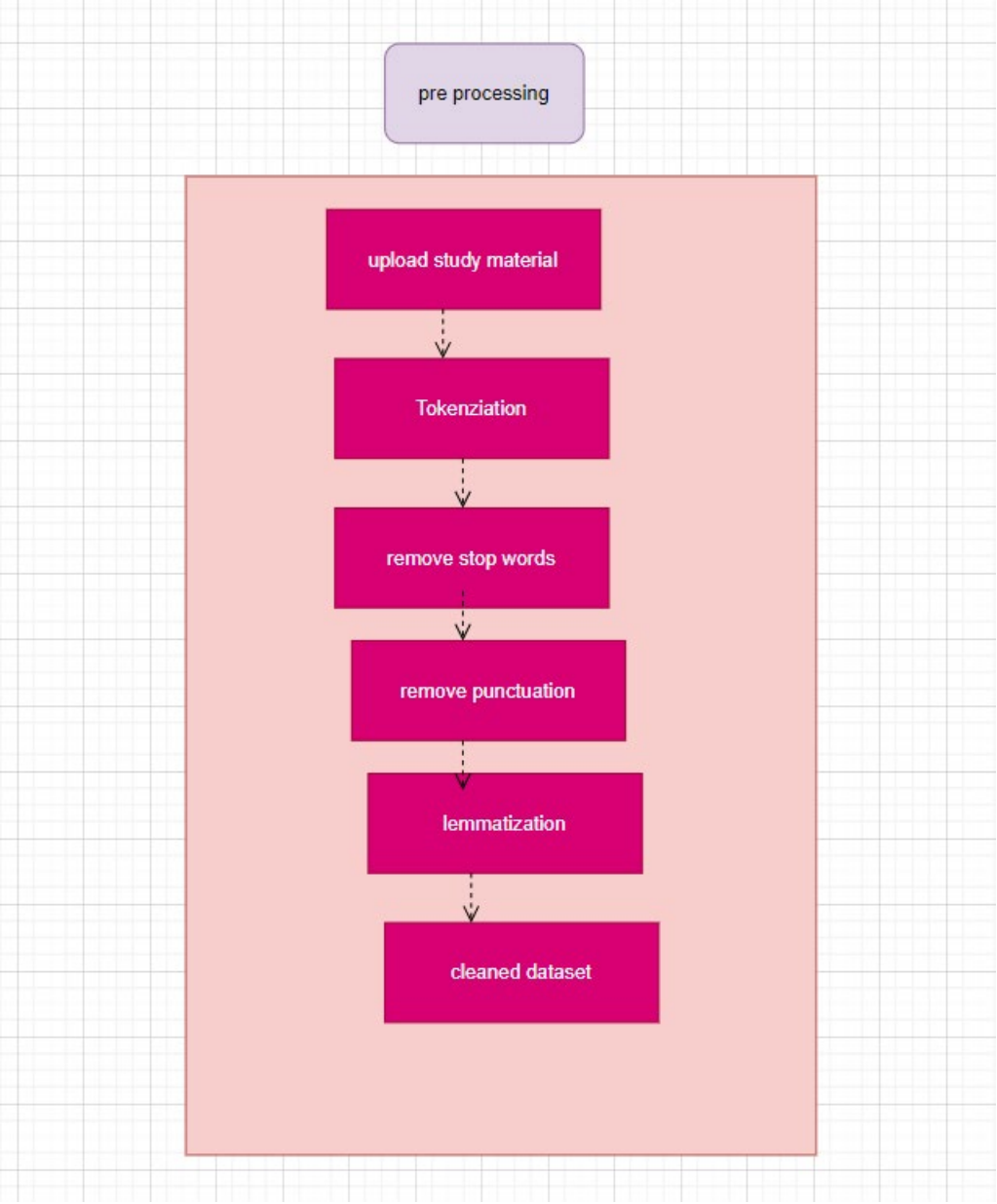


Figure 3:Preprocessing Diagram

Tools for preprocessing

1. **NLTK (Natural Language Toolkit):**
    - **Description:** NLTK is a comprehensive library for natural language processing in Python. It includes modules for tokenization, stemming, lemmatization, stopword removal, and more.
    - **Link:** NLTK

2. Scikit-learn is a machine-learning library for Python. It includes tools for text feature extraction, such as TF-IDF vectorization, which is commonly used in text preprocessing.

    - **Link:** https://scikit-learn.org/stable/modules/preprocessing.html

3. **Transformers (Hugging Face):**

    https://huggingface.co/docs/transformers/preprocessing

2.3.3 Keyword Extraction

The cornerstone and heart of natural language processing is keyword extraction, which finds extensive use in knowledge mapping, text summarization, text categorization, and related domains (Li and Ning, 2022). Keyword extraction is the process of extracting certain words or phrases from a document that best fits its theme. A document is simply a collection of words. With the proliferation of information available today, automatic keyword extraction from text can help users find the primary material of a document quickly and effectively.

1. **TF-IDF Principal Analysis**

    By multiplying a word's term frequency (TF) by the inverse document frequency (IDF), TF-IDF evaluates a word's importance in a document. Words exclusive to an article are indicated by high TF-IDF values, which facilitate classification. Word frequency in a document is measured by TF, whereas a word's overall relevance across papers is determined by IDF.

$$w_t = tf \times idf = tf \times \lg\left(\frac{N}{n} + 0.01\right)$$

Figure 4 TF-IDF Evaluation Formula (Schuh, G. et al,. 2018)

2. **Retrieval Augmented Generation (RAG)**

    Retrieval Augmented Generation (RAG) improves keyword extraction. Retrieval models are first used to obtain pertinent documents. After that, a generation model uses the text it has retrieved to generate clear

Figure 5 Process of RAG (Takyar, A., 2023)

and useful keywords (HeidiSteen, 2023) . This method enhances the relevance and accuracy of extracted keywords by fusing the expressiveness of generation with the precision of retrieval.

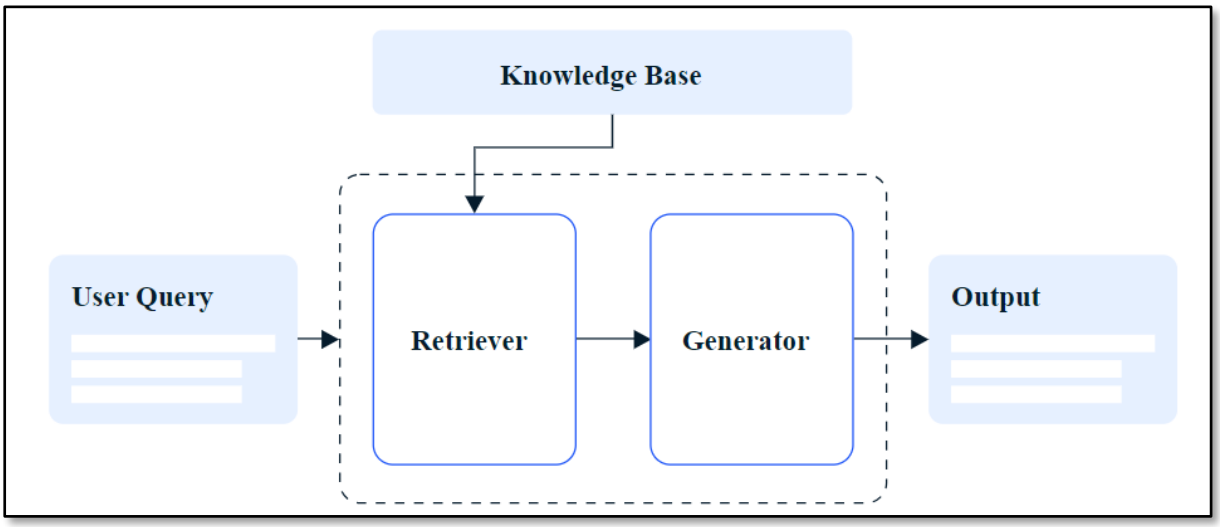


Figure 6:Retrieval Process

Retrieval Process

- **Data sources:** RAG begins with gaining access to external data sources, such as websites, APIs, databases, documents, and other structured information repositories. These data sources could provide a wealth of information, including current statistics and knowledge related to the topic.
- **Chunking:** These sources frequently produce data that is too big to handle all at once. It is therefore divided into smaller, easier-to-manage chunks. Each chunk can be considered a self-contained unit that represents a portion of the data.
- **Conversion to vectors:** Next, each chunk's text is transformed into vectors, which are numerical representations. Numerical sequences called vectors are used to represent the text's semantic significance. The computer can now comprehend the connections between the ideas in the text thanks to this conversion.
- **Metadata:** Each time the data is processed and divided into chunks, metadata is generated and linked to it. The source, the context, and other pertinent

information are all contained in this metadata. It is employed as a source and citation.

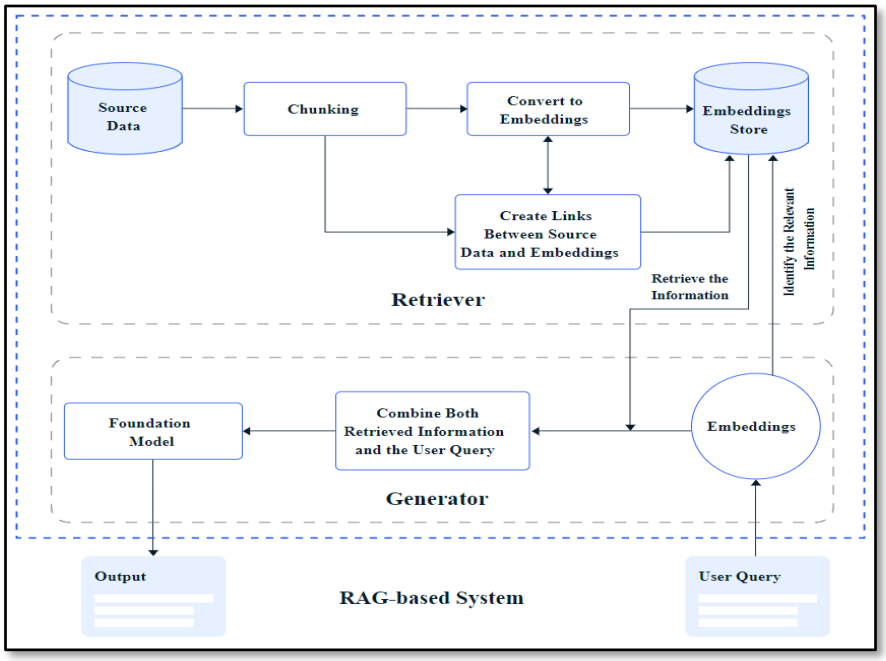


Figure 7 Retrieval Process Model (Takyar, A., 2023)

2.3.4 User interface design

The collaborative UX/UX design of the LumosLore educational software project is made possible via Figma. Its cloud-based platform enables stakeholders, engineers, and designers to collaborate in real time. The interactive mockups that can be created with Figma's prototyping tools enable the testing of user interactions and guarantee an easy-to-use learning environment. A consistent user experience across a variety of devices is ensured by the tool's support for responsive design. Using component libraries makes design coherence better and management easier. Furthermore, early feedback is made possible by Figma's user testing capabilities, which promote an iterative design process that improves LumosLore's user interface based on actual interactions. Figma is, all things considered, an invaluable tool for creating a user-centered, responsive, and cooperative learning environment.

Principal user interfaces to be present.

Login/Registration

Dashboard

Upload Study Materials

: Interface for Questionnaires

Feedback & Report Display:

Settings:

2.3.5 Awnser Evaluation

Several complex NLP techniques and frameworks will be used in an intricate way to create LumusLore's Answer Evaluation model. A flexible open-source machine learning framework called TensorFlow will be used to train and implement the question-answering models. LumusLore's QA system has a strong basis thanks to its extensive variety of natural language processing (NLP) capabilities (Otten, N.V., 2023), which include sentiment analysis, language translation, and text production.

Modern transformer-based models, such as BERT and GPT-3, will be crucial (Höhn, F., 2022). These models, which have been pre-trained on large text corpora, perform exceptionally well on a number of NLP benchmarks. Developers guarantee that the models perform well in the complexities of answer evaluation tasks by fine-tuning them on particular datasets customized to LumusLore's environment.

The open-source platform from Hugging Face is a great tool for combining and perfecting these models. Its library of pre-trained models makes implementation easier and enables us to quickly adjust and optimize for LumusLore's particular needs.

Popular open-source library SpaCy will handle named entity recognition, tokenization, and lemmatization to improve text processing skills. Furthermore, the Python libraries NLTK and OpenNLP, which are well-known for their extensive NLP features (Lokare, G., 2023), will enhance LumusLore's Answer Evaluation model by helping with text pre-processing chores.

This combination of strong frameworks and instruments guarantees a smart, precise, and perceptive method of assessing student responses. LumusLore seeks to offer a smooth and efficient learning experience, encouraging a deeper comprehension of the subject matter for its users by utilizing the advantages of these NLP techniques.

2.4 Chapter Summary

This chapter briefly discusses the existing work done in the field and analysis of the present state of the field. It gives an in-depth understanding of similar products and technologies and tools and implementation plan which deeply covers the subtopics Data collection, Preprocessing, Keyword extraction, User Interface Design, Answer evaluation, Machine learning model and under preprocessing section the author give deep explanation about introducing NLTK, Scikit-learn, and Transformers in the project LumosLore. Throughout the chapter, the author ensured that the results of the study and their own analysis and reviews commended both well by combining them with ease.

Chapter 3: Methodology

3.1. Chapter Overview

This chapter outlines the systematic approach adopted in the project's development, encompassing development, design, and project management methodologies. It begins with an overview, setting the stage for the detailed exploration of the development methodology, design methodology, and project management methodology. The teamwork breakdown structure (WBS) is introduced to delineate the distribution of tasks among team members, fostering collaboration and efficiency. A Gantt chart diagram further visualises the project timeline and takes dependencies. In addition, the chapter highlights the strategic use of project management and collaboration software to streamline communication and enhance coordination among team members through the project lifecycle. Overall, this chapter serves as a comprehensive guide to the methodologies and tools employed, ensuring a structure and effective approach to project execution.

3.3. Development Methodology

A crucial aspect that influences the entire software development process is the selection of a development approach. It offers an organized structure that directs the planning, carrying out, and overseeing of tasks during the project. Different approaches provide different techniques and philosophies (Sharma, 2015). Agile, Scrum, Kanban, Feature-Driven Development, and DevOps are some of these. Agile development approaches, like Scrum and Kanban, emphasize flexibility and iterative development (Albers et al., 2019), which promotes teamwork and response to shifting needs. While DevOps unifies development and operations for continuous delivery, feature-driven development concentrates on delivering certain functionalities. In contrast, the Waterfall model employs a sequential approach, requiring the completion of each phase before proceeding to the next. Various aspects such as project size, complexity, and degree of requirement uncertainty influence the choice of development methodology, which guarantees a customized and efficient approach to software development.

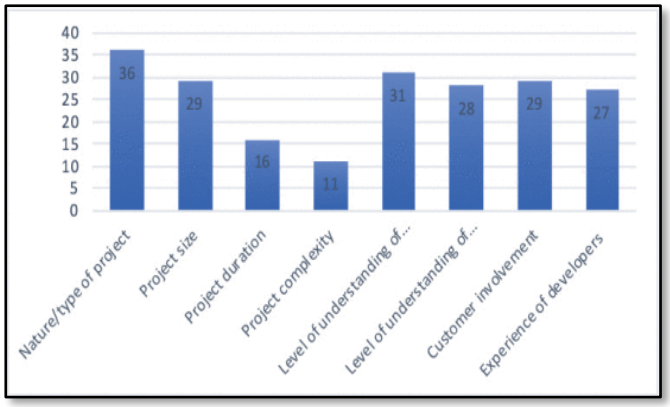


Figure 8 Things need to consider for the project management (Cooper, R.G., 1990)

Agile and Waterfall are two well-liked project management techniques. A more conventional method of project management that follows a linear flow is called waterfall. Agile, in contrast, values iterative development. It calls for quick thinking and extreme adaptability.

Development by waterfall is a sequential process that follows a linear path, Planning is the first step in the sequential process of waterfall development, which moves forward via requirements analysis, design, implementation, testing, deployment, and maintenance. While its methodical approach works well for projects with well-defined requirements (Mushashu and Mtebe, 2019), its rigidity causes problems in situations that are dynamic. Misunderstandings could result from the slow reaction to changes and the incomplete incorporation of client input. Notwithstanding its shortcomings, Waterfall is nevertheless relevant in situations where consistency and a clear plan are more important than adaptability. For these kinds of situations, it is appropriate.

The agile development process emerged as a flexible and adaptable alternative to the traditional Waterfall methodology. Agile places a high value on adaptability and working together with customers to enable iterative development cycles that adjust to changing requirements (Schuh et al., 2018). Agile promotes continuous improvement as opposed to Waterfall's Sequential structure, enabling frequent reevaluation and modifications throughout the development process. Its iterative design facilitates the delivery of functional increments more quickly, encouraging frequent and early client input. The Agile approach is replacing the traditional Waterfall paradigm in modern software development processes because it offers a flexible and customer-focused framework and performs well in contexts of dynamic or unknown needs (Baschin, Inkermann, and Vietor, 2019).

Table 5 Waterfall vs Agile

Aspect	Waterfall	Agile
Life Cycle	A linear sequential model	Methodology that utilizes continuous iteration
Rigidity	Developed using strict, hierarchical model	Can be built flexible
Collaboration	Least flexible model, which follows a set of processes and prevents team collaboration	Highly collaborative model which produces superior results
Process	There are various stages to the development process	There are sprints throughout the entire development process
Changes	Once the development commences, this prohibits the changes	Modification can be made even after the preliminary planning is finished
Software Development	Development finished as a single project or deliverable using this model	Development process can be a collection of numerous projects or deliverables
Testing	Testing phase follows the build phase	Testing can happens during the same iteration as developing or programming

By choosing the Agile technique for LumusLore, the end user guarantees an agile and constantly adapting software development process. The advantage of agile's iterative cycles, inside each sprint, is that problems can be found and fixed quickly. Throughout the development process, this user-centric approach places a high priority on testing and communication, creating a dynamic environment that improves the project's overall efficiency and product quality.

3.4 Design Methodology

Methodical methods of problem-solving are among the design techniques that are necessary for the development of successful solutions. Two noteworthy methods under this paradigm are the **Structured Systems Analysis and Design** (SSAD) and the **Systems Analysis and Design Methodology** (SADM).). Software Architecture and Development Methodology (SADM) ensures extensibility, maintainability, and stability in software systems.

Object-Oriented Analysis and Design (OOAD), which uses concepts from object-oriented programming to system analysis and design, is another significant paradigm. OOAD models systems through the use of objects and their interactions, emphasizing flexibility over traditional functions and logic. The process includes steps such as object identification, class definition, relationship and interface definition, and implementation.

Advantages

Table 6 Advantages OOAD vs SSAD

Aspect	OOAD	SSAD
Flexibility and Reusability	promotes modular design, which improves code component reusability.	encourages orderly, structured design, which makes code reuse easier.
Maintenance	Easily adjustable to changes, which facilitates easier management of updates and maintenance.	Maintenance is made easier with a structured strategy, especially for small to medium-sized enterprises.
Encapsulation and Abstraction	promote encapsulation, which reduces complexity by keeping implementation specifics hidden.	highlights abstraction to conceal implementation specifics, which helps to control system complexity.
Complexity Management	uses the concepts of object-oriented design to control	uses elements from procedural programming to

	system complexity.	offer an organized method of handling complexity.

Disadvantages of OOAD and SSAD

Table 7 Disadvantages OOAD vs SSAD

Aspect	OOAD	SSAD
Learning Curve	greater learning curve for programmers who are not familiar with object-oriented ideas.	usually has a smaller learning curve, making it more approachable for those who are familiar with procedural programming.
Project Size	Could add overhead for small projects because of the extra abstraction and structure..	For larger projects, it might become cumbersome and complex, which would affect maintainability.
Performance	Performance is a little bit worse because of object-oriented features	Because of its more straightforward structure, it typically performs better for smaller tasks.
Adaptability to Legacy Systems	can encounter difficulties when utilizing procedural design to integrate with legacy systems	Proven interoperability via procedural design with legacy systems.

The rationale behind the adoption of object-oriented Analysis and Design (OOAD) as the development technique for LumosLore is its customized fit for the needs of the educational application. The focus placed by OOAD on developing systems that are simple to extend and maintain aligns with how educational technologies are developing. Since LumosLore might eventually need to be updated, or improved, OOAD offers a strong framework for methodical analysis, design, and execution. The method's structured approach makes it easier to define

classes, connections, interfaces, and objects. It also helps create a cohesive implementation strategy, making LumosLore fast, user-friendly, and flexible enough to meet changing educational needs.

In conclusion, OOAD's natural compatibility with the complexities of educational software, its capacity to build adaptable and maintainable systems, and its methodical strategy fit LumosLore's continuing development requirements.

3.5 Project Management Methodology

Methodologies for project management are organized frameworks that direct the planning, carrying out, and overseeing of projects (Cooper, 1990). Different approaches meet different project requirements. While agile systems like Scrum and Kanban prioritize flexibility and iterative development, traditional approaches like Waterfall follow a linear series of stages. For complex projects, Prince2 offers an organized, process-driven framework (Types of Project Management: Methodologies, Industries, and More, 2023). The goals of lean project management are waste reduction and value maximization.Project managers can select the most appropriate methodology based on the objectives, size, and complexity of the project by selecting from a variety of techniques offered by each methodology, such as the Critical Path Method (CPM) and Critical Chain Project Management (CCPM).

Agile Scrum was chosen as the project management style for the LumusLore project because it fits with the project's iterative and dynamic development process. Flexibility, regular cooperation, and adaptability to changing needs are given priority in this decision. The Agile Scrum framework is improved by using Jira as the project management platform since it offers strong capabilities for sprint planning, backlog management, and real-time collaboration. Jira's intuitive user interface and extensive tracking features guarantee effective task management and smooth team communication. The LumusLore team is able to produce software that is extremely responsive, customer-focused, and quality-driven because of the integration of Jira and Agile Scrum.

3.6 Team Work Breakdown Structure (WBS)

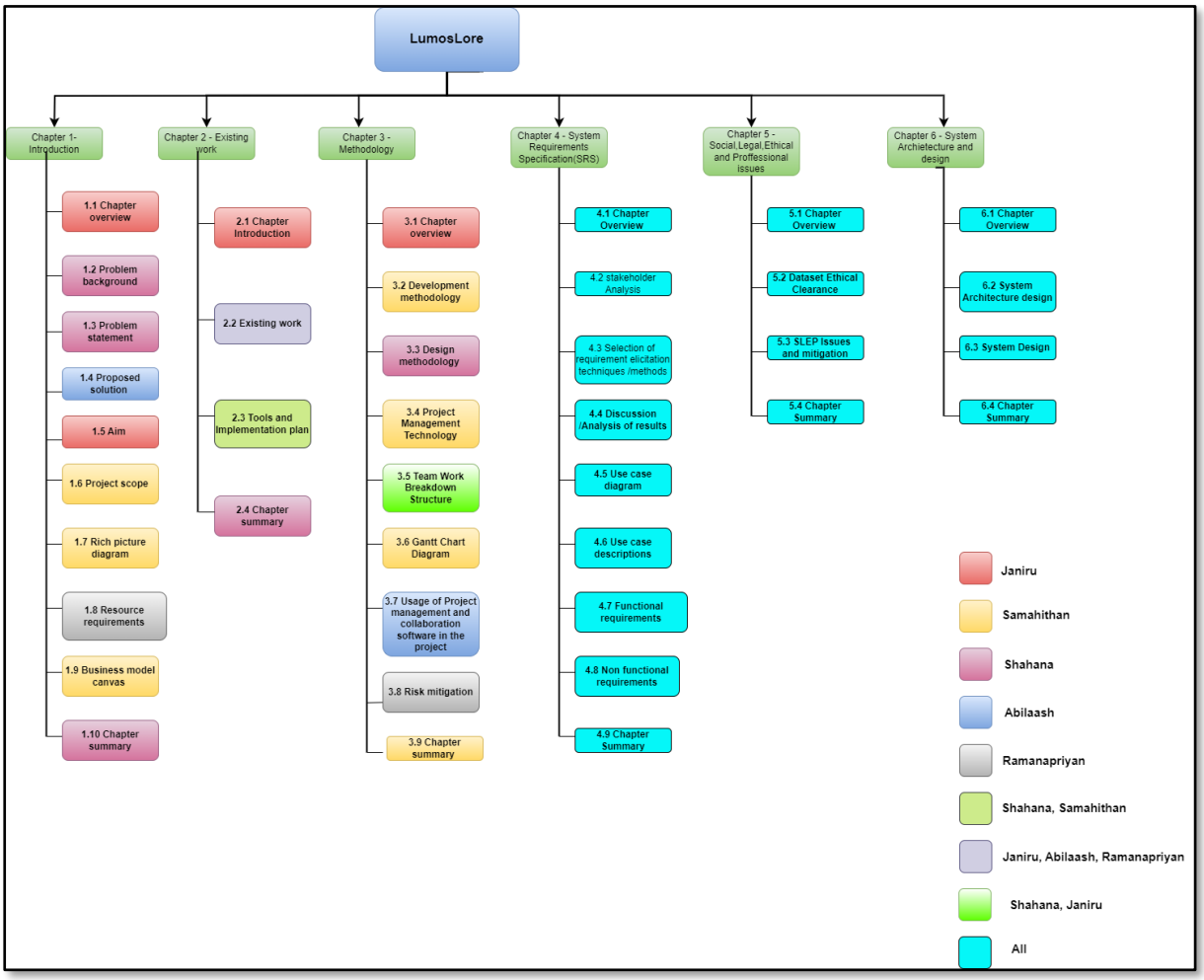


Figure 9 Work Breakdown Structure

3.7 Gantt Chart Diagram

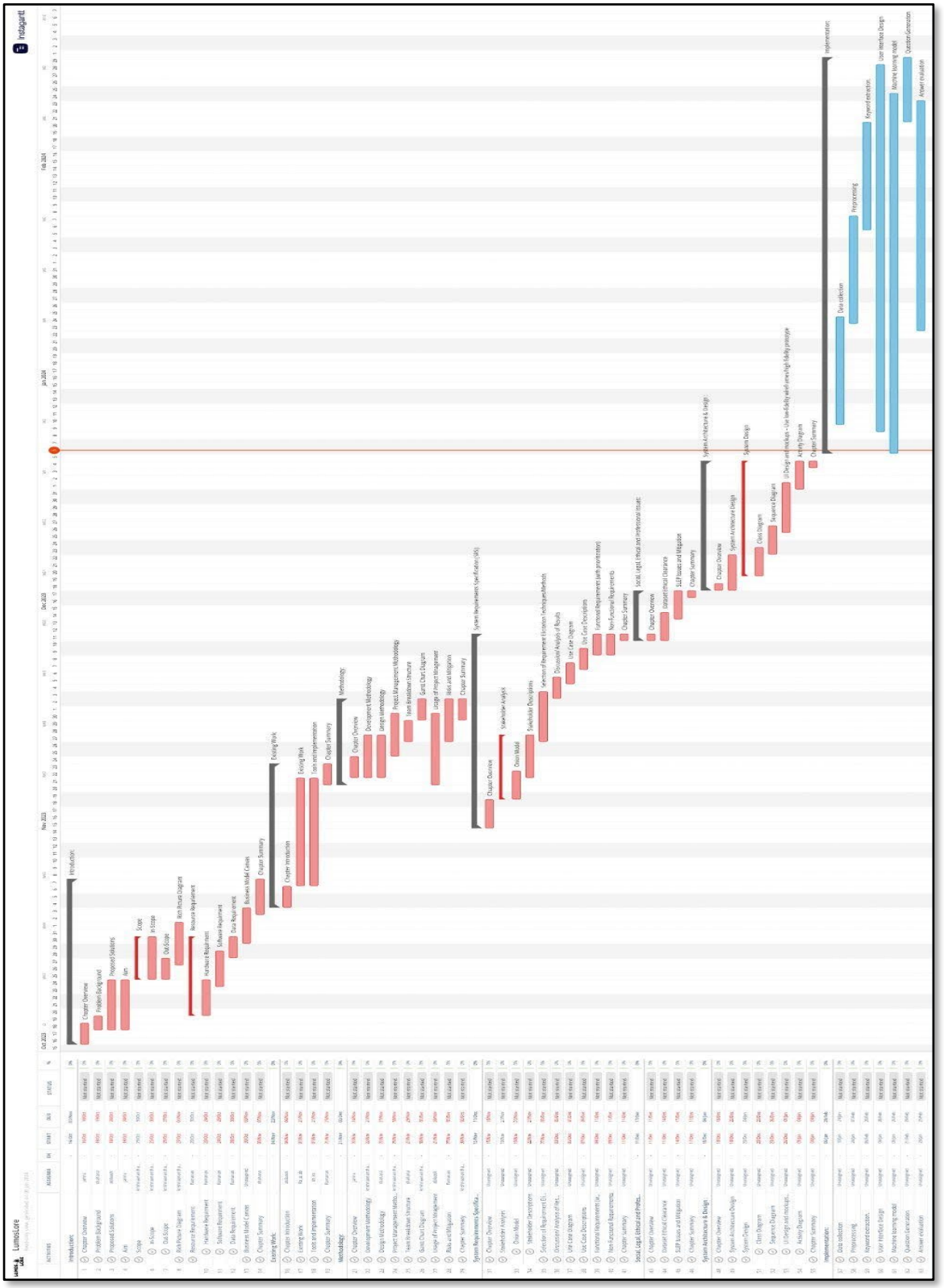


Figure 10 Gantt Chart

3.8 Usage of Project Management and Collaboration Software In the project

3.8.1 Jira

Jira Software is a widely used project management solution that helps teams in effectively arranging, guiding, and tracking their projects. Although teams that develop software utilise it regularly, it may be applied to a wide range of initiatives and companies(Atlassian,2023). With configurable techniques, Jira software offers teams to develop tasks, Prioritise work, analyse progress and collaborate easily on a single platform(Atlassian, n.d). Several organisations seeking ways to improve their process for managing projects utilise it due to its flexibility, wide range of personalization options, and interfaces with other applications.

Refer to <Appendix A>

3.8.2 Google Meetings

Google Meet is a collaboration and video conferencing tool that includes online meetings, communications, and visual interaction (support.google.com, n.d.). It serves an easy-to-use platform for individuals and organisations to meet via audio or video chats, allowing faraway collaborators and communication. Google Meet integrates smoothly with other Google Workspace applications, making it achievable to exchange documents using Google Drive and arrange meetings using Google Drive and arrange meetings using Google Calendar, Screen Sharing, live captioning, and conducting big meetings with numerous attendees are some of its benefits(Devoteam G Cloud, n.d.). Google Meet is a useful application for personal as well as professional virtual meetings as a result of its intuitive UI and compatibility with various devices.

Refer to <Appendix B>

3.9 Risks and Mitigation

Table 8 Risk and Mitigation

Risk Item	Severity	Frequency	Mitigation Plan
Insufficient understanding of the technical aspects.	**5**	**3**	**Get help from the experts and do self-study.**
Absence of prior research, datasets and additional sources on the relevant topic.	**4**	**5**	**Check in advance before committing to the project that required datasets and other sources are available.**
Malfunctions in the software or hardware components.	**4**	**3**	**Keep track of the data backups and use cloud storage, having extra devices in case of failure.**
Delays brought on by unforeseen events, including illnesses.	**3**	**1**	**Don't wait until the deadline. Do the most part when you are free without hesitation.**
Project requirements have grown significantly, leading to an excessive expansion in scope.	**4**	**2**	**Carefully assessing new requirements, setting change control, and keeping updated with stakeholders to properly manage scope expansions**
Overdue work because of the project's complexity.	**4**	**2**	**Don't wait until the deadline. Do the most part when you are free without hesitation.**
Problems arise throughout the execution.	**3**	**2**	**Get assistance from experts, having an alternate software.**

3.10 Chapter Summary

The LumusLore project thoroughly describes its process in this chapter, focusing on three essential elements: development, design, and project management. The team chose Agile as their development technique since it promotes flexibility and iterative development. In addition to this decision, the design methodology of Object-Oriented Analysis and Design (OOAD) was adopted, guaranteeing a methodical and modular approach to system architecture. Agile Scrum is chosen to facilitate responsiveness and participation in project management. The chapter explores the reasoning behind each decision, tying them in with LumusLore's goals for effective, client-focused, and high-quality development.

Turning our attention to project management methods and technologies, LumusLore uses Jira to improve the application of Agile Scrum. Jira's extensive feature set makes backlog management and sprint planning easier. The team's communication strategy incorporates Google Meets and WhatsApp to guarantee real-time collaboration. Tasks are broken down into a Work Breakdown Structure (WBS) for a thorough perspective, and dependencies and timelines are shown visually in a Gantt chart. Together with methodologies, these tools establish a transparent framework that facilitates productive teamwork and efficient project management for LumusLore.

References

Albers, A. et al. (2019). A systematic approach to situation-adequate mechatronic system development by ASD - Agile Systems Design. Procedia CIRP, 84, 1015–1022. Available at https://doi.org/10.1016/j.procir.2019.03.312. [Accessed 5 January 2024].

ASGEN: ANSWER-CONTAINING SENTENCE GENERATION TO PRE-TRAIN QUESTION GENERATOR FOR DATA AUGMENTATION IN QUESTION ANSWERING. (n.d.). Available at: https://openreview.net/pdf?id=H1lFsREYPS [Accessed 4 Jan. 2024].

Atlassian (2023). Welcome to Jira Software. [online] Atlassian. Available at: https://www.atlassian.com/software/jira/guides/getting-started/introduction#what-is-jira-software. [Accessed 5 January 2024].
Atlassian (n.d.). Jira Software for Teams. [online] Atlassian. Available at: https://www.atlassian.com/software/jira/guides/getting-started/who-uses-jira#for-project-management-teams. [Accessed 5 January 2024].

Baschin, J., Inkermann, D. and Vietor, T. (2019). Agile Process Engineering to support Collaborative Design. Procedia CIRP, 84, 1035–1040. Available at https://doi.org/10.1016/j.procir.2019.05.010. [Accessed 5 January 2024].

Bradbury, N. (2016). Attention span during lectures: 8 seconds, 10 minutes, or more? Advances in Physiology Education, [online] 40(4), pp.509–513. Available at: https://doi.org/10.1152/advan.00109.2016 [Accessed 2 Jan. 2024].

Cooper, R.G. (1990). Stage-gate systems: A new tool for managing new products. Business Horizons, 33 (3), 44–54. Available at https://doi.org/10.1016/0007-6813(90)90040-I. [Accessed 5 January 2024].

DeepLearning.AI (2023). Natural Language Processing (NLP) - A Complete Guide. [online] www.deeplearning.a Available at:
Devoteam G Cloud. (n.d.). The 10 most-used features in Google Meet. [online] Available at: https://gcloud.devoteam.com/blog/the-10-most-used-features-in-google-meet/ [Accessed 5 January 2024].

GeeksforGeeks. (2021). Natural Language Processing - Overview. [online] Available at: https://www.geeksforgeeks.org/natural-language-processing-overview/. [Accessed 5 January 2024].

Gong, H., Pan, L. and Hu, H. (2022). KHANQ: A Dataset for Generating Deep Questions in Education. [online] ACLWeb. Available at: https://aclanthology.org/2022.coling-1.518/ [Accessed 2 Jan. 2024].

HeidiSteen (2023). RAG and generative AI - Azure Cognitive Search. [online] learn.microsoft.com. Available at: https://learn.microsoft.com/en-us/azure/search/retrieval-augmented-generation-overview. [Accessed 6 Jan. 2024].

Höhn, F. (2022). Natural Language Question Answering Systems. [online] SEEBURGER Blog. Available at: https://blog.seeburger.com/natural-language-question-answering-systems-get-quick-answers-to-concrete-questions/. [Accessed 7 Jan. 2024].

https://support.google.com/meet/answer/12387657?hl=en.

https://www.deeplearning.ai/resources/natural-language-processing/.

huggingface.co. (n.d.). Preprocess. [online] Available at: https://huggingface.co/docs/transformers/preprocessing [Accessed 2 Jan. 2024].

Li, Y. and Ning, H. (2022). Keywords extraction method based on two-way feature fusion. 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). August 2022. 127–130. Available at https://doi.org/10.1109/MLISE57402.2022.00033 [Accessed 6 January 2024].

Lokare, G. (2023). Question Answering with Hugging Face Transformers: A Beginner's Guide. [online] Medium. Available at: https://medium.com/@lokaregns/question-answering-with-hugging-face-transformers-a-beginners-guide-487ae1a91b9a [Accessed 7 Jan. 2024].

Microsoft Research. (n.d.). NewsQA Dataset. [online] Available at: https://www.microsoft.com/en-us/research/project/newsqa-dataset [Accessed 4 Jan. 2024].

Mushashu, E.T. and Mtebe, J.S. (2019). Investigating Software Development Methodologies and Practices in Software Industry in Tanzania. 2019 IST-Africa Week Conference (IST-Africa). May 2019. Nairobi, Kenya: IEEE, 1–11. Available at https://doi.org/10.23919/ISTAFRICA.2019.8764884. [Accessed 5 January 2024].

NLTK (2009). Natural Language Toolkit — NLTK 3.4.4 documentation. [online] Nltk.org. Available at: https://www.nltk.org/ [Accessed 2 Jan. 2024].

Otten, N.V. (2023). Top 5 Ways To Implement Question-Answering Systems In NLP & A List Of Python Libraries. [online] Spot Intelligence. Available at: https://spotintelligence.com/2023/01/20/question-answering-qa-system-nlp/#:~:text=Question%20answering%20(QA)%20is%20a [Accessed 7 Jan. 2024].

paperswithcode.com. (n.d.). Papers with Code - Machine Learning Datasets. [online] Available at: https://paperswithcode.com/datasets?task=question-generation [Accessed 4 Jan. 2024].

paperswithcode.com. (n.d.). Papers with Code - Machine Learning Datasets. [online] Available at: https://paperswithcode.com/datasets?task=question-answering&page=1 [Accessed 4 Jan. 2024].

paperswithcode.com. (n.d.). Papers with Code - Machine Learning Datasets. [online] Available at: https://paperswithcode.com/datasets?task=text-generation [Accessed 4 Jan. 2024].

paperswithcode.com. (n.d.). Papers with Code - Machine Learning Datasets. [online] Available at: https://paperswithcode.com/datasets?task=text-classification&page=1 [Accessed 4 Jan. 2024].

paperswithcode.com. (n.d.). Papers with Code - Machine Learning Datasets. [online] Available at: https://paperswithcode.com/datasets?task=token-classification [Accessed 4 Jan. 2024].

Scepanovic, S., Devedzic, V., & Kraljevski, I. (2011). E-LEARNING BENCHMARKING: METHODOLOGY AND TOOLS REVIEW. The Second International Conference on e-Learning (eLearning-2011), Belgrade, Serbia. Available at:https://www.researchgate.net/publication/281269775_E-LEARNING_BENCHMARKING_METHODOLOGY_AND_TOOLS_REVIEW [Accessed 2 Jan 2024]

Schuh, G. et al. (2018). Identification of Agile Mechanisms of Action As Basis for Agile Product Development. Procedia CIRP, 70, 19–24. Available at https://doi.org/10.1016/j.procir.2018.02.007. [Accessed 5 January 2024].

Scikit-learn.org. (2019). 5.3. Preprocessing data — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/preprocessing.html [Accessed 2 Jan. 2024].

Sharma, K. (2015). Top 12 Software Development Methodologies. TatvaSoft Blog. Available at https://www.tatvasoft.com/blog/top-12-software-development-methodologies-and-its-advantages-disadvantages/ [Accessed 5 January 2024].

support.google.com. (n.d.). Use Google Meet features in the new Meet app - Google Meet Help. [online] Available at:

Takyar, A. (2023). What is Retrieval Augmented Generation (RAG)? [online] LeewayHertz - AI Development Company. Available at: https://www.leewayhertz.com/what-is-retrieval-augmented-generation/ [Accessed 6 Jan. 2024].

Types of Project Management: Methodologies, Industries, and More. (2023). Coursera. Available at https://www.coursera.org/articles/types-of-project-management [Accessed 6 January 2024].

www.kaggle.com. (n.d.). Question-Answer Dataset. [online] Available at: https://www.kaggle.com/datasets/rtatman/questionanswer-dataset [Accessed 7 Jan. 2024].

Xia, Y., Hu, Y., Wu, C., Yang, L. and Lei, M. (2022). Challenges of online learning amid the COVID-19: College students' perspective. Frontiers in Psychology, 13. Available at: https://doi.org/10.3389/fpsyg.2022.1037311 [Accessed 2 Jan. 2024].

Appendix

Appendix A

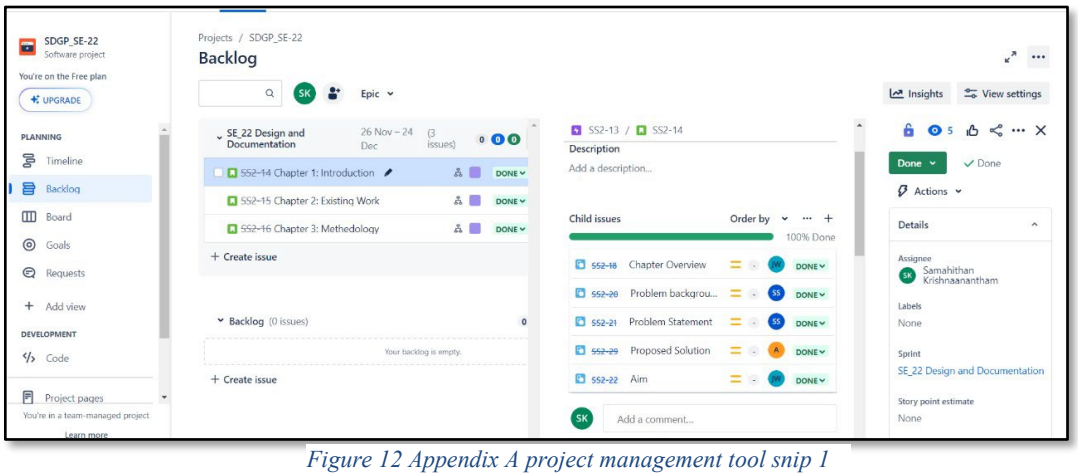


Figure 12 Appendix A project management tool snip 1

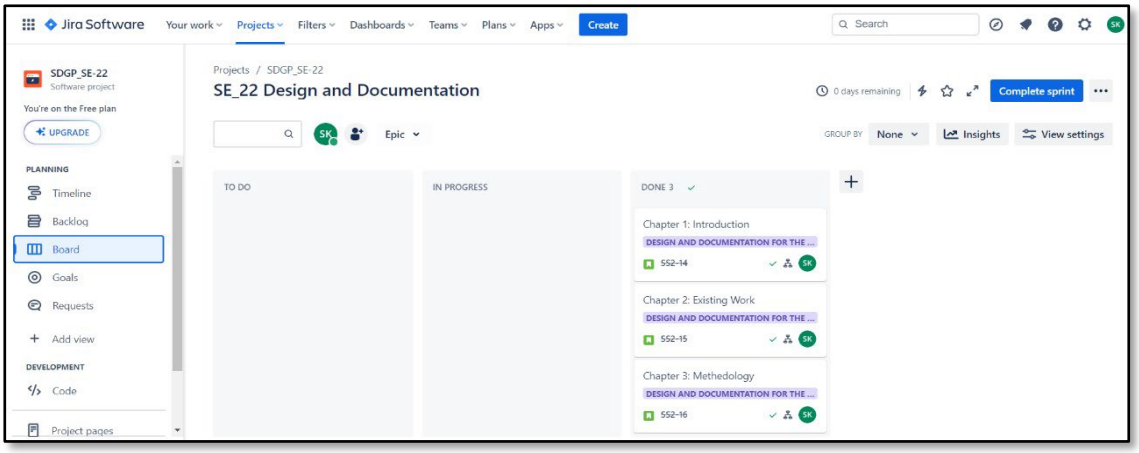


Figure 11 A project management tool snip 2

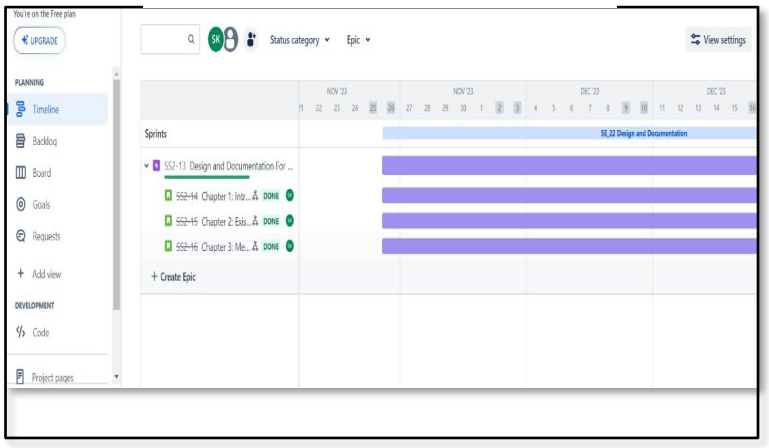


Figure 13 A project management tool snip 3

Appendix B

Table 9:Appendix Table

Date and Time Duration	Discussed about	Screenshot Evidence
24/09/2023 45 Minutes	Team Members Introductions & Team Leader Selection	Appendix B.1
10/10/2023 1 Hour 30 Minutes	Sharing new project Ideas	Appendix B.2
26/10/2023 1 Hour	Finalise the topic and discuss the scope of the topic	Appendix B.3
14/11/2023 45 Minutes	Dividing Project Proposal topics	Appendix B.4
21/11/2023 1 Hour 45 Minutes	Dividing Project report topics	Appendix B.5
26/11/2023 2 Hours	Discuss the research papers and list down all key points	Appendix B.6
03/01/2024 1 Hour 45 Minutes	Review the progress and giving ideas to improve the works	Appendix B.7
07/01/2024 30 Minutes	Finalise the Report	Appendix B.8

Appendix B.1

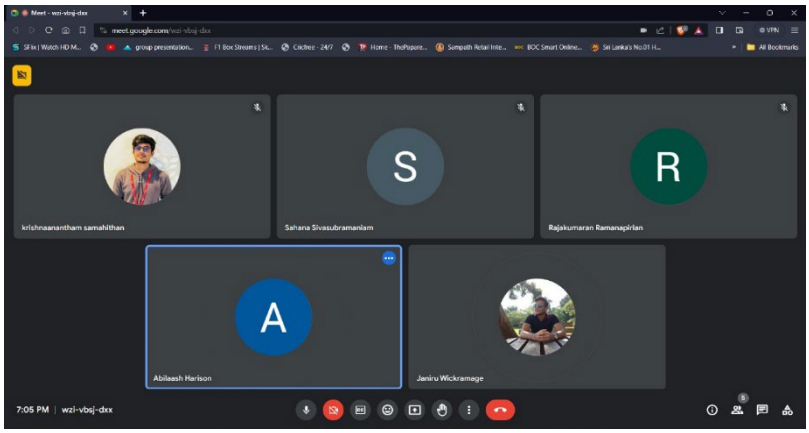


Figure 14 Appendix B.1 meeting evidence

Appendix B.2

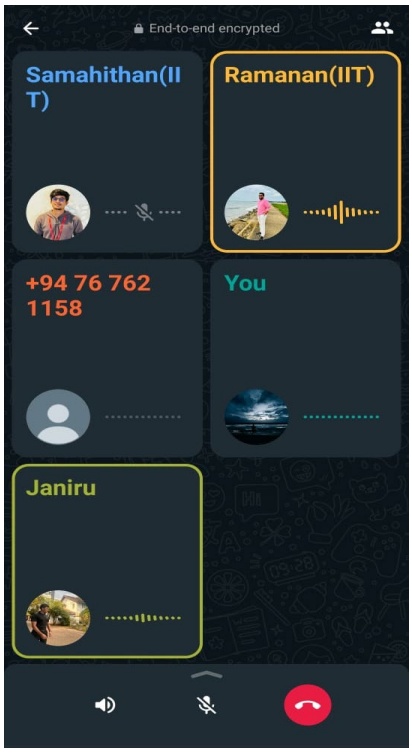


Figure 16:Appendix B.2 meeting evidence

Appendix B.3

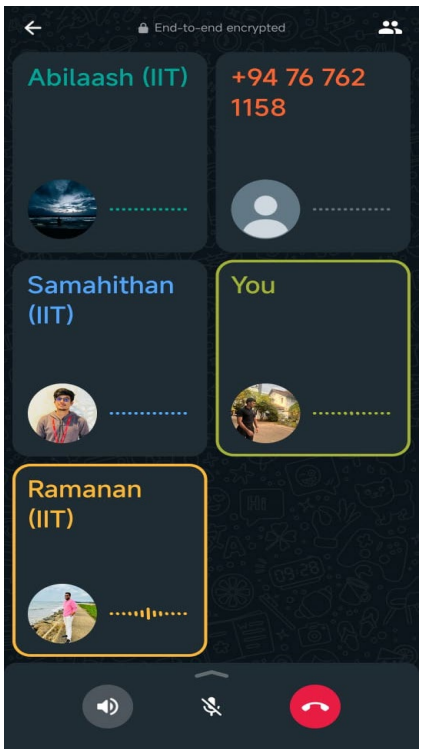


Figure 15:Appendix B.3 meeting evidence

Appendix B.4

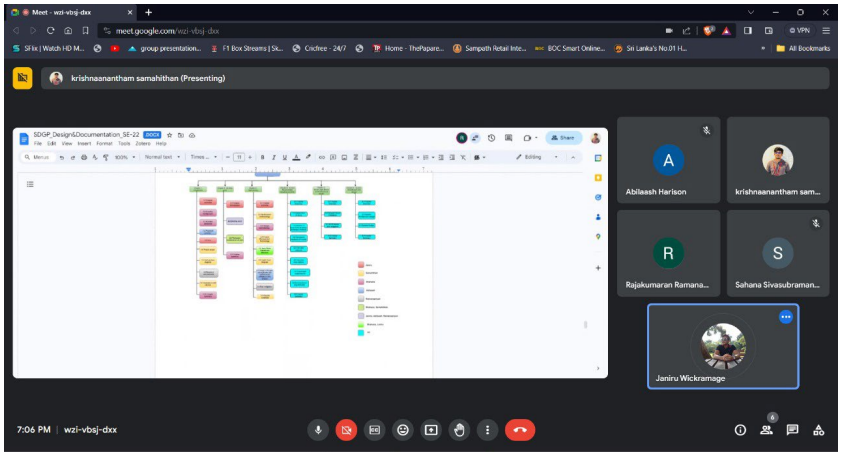


Figure 17:Appendix B.4 meeting evidence

Appendix B.5

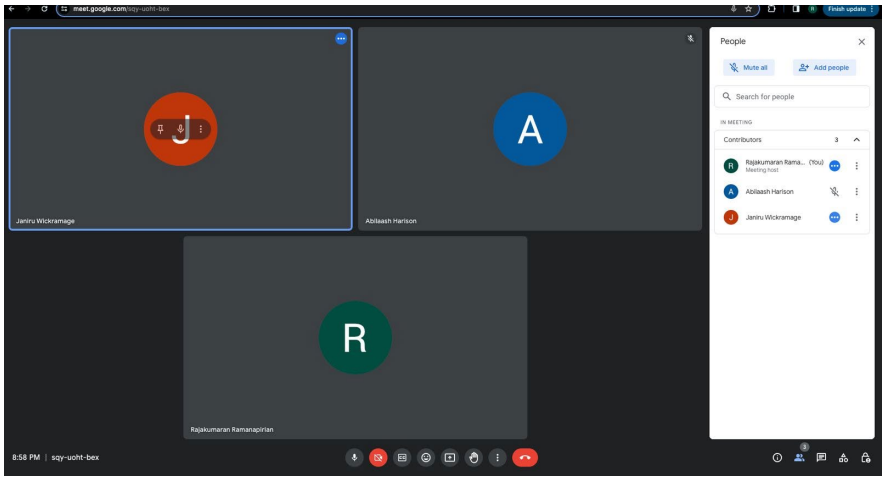


Figure 18:Appendix B.5 meeting evidence

Appendix B.6

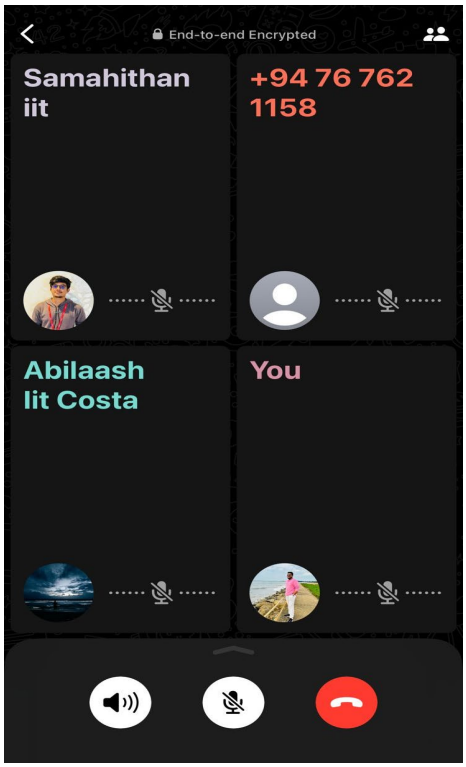


Figure 19:Appendix B.6 meeting evidence

Appendix B.7

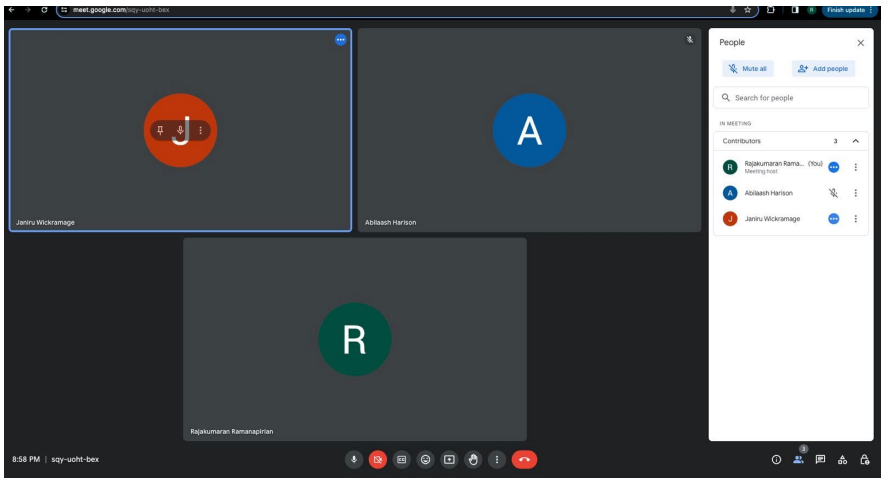


Figure 20:Appendix B.7 meeting evidence

Appendix B.8

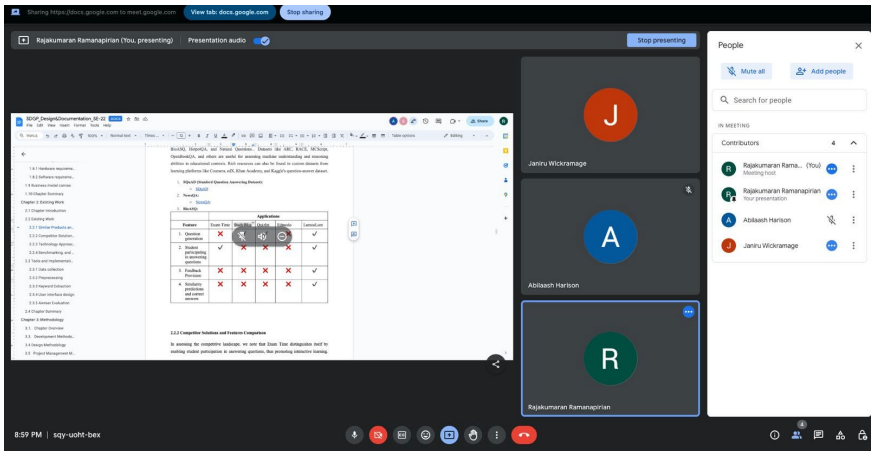


Figure 21:Appendix B.8 meeting evidence