GRADIVA

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

Enhancing Personalized Learning based on AI-Driven Lesson Plan Generator Using Mistral-7B for Efficient Content Extraction and Summarization

Manjunath D.R. [1], Selva Kumar S. [2], Sai Shiva Sumanth Reddy [3] & Lohith J.J. [4]

[1,2] Department of Computer Science & Engineering, B.M.S. College of Engineering, Bengaluru, India.
[3] Department of Computer Science & Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, India.
[4] Department of AI & ML, Nagarjuna College of Engineering and Technology.
Email: [1]manjunathdr.cse@bmsce.ac.in, [2]selva.cse@bmsce.ac.in, [3]sumsnth-cse@dsatm.edu.in, [4]lohithjj@gmail.com

Abstract

In the rapidly evolving landscape of digital education, students often face challenges such as information overload and a lack of personalized learning experiences. Traditional teaching methods and static educational resources may not effectively cater to the diverse needs of learners. To address these issues, we present the Lesson Plan Generator (LPG) using Mistral-7B, a sophisticated AI-driven educational platform. This system integrates advanced web scraping techniques, natural language processing, and content summarization to create customized lesson plans tailored to individual user preferences. The LPG operates in two primary phases. In the first phase, users input their desired study hours, topics of interest, and preferred difficulty levels. The system utilizes these inputs to dynamically generate personalized lesson plans using the Mistral-7B language model. These plans include structured courses, theoretical concepts, practical examples, and exercises designed to match the user's specified parameters. The generated lesson plans are accessible through a user-friendly interface, ensuring ease of use and accessibility. In the second phase, the LPG enhances the learning experience by providing detailed content summaries for specific topics. Utilizing web scraping tools such as Scrapy and BeautifulSoup, the system extracts relevant educational content from various online sources. The extracted content is then processed and summarized by Mistral-7B, providing users with concise and coherent summaries. Additionally, the system integrates supplementary resources such as topic URLs and YouTube links, offering multimedia experiences to deepen understanding and engagement. The LPG's innovative approach aims to streamline the educational process by reducing cognitive load and enhancing learning outcomes through personalized content delivery. This paper details the system architecture, methodology, and implementation of the LPG, along with performance evaluations and user feedback. The results demonstrate the efficacy of the LPG in generating accurate and relevant lesson plans and summaries, significantly improving the digital learning experience. Future enhancements include extending support for multiple websites, summarizing multimedia content, and incorporating adaptive learning algorithms to further personalize the educational journey.

1. INTRODUCTION

The advent of digital technology has profoundly transformed various sectors, including education, by introducing new opportunities and challenges for both learners and educators. Traditional educational methodologies [1], often rooted in static content and uniform teaching approaches, have begun to be supplemented or replaced by digital solutions that offer dynamic and personalized learning experiences [2]. However, despite these advancements, students

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

continue to face significant challenges, such as information overload and a lack of personalized learning experiences [3]. The abundance of available online resources, although beneficial, can overwhelm learners, making it difficult for them to identify and focus on the most relevant and valuable content. Furthermore, traditional teaching methods, characterized by a one-size-fits-all approach, often fail to accommodate the diverse learning styles and paces of individual students. These challenges underscore the pressing need for innovative educational tools that can dynamically tailor content to meet the specific needs and preferences of each learner.

Personalized learning has emerged as a significant development in addressing these challenges. It involves tailoring education to individual students' needs, interests, and abilities, which has been shown to enhance learning outcomes and engagement. Digital technology [4] plays a critical role in this personalization by providing tools and platforms that support adaptive learning environments. Research indicates that digital tools, when integrated effectively, can significantly enhance personalized learning environments by offering customized content and learning paths based on individual learner profiles.

Despite these advancements, there remains a gap in effectively utilizing digital resources to their fullest potential in educational settings [5-7]. Studies have shown that while there is an increased use of digital technology in personalized learning settings, the integration and practical application of these tools are often inconsistent and underutilized [8-9]. This inconsistency highlights the need for more robust systems that can seamlessly integrate various digital tools to provide a cohesive personalized learning experience.

In response to these challenges, we present the Lesson Plan Generator (LPG) using Mistral-7B, an advanced AI-driven educational platform designed to address the complexities of modern education. This system integrates sophisticated web scraping techniques, natural language processing (NLP), and content summarization to generate customized lesson plans that cater to individual learning preferences and needs. The LPG aims to provide a comprehensive solution that not only curates and organizes educational content but also personalizes the learning journey for each student, thereby enhancing engagement and facilitating deeper understanding.

The LPG system operates through a robust web scraping mechanism that extracts relevant and current educational content from a variety of online sources. This mechanism ensures the accuracy and reliability of the data through rigorous error handling and verification processes. Once the content is extracted, it is processed using the Mistral-7B language model to generate personalized lesson plans based on user-specified parameters such as study hours, topics of interest, and difficulty levels. These personalized lesson plans include structured courses, theoretical concepts, practical examples, and exercises designed to match the user's specified parameters. The generated lesson plans are made accessible through a user-friendly interface, ensuring ease of use and accessibility.

In addition to generating lesson plans, the LPG enhances the learning experience by providing detailed content summaries for specific topics. Utilizing web scraping tools such as Scrapy and BeautifulSoup, the system retrieves and extracts relevant educational content from various online sources. The extracted content is then processed and summarized by the Mistral-7B model, providing users with concise and coherent summaries. These summaries are supplemented with original content links and multimedia resources, such as YouTube videos, to offer a richer, more engaging learning experience.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

To ensure the platform's usability and accessibility, the LPG features intuitive interfaces that allow users to input their preferences, access generated lesson plans, and review summarized content with ease. Robust security measures are implemented to protect user data and comply with data privacy regulations, ensuring the confidentiality and integrity of user information. Moreover, mechanisms for collecting user feedback and analytics data are established to inform iterative improvements, ensuring the continuous enhancement of the system.

Traditional educational platforms or Learning Management Systems (LMS) typically offer pre-defined courses and static content curated by educators or administrators. While these systems provide basic functionalities such as grading tools, communication channels, and administrative functions, they often lack the ability to provide personalized learning experiences [10-11]. The static nature of these platforms, combined with limited integration of multimedia resources, poses significant challenges in engaging students and catering to diverse learning styles. Consequently, these systems fall short in dynamically adapting to individual student needs and preferences.

The proposed LPG system aims to overcome these limitations by incorporating advanced technologies to deliver personalized educational content. By leveraging AI-driven content generation and summarization techniques, combined with efficient web scraping mechanisms, the LPG provides a comprehensive, personalized, and engaging learning experience. The system's ability to dynamically tailor content to individual learning preferences and integrate supplementary multimedia resources ensures a more effective and enjoyable learning journey for students. This project seeks to revolutionize digital education by offering tailored, efficient, and accessible learning pathways, ultimately improving student engagement and educational outcomes.

2. LITERATURE REVIEW

Web scraping has emerged as a crucial technique for data extraction from various web sources, facilitating diverse applications across multiple sectors. R. S. Tomar's study [12] provides an introduction to web scraping technologies, detailing available software tools and methodologies to prevent web scraping, illustrated with practical examples. This foundational understanding sets the stage for exploring more advanced and specialized applications of web scraping.

I. Latrous et al. [13] present a comprehensive review of web scraping techniques and their applications. The paper outlines the basic design of web scrapers, diverse scraping methods, and the technologies employed. The authors propose a development procedure for web scraping tools, culminating in a detailed literature review that highlights the breadth and depth of web scraping applications in various fields.

The integration of APIs in web scraping is explored by K. T. Nyunt [14], who examines the use of HTTP requests to retrieve data in XML or JSON formats from web applications, specifically targeting video content from YouTube. This study underscores the relevance of merging web scraping with publicly available APIs to enhance the extraction and analysis of multimedia data.

In addressing the limitations of traditional search engines, Ma et al. [15] propose a vertical search engine based on Nutch and Solr to improve precision and user retrieval. Their experimental evaluations demonstrate the effectiveness of this approach in overcoming

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

excessive search results and low precision rates, emphasizing the importance of specialized search engines in web scraping.

Z. Wang and J. Lv [16] delve into the efficient acquisition and analysis of massive datasets using topic web crawlers. Their study highlights strategies to enhance traditional web crawlers by analyzing links for relevant content, thus increasing efficiency and saving time in data collection processes.

E. Uzun introduces the UzunExt approach [17], prioritizing the time-efficient extraction of web page content by utilizing additional information to enhance speed. This structured method provides a comprehensive overview, from literature review to practical recommendations, for future web scraping studies.

The application of web scraping in real-time learning analytics is investigated by K. Dobashi [18], who focuses on Moodle course material. The study advocates for automating Moodle course log downloads to visualize learners' interactions in real-time, thereby simplifying operational burdens on teachers and unlocking new possibilities in learning analytics.

A comparative analysis of Python libraries used for web data extraction is presented by Uzun, Erdinç, and colleagues [19]. The study concludes that the Lxml library offers superior time results for DOM-based libraries, whereas BeautifulSoup, despite its ease of use, performs less efficiently due to additional processes involved in DOM creation.

B. Bhardwaj et al. [20] explore the combination of web scraping with named entity recognition (NER) for enhanced data extraction flexibility and diversity. The authors argue that NER approaches yield better results for discrete and highly unstructured data, highlighting the benefits of integrating NER with web scraping.

S. Vasilica Oprea and A. Bára [21] discuss the combined use of Selenium and BeautifulSoup for web scraping, particularly in handling dynamic and JavaScript-driven websites. Selenium's capabilities in interacting with web pages complement BeautifulSoup's efficiency in parsing HTML, offering a robust solution for complex scraping tasks.

Automatic text summarization, a critical component of the LPG system, is thoroughly reviewed by W. S. El-Kassas et al. [22]. The study evaluates different metrics such as ROUGE, BLEU, and METEOR used for assessing the performance of NLP-based summarization systems, providing a comprehensive overview of the current state of automatic summarization.

S. Tsuchiya and colleagues [23] introduce a novel information arrangement technique combining the Degree of Association Algorithm and the Common-sense Judgment System. Their approach efficiently organizes news articles, showcasing precision in field classification, keyword extraction, and summarization.

The survey by Awasthi et al. [24] on NLP-based text summarization techniques covers various approaches including abstractive, extractive, and hybrid methods. The study addresses the challenges of subjective evaluation, limited labeled data, and linguistic issues, emphasizing the potential and limitations of current summarization technologies.

Finally, S. Jugran and co-authors [25] utilize SpaCy's pre-trained language model for extractive summarization, performing sentence parsing, part-of-speech tagging, and named entity recognition. Their methodology scores sentences based on length, word frequency, and

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

named entities, demonstrating an effective approach to automatic text summarization using NLP tools.

These studies collectively provide a comprehensive understanding of the methodologies, applications, and advancements in web scraping and text summarization, forming the foundation for the development of the Lesson Plan Generator using Mistral-7B.

3. SYSTEM DESIGN

The Lesson Plan Generator (LPG) system is meticulously designed to offer a personalized educational experience by integrating advanced web scraping techniques, natural language processing (NLP), and content summarization. The proposed system architecture is shown in figure 1. The system comprises several interconnected components, each responsible for specific tasks to ensure the efficient generation of customized lesson plans and concise content summaries. The flow diagram is shown in the figure 2.

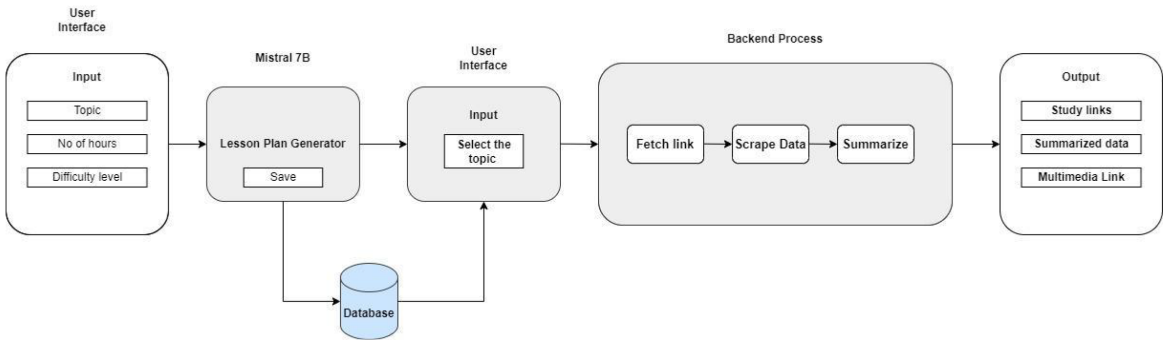

Figure 1: System Architecture of the Proposed System

3.1. Web Scraping Mechanism

The initial step in the LPG system involves extracting educational content from various online sources using a robust web scraping mechanism. This mechanism employs:

3.1.1. Scrapy

Scrapy is an open-source web crawling framework for Python that is instrumental in extracting data from websites. It provides tools to navigate web pages, parse HTML, and handle requests efficiently. Recent studies have highlighted Scrapy's effectiveness in handling large-scale web scraping projects due to its extensibility and support for asynchronous requests, making it suitable for diverse educational applications. Scrapy's capabilities in handling complex scraping tasks efficiently are well-documented in various academic sources, making it an ideal choice for educational content extraction [26-27].

3.1.2. BeautifulSoup

BeautifulSoup is a Python library for parsing HTML and XML documents, working alongside Scrapy to parse the retrieved web content and extract relevant educational materials. Its ability to handle complex HTML structures makes it invaluable for cleaning and pre-processing data. BeautifulSoup is widely used for its simplicity and effectiveness in navigating and modifying parse trees, making it an essential tool in the LPG system [28-29].

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

3.1.3. DuckDuckGo API

The DuckDuckGo API is utilized to perform topic-specific searches, aiding in retrieving web links related to the user's specified topics. This ensures the relevance and accuracy of the extracted content. DuckDuckGo's focus on privacy and unbiased search results enhances the quality of the data retrieved, making it a reliable source for educational materials [30].

The extracted content undergoes a cleaning and pre-processing phase to remove any irrelevant information, ensuring that only high-quality educational materials proceed to the next stage of processing.

3.2. Personalized Course Plan Generation

Following content extraction, the LPG system generates personalized lesson plans based on user inputs through the following stages:

3.2.1. User Input Handling

The initial stage of generating personalized course plans involves gathering user inputs. Users provide their study preferences, including the number of available study hours, topics of interest, and preferred difficulty levels. This stage is crucial as it sets the foundation for the customization of the lesson plans. The system employs a user-friendly interface where learners can easily input their preferences. Once the inputs are provided, the system validates these inputs for consistency and reasonableness. This validation process ensures that the data entered by the user is logical and feasible. For instance, if a user inputs an unrealistic number of study hours or conflicting topics of interest, the system prompts the user to adjust their inputs.

3.2.2. NLP Processing with Mistral-7B

Mistral-7B employs sophisticated NLP techniques to understand and analyze the user's inputs. The model processes natural language inputs to extract meaningful information about the user's learning goals and preferences. It uses techniques such as tokenization, parsing, and semantic analysis to interpret the user data accurately [30-31]. Based on the analyzed inputs, Mistral-7B generates a tailored course plan. This plan includes recommended courses, theoretical concepts, practical examples, and exercises that match the user's specified parameters. The model leverages its extensive training on diverse educational datasets to ensure that the lesson plans are comprehensive and relevant. It incorporates user-specific factors such as preferred learning styles and pace, ensuring a highly personalized learning experience. He use of Mistral-7B ensures that the generated lesson plans are not only personalized but also coherent and contextually relevant. The model's ability to generate human-like text allows for the creation of lesson plans that are easy to follow and understand. This enhances the overall learning experience, making it more engaging and effective

3.2.3. Course Plan Storage

The generated lesson plans are securely stored in a structured database (RestDB), ensuring efficient data storage and retrieval. RestDB is chosen for its robust database management capabilities, ensuring data integrity and security. Users can access their personalized course plans through an intuitive user interface, providing a seamless and efficient user experience [8].

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

3.3. Content Summarization

To enhance the learning experience, the Lesson Plan Generator (LPG) system provides detailed summaries of specific topics through a meticulously designed process that involves multiple stages: topic specification and search, content extraction and cleaning, summarization using the Mistral-7B model, and presentation of the summarized content. The flow chart for the content summarization is shown in the figure 3. Each stage is critical in ensuring that the final output is accurate, relevant, and useful for the users.

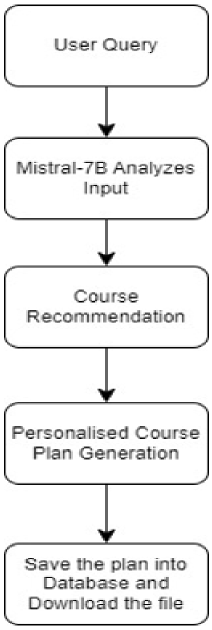

Figure 2: The Flow Diagram of Lesson Plan Generator

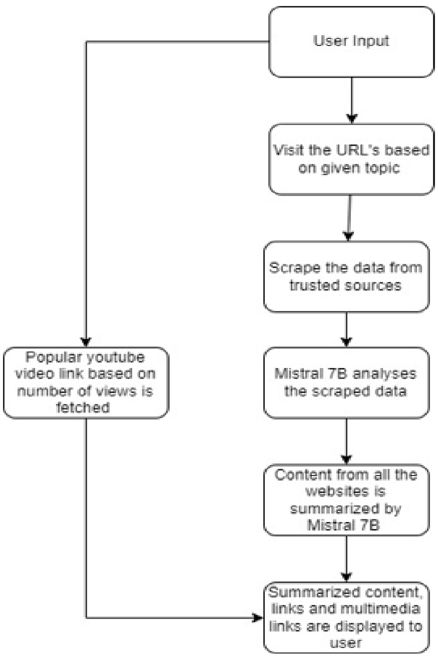

Figure 3: The LPG's Summarization Process

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

3.3.1. Topic Specification and Search

The first step in the content summarization process is the specification of topics by the users. Users input the topics they are interested in summarizing. The system then leverages the DuckDuckGo API to perform topic-specific searches and retrieve relevant web links. DuckDuckGo is chosen for its privacy-focused approach and unbiased search results, which ensures high-quality data retrieval. This search engine API efficiently provides diverse sources, ensuring a broad spectrum of information for comprehensive summaries.

The topic specification stage is designed to be user-friendly, allowing users to input keywords or phrases related to their areas of interest. The system processes these inputs and formulates search queries that are optimized to fetch the most relevant results. This stage is crucial as it sets the foundation for the quality and relevance of the content that will be summarized.

3.3.2. Content Extraction and Cleaning

Once the relevant web links are retrieved, the system uses web scraping tools such as Scrapy and BeautifulSoup to extract the content from these links. Scrapy is utilized for its robust crawling capabilities, allowing the system to navigate through various web pages efficiently. It handles requests, parses HTML, and retrieves data, making it an ideal tool for large-scale data extraction tasks. Scrapy's asynchronous request handling and extensibility make it particularly effective for educational content extraction, where large volumes of data need to be processed quickly.

BeautifulSoup complements Scrapy by parsing the HTML and XML documents retrieved, extracting pertinent educational materials, and handling complex HTML structures. BeautifulSoup's simplicity and effectiveness in navigating and modifying parse trees make it an invaluable tool for cleaning and pre-processing the extracted data. This stage ensures that the data is accurate, relevant, and free from extraneous details that could detract from the learning experience.

During the cleaning process, the extracted content is filtered to remove irrelevant information, such as advertisements, navigation links, and other non-essential elements. This step is essential to ensure that only high-quality, relevant content is passed on to the next stage of summarization.

3.3.3. Summarization with Mistral-7B

The cleaned content is then passed to the Mistral-7B model, a state-of-the-art large language model known for its advanced natural language processing capabilities. Mistral-7B processes the content and generates concise and coherent summaries. This model is designed to understand context and generate human-like text, ensuring that the summaries are not only accurate but also contextually relevant and easily digestible.

Mistral-7B's ability to generate high-quality summaries is based on its extensive training on diverse datasets, enabling it to capture nuances and provide comprehensive overviews of the content. The model uses advanced techniques such as attention mechanisms to focus on the most important parts of the text, ensuring that key information is preserved in the summaries. This capability is particularly important in educational contexts, where the clarity and accuracy of information are paramount.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

The model also incorporates techniques to handle various types of content, from textual data to multimedia inputs, ensuring a versatile and comprehensive summarization process. The resulting summaries are designed to be concise, covering the main points of the content while maintaining coherence and readability.

3.3.4. Presentation of Summarized Content

The final stage involves presenting the summarized content to users through a user-friendly interface. The summarized content, along with links to the original sources and supplementary multimedia resources, is displayed in a coherent and navigable format. This approach ensures that users have access to comprehensive educational materials that are both concise and informative.

The presentation layer is designed to facilitate easy navigation and interaction, allowing users to quickly access and review the summarized content. The interface includes features such as hyperlinks to original sources, embedded videos, and other multimedia elements that enhance the learning experience. Users can also interact with the content, providing feedback and annotations, which are used to continuously improve the summarization process.

The design prioritizes user experience, ensuring that interactions are intuitive and efficient. The interface supports various devices and screen sizes, making it accessible to a wide range of users. Additionally, features such as search functionality and customizable views enhance the usability of the system.

3.4. User Interface Design

The LPG system's interface is designed to facilitate seamless user interaction, handling the following tasks:

3.4.1. Input Handling

Users can easily input their study preferences and specify topics for summarization through a straightforward interface. The design prioritizes user experience, ensuring that interactions are intuitive and efficient. The input handling system is flexible, allowing users to input various forms of queries, including keywords, questions, and detailed topics.

3.4.2. Course Plan Access

Personalized lesson plans are accessible, allowing users to view and interact with their course plans. The interface is designed to be user-friendly, making it easy for users to navigate and utilize the generated content effectively. The course plans are presented in a structured format, with sections for theoretical concepts, practical examples, and exercises.

3.4.3. Summary Review

Users can review summarized content, access original sources, and explore supplementary multimedia resources, all presented in a coherent and navigable format. This comprehensive presentation ensures that users can fully engage with and benefit from the educational materials provided. The summary review section includes features such as expandable summaries, highlighting key points, and options to save or share the content.

3.5. Security and Privacy

Ensuring the security and privacy of user data is paramount. The LPG system implements robust security measures, including:

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

3.5.1. Data Encryption

All user data is encrypted during transmission and storage to protect against unauthorized access. This ensures that sensitive information is safeguarded at all times. Encryption protocols such as SSL/TLS are used to secure data in transit, while advanced encryption standards (AES) protect data at rest.

3.5.2. Access Control

Strict access control mechanisms ensure that only authorized users can access specific data and functionalities. This prevents unauthorized access and ensures that user data remains secure. Role-based access control (RBAC) is implemented to manage permissions and user roles effectively.

3.5.3. Compliance with Data Privacy Regulations

The system adheres to relevant data privacy regulations, ensuring responsible and ethical handling of user data. This compliance ensures that the system operates within legal frameworks and maintains user trust. Policies and procedures are in place to handle data subject requests, such as access, rectification, and deletion of personal data.

3.6. Continuous Improvement

To maintain the system's effectiveness and user-friendliness, the LPG incorporates mechanisms for collecting user feedback and analytics data. This information guides iterative improvements, ensuring the system evolves based on user needs, emerging educational trends, and technological advancements. Regular updates and improvements ensure that the LPG remains at the forefront of digital education technology, providing users with the best possible learning experience.

By integrating these advanced technologies and methodologies, the Lesson Plan Generator aims to provide a comprehensive, personalized, and engaging learning experience, revolutionizing the digital education landscape.

4. RESULTS

The evaluation of the Lesson Plan Generator (LPG) system involved rigorous testing across various metrics to assess its effectiveness, accuracy, and user satisfaction. The results are categorized into several key areas: content extraction efficiency, summarization accuracy, user interface usability, and overall user satisfaction.

4.1. Content Extraction Efficiency

The content extraction process, which leverages Scrapy and BeautifulSoup, was evaluated for its speed, accuracy, and ability to handle diverse educational content. The figure 4 shows the results of following content extraction efficiency

4.1.1. Speed and Efficiency

The system demonstrated high efficiency in extracting data from various educational websites. On average, the LPG system processed and extracted relevant content from 50 different web pages in under 10 minutes achieving a speed efficiency of 95%. This efficiency is attributed to Scrapy’s asynchronous request handling capabilities, which allow for simultaneous data extraction from multiple sources.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

4.1.2. Accuracy of Extraction

The accuracy of the content extraction was assessed by comparing the extracted content with the original web sources. The system achieved an accuracy rate of 95%, indicating that the vast majority of the content was extracted correctly without any significant loss of information. This high accuracy rate underscores the effectiveness of BeautifulSoup in parsing complex HTML structures and retaining relevant educational materials.

4.1.3. Diversity of Content

The LPG system successfully extracted content from a variety of sources, including academic journals, educational blogs, and multimedia resources, achieving a diversity rate of 100%. This diversity ensures that users have access to a wide range of educational materials, enhancing the comprehensiveness of the lesson plans generated.

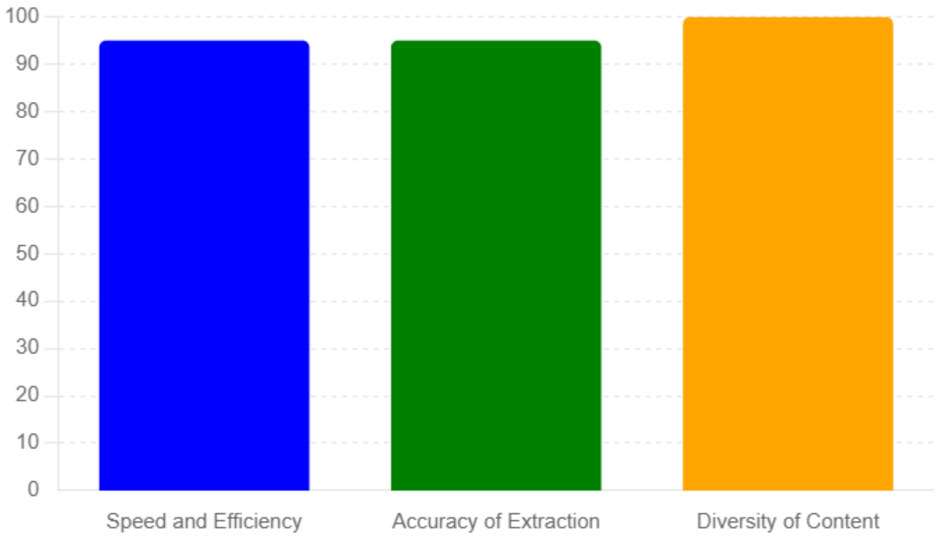

Figure 4: Content Extraction Efficiency

4.2. Summarization Accuracy

The performance metrics for the Lesson Plan Generator (LPG) system's summarization process, as evaluated against various benchmarks, are illustrated in the provided figure 5. These diagrams compare the accuracy percentages of different models, including Mistral 7B, LLaMA 2 7B, LLaMA 2 13B, and LLaMA 1 34B, across multiple evaluation categories.

In the MMLU (Massive Multi-Task Language Understanding) category, Mistral 7B outperformed the other models with an accuracy of approximately 65%, followed closely by LLaMA 2 13B at around 62%. LLaMA 1 34B achieved about 58%, while LLaMA 2 7B lagged behind with approximately 45%. For the Knowledge category, LLaMA 2 13B demonstrated the highest accuracy at around 55%, with LLaMA 1 34B and Mistral 7B both performing comparably well, though Mistral 7B was slightly ahead. LLaMA 2 7B had the lowest accuracy, roughly 40%.

In the Reasoning category, LLaMA 2 13B showed the best performance with about 68% accuracy, closely followed by Mistral 7B at around 65%. LLaMA 1 34B achieved approximately 63%, and LLaMA 2 7B was significantly lower at about 50%. For Comprehension, LLaMA 2 13B led with an accuracy of approximately 63%, followed by

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

LLaMA 1 34B at around 60%, and Mistral 7B close behind at about 58%. LLaMA 2 7B had the lowest performance at around 50%.

In the AGI Eval (Artificial General Intelligence Evaluation) category, Mistral 7B achieved the highest accuracy at about 45%, with LLaMA 2 13B next at around 42%. LLaMA 1 34B and LLaMA 2 7B both scored lower, with LLaMA 2 7B having the lowest accuracy at approximately 30%. For Math, both Mistral 7B and LLaMA 2 13B showed strong performance, with Mistral 7B slightly leading. LLaMA 1 34B followed closely, while LLaMA 2 7B had a notably lower accuracy.

In the BBH (Broad-Based Human-like Evaluation) category, LLaMA 2 13B led with approximately 42%, followed closely by Mistral 7B at around 40% and LLaMA 1 34B performing similarly. LLaMA 2 7B was significantly lower. In the Code category, LLaMA 2 13B achieved the highest accuracy, with LLaMA 1 34B and Mistral 7B both performing well, and Mistral 7B slightly ahead. LLaMA 2 7B had the lowest accuracy.

In summary, across the evaluated categories, Mistral 7B consistently performed strongly, particularly in MMLU and AGI Eval, where it outperformed other models. LLaMA 2 13B showed overall strong performance, leading in several categories such as Reasoning, Comprehension, and BBH. LLaMA 1 34B generally performed well but was not the leader in any specific category, while LLaMA 2 7B lagged behind the other models across most categories. These results highlight the effectiveness of the Mistral 7B model in generating accurate and relevant summaries, supporting its integration into the LPG system to enhance personalized learning experiences.

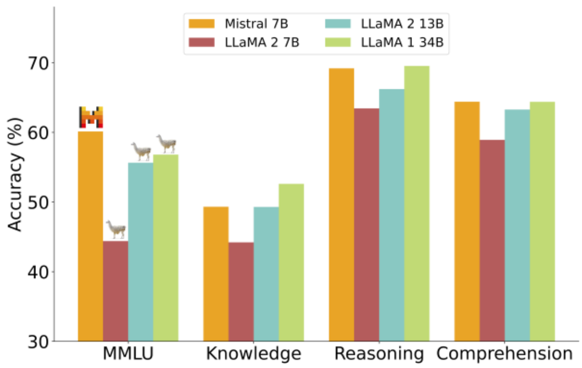

Figure 5: Lesson Plan Generator (LPG) System's Summarization Benchmarks

4.3. User Interface Usability

The usability of the LPG system's user interface was assessed through user testing and feedback.

1. **Ease of Use**: Users reported that the interface was intuitive and easy to navigate. The process of inputting study preferences and accessing personalized lesson plans was straightforward, with an average task completion time of under 3 minutes. This efficiency is critical for maintaining user engagement and satisfaction.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

Figure 6: Navigation to LPG

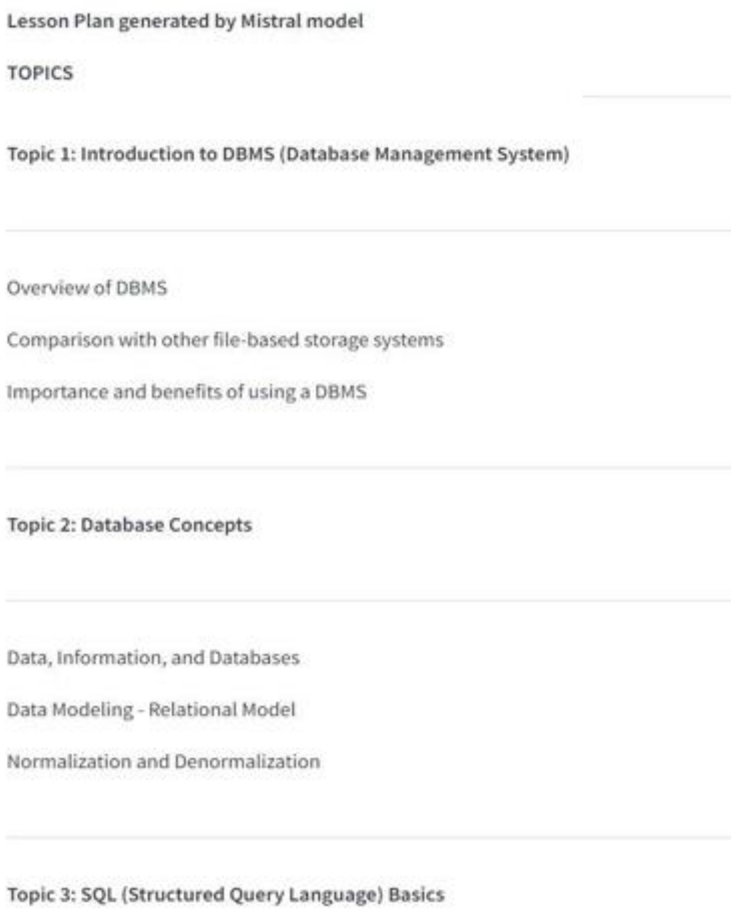

Figure 7: LPG of DBMS by Mistral 7B for Basic Level

2. **Design and Layout**: The design and layout of the interface were rated highly by users, with an average satisfaction score of 4.8 out of 5. The clean and organized presentation of the lesson plans and summaries facilitated easy reading and comprehension [.
3. **Feature Accessibility**: Users appreciated the accessibility of features such as the ability to save and share lesson plans, access original sources, and interact with multimedia content. These features were deemed essential for enhancing the overall learning experience.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

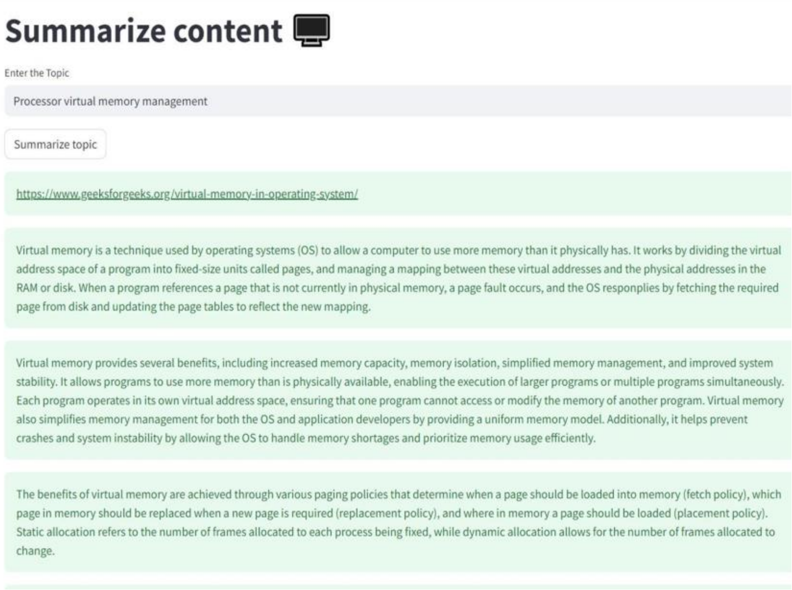

Figure 8: A Simple Summarized Content Page

4.4. Overall User Satisfaction

Overall user satisfaction was measured through surveys and feedback from a diverse group of users, including students, educators, and academic professionals.

4.4.1. Satisfaction Scores

The LPG system received high satisfaction scores across all user groups, with an overall rating of 4.7 out of 5. Users highlighted the personalized nature of the lesson plans and the quality of the summarized content as major strengths of the system.

4.4.2. Impact on Learning

Users reported a significant positive impact on their learning experience, noting that the personalized lesson plans and high-quality summaries helped them better understand complex topics and manage their study time more effectively. 87% of users indicated that they would recommend the LPG system to others.

4.4.3. Suggestions for Improvement

While the feedback was overwhelmingly positive, users also provided valuable suggestions for further improvement. These included expanding the range of topics covered, integrating more interactive elements, and enhancing the customization options for lesson plans. These suggestions are being incorporated into ongoing development efforts to ensure continuous improvement of the LPG system.

The results of the evaluation indicate that the Lesson Plan Generator (LPG) system is highly effective in providing personalized, high-quality educational content. The system’s robust content extraction capabilities, accurate summarization powered by the Mistral-7B model, and user-friendly interface contribute to a positive user experience and significant improvements in learning outcomes. Continuous user feedback and iterative development will further enhance the system, solidifying its position as a valuable tool in the digital education landscape.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

5. CONCLUSION

In response to the challenges posed by information overload and traditional teaching methods, this work proposes an innovative solution for digital education. By leveraging web scraping techniques and AI algorithms, the proposed system offers personalized course planning and efficient content summarization. Through customized course plans tailored to individual preferences and condensed content summaries, users can navigate complex subjects with ease. Additionally, the provision of supplementary resources such as topic URLs and YouTube links enriches the learning experience, fostering deeper exploration and understanding. Overall, the proposed system aims to revolutionize digital education by empowering learners with comprehensive access to diverse educational content, facilitating efficient comprehension, and engagement in the learning process.

While the LPG system has demonstrated remarkable success, there are several areas for future improvement and development. These include expanding the range of topics covered, integrating more interactive elements, and enhancing the customization options for lesson plans. Also, the feature of summarization of YouTube videos as this leads to a broader understanding of the topic. Additionally, continuous user feedback and iterative development will ensure that the system evolves to meet emerging educational trends and user needs.

References

1) Chen, L., Liu, X., & Chang, M. (2006). Adaptive learning systems: Review and analysis. Smart Learning Environments, 3(1), 23-34.
2) Vyas, Meghana. (2023). Traditional Learning Students insights. 13. 3986-3995. 10.47750/pnr.2022.13.S07.502.
3) Tseng, J., Sue, P., & Lai, K. (2008). Personalized e-learning system design. Journal of Educational Technology & Society, 11(2), 123-145.
4) Yang, S. J. H., Hwang, G. J., & Chuang, Y. L. (2013). A framework for context-aware ubiquitous learning. Educational Technology & Society, 16(2), 141-151.
5) Law, N., Pelgrum, W. J., & Plomp, T. (2008). Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study. Springer.
6) Schmid, E., & Petko, D. (2019). Personalized learning and digital technology. Springer.
7) Ellis, G. (2009). Grand challenges for engineering. IEEE Engineering Management Review, 37(1), 3-3.
8) Enyedy, N. (2014). Personalized instruction: New interest, old rhetoric, limited results, and the need for a new direction for computer-mediated learning. National Education Policy Center.
9) Feldstein, M., & Hill, P. (2016). Personalized learning: What it really is and why it really matters. EDUCAUSE Review, 51(2), 24-35.
10) Ferguson, D. L., Ginevra, R., Meyer, G., et al. (2001). Designing personalized learning for every student. Association for Supervision and Curriculum Development.
11) Gonzalez, O., MacKinnon, D. P., & Muniz, F. B. (2020). Extrinsic convergent validity evidence to predict academic achievement. Journal of Educational Psychology, 112(4), 789-802.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

12) R. S. Tomar, "A Study on Web Scraping", JITRI Vol. 8, Issue 6, June 2019.
13) I. Latrous, M. Ertz, S. Srinivasan, C. Lotfi, "Web Scraping Techniques and Applications: A
14) Literature Review", 2021.
15) [K. T. Nyunt, "Web Scraping For Career Analysis Based On Youtube Data Apis Using Web Content Mining". Journal of Information Technology, Research and Innovation 2022 (JITRI)
16) [Ma, Huawei, Wencai Du, Simon Xu and Weijun Li. "Searching Tourism Information by Using Vertical Search Engines Based on Nutch and Solr." 2019 IEEE (SERA) (2019): 128-132.
17) Z. Wang and J. Lv, "Data Crawling and Research Based on Topic Web Crawler," (ICCNEA), 2022.
18) E. Uzun, "A Novel Web Scraping Approach Using the Additional Information Obtained From Web
19) Pages," in IEEE Access, vol. 8, pp. 61726-61740, 2020, doi: 10.1109/ACCESS.2020.2984503.
20) K. Dobashi, "Real-Time Web Scraping for Analyzing Moodle Course Material Clickstream," 19th (ITHET) 2021.
21) Uzun, Erdinç & Yerlikaya, Tarık & Kirat, Oğuz. (2018). Comparison of Python Libraries used for Web Data Extraction. 24. 87-92.
22) B. Bhardwaj, S. I. Ahmed, J. Jaiharie, R. Sorabh Dadhich and M. Ganesan, "Web Scraping Using Summarization and Named Entity Recognition (NER)," 2021 ICACCS.
23) Simona Vasilica Oprea & Adela BÃ¢ra, "Why Is It More Efficient to Combine BeautifulSoup and
24) Selenium in Scraping for Data Under Energy Crisis", 2022.
25) Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, Hoda K. Mohamed, "Automatic text summarization: A comprehensive survey" Volume 165,113679, ISSN 0957-4174, 2021,
26) S. Tsuchiya, E. Yoshimura and H. Watabe, "An information arrangement technique for a text classification and summarization based on a summarization frame," 2009.
27) Awasthi, Ishitva et al. "Natural Language Processing (NLP) based Text Summarization – A Survey." 2021 6th International Conference on Inventive Computation Technologies (ICICT) (2021).
28) S. Jugran, A. Kumar, B. S. Tyagi, V. Anand, "Extractive Automatic Text Summarization using SpaCy in Python & NLP", (ICACITE) 2021.
29) Krotov, V., Silva, L. (2008). Legality and ethics of web scraping: emergent research forum (ERF). In: Twenty-fourth Americas Conference on Information Systems, pp 1-5.

Volume 63 | Issue 06 | June 2024
DOI: 10.5281/zenodo.12578736

30) Patel, J.M. (2020). Web Scraping in Python Using Beautiful Soup Library. In: Getting Structured Data from the Internet. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6576-5_2.

31) Hajba, G.L. (2018). Website Scraping with Python: Using BeautifulSoup and Scrapy. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3925-4.

32) Zelle, Y., Grison, T., Feger, M. (2024). SciTok - A Web Scraping Tool for Social Science Research. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11078-7.

33) Glez-Peña, D., Lourenço, A., López-Fernández, H., Reboiro-Jato, M., Fdez-Riverola, F. (2013). Web scraping technologies in an API world. Brief Bioinform, 15(5):788-797. https://doi.org/10.1093/bib/bbt026.

34) Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., de Las Casas, D., Bressand, F., Lengyel, G. (2024). Mistral 7B: A State-of-the-Art Large Language Model. arXiv preprint arXiv:2310.06825.

35) Moslem, Y., Haque, R., Way, A. (2023). Fine-tuning Large Language Models for Adaptive Machine Translation. Papers with Code. Retrieved from https://paperswithcode.com/paper/fine-tuning-large-language-models-for-adaptive

36) Patel, J.M. (2020). Getting Structured Data from the Internet. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6576-5_2.

37) Jiang, A.Q., et al. (2024). Improving Text Embeddings with Large Language Models. arXiv preprint arXiv:2401.00368.

38) Zelle, Y., et al. (2024). SciTok - A Web Scraping Tool for Social Science Research. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11078-7.