

Automated Doubt Resolution System for Students

1st Christa Carmel
Scholar

Department of Computer Science
Sacred Heart College Thevara,Kerala
christacarmel7@gmail.com

2nd Angel Mary D Coutha
Scholar

Department of Computer Science
Sacred Heart College Thevara,Kerala
angelmary3505@gmail.com

3rd C.Vishnu Mohan
Assistant Professor

Department of Computer Science
Sacred Heart College Thevara,Kerala
vishnumohan@shcollege.ac.in

Abstract—This project is to develop an Automated Doubt Resolution System, capable of providing students' doubts to be solved accurately, fast, with the context of academic support. Then the system will be capable of leveraging techniques from artificial intelligence (AI), natural language processing (NLP), and deep learning, and combining the semantic similarity models like BERT and SBERT to develop the project's search, recommendation, and classification capabilities so that it uses a range of educational materials like textbooks, lecture notes, online forums to classify queries, recognize the intent of the query, and perform semantic search. Ultimately, the system has a feedback loop that will allow students to rate the provided responses to improve accuracy and relevancy continuously. The research and methodology are to undertake data collection, data type definition and preprocessing, data analysis through the development of semantic encodings, and possible outcome classification/model variation so that the system utilizes useful information or develops and provides accurate and supportive answers to complex questions. The work can be implemented as a prototype on a laptop or sophisticated server through the use of Python and existing NLP frameworks/libraries. The evaluation methods for the project are standard metrics from machine learning and information retrieval evaluations against benchmark datasets. This project departs from traditional tutoring systems to investigate a system that provides a combination of Semantic retrieval and active feedback loops, and generative models will create a scalable, autonomous, and intelligent Automated Doubt Resolution System that will enhance student engagement and learning outcomes in virtual education contexts

Index Terms—Artificial Intelligence, Natural Language Processing, Deep Learning, Machine Learning, Semantic Search, BERT, Virtual Teaching Assistant, Automated Tutoring, Query Classification

I. INTRODUCTION

Learners need to receive timely feedback on their uncertainty for effective learning today. Face-to-face timetabled learning is normally limited by the time availability of teachers and does not normally yield timely feedback for learner uncertainty. If learning is asynchronous (offline, not face-to-face or self-started) or online self-study (independent or peer facilitated), it is even slower. Accordingly, there is an urgent need for a real-time solution, or at least a scalable solution for learning in some future context to ask some sort of question and be able to receive feedback right there and then, rather than waiting for a response with other forms of asynchronous

communication. Long delays when learning begins to learn, stopping learning for some time frame, potentially reducing motivation to learn independently.

Advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and semantic search have enabled the creation of automated doubt-resolving systems that can analyze a learner's natural language query and present accurate contextually aware explanations. These models only differ from static FAQ and contextually unaware rule-based bots in that they are capable of semantic similarity and adaptive learning over time, meaning they offer a deeper understanding of the learner's context and support offered.

Though advancements are being made, the various systems still cannot handle ambiguous queries, ambiguities in terms, or different levels of learners. A student can vary in terms of academic expertise and experience. Although students have varying backgrounds, current systems are unable to provide effective personalized explanations to students who have differing backgrounds in their academic expertise and experience.

This project addresses these challenges through a three-layer hybrid methodology. The first layer employs semantic similarity models like BERT and SBERT to interpret queries. The second integrates a knowledge base of curated educational resources to provide accurate responses. The third introduces adaptive feedback, where student ratings refine future answers. This system reduces manual dependency while supporting personalized, scalable, and self-directed learning in both virtual and traditional educational settings.

II. RELATED WORKS AND BACKGROUND

In the history of AI, conversational agents have evolved significantly from early rule-based systems to today's sophisticated data-driven models. The shift was driven by Seq2Seq architectures, which use encoder and decoder frameworks, now predominantly built with Transformers and their attention mechanisms, to process and generate sequences. These models [1] are underpinned by neural network concepts like back propagation and are trained on language representations that have advanced embeddings to deep contextual models like BERT and GPT. Scotti, Sbattella & Tedesco reviews three primary language modeling approaches—causal, bi-directional,

and transducer models—and explores how these are adapted for dialogue. It also involves into more advanced techniques, such as hierarchical and latent variable models that capture high-level dialogue structure and improve diversity, as well as crucial decoding strategies like beam search and contrastive search, which are used for controlled and coherent generation.

Intelligent Tutoring Systems (ITS) can be considered as a key solution for the global education crisis of over 250 million children out of school. The AI-powered platforms like Duolingo and Cognitive Tutor are designed to personalize learning by supporting to a student's individual needs, offering immediate feedback, and self-regulation. Previous research by Kulik & Fletcher [2] has shown that ITSs can be as effective as human tutors, especially when they are designed with solid principles and include features like guided practice and adaptivity. This literature has significant gaps including many studies are short-term, conflate different age groups, and often overlook critical ethical considerations such as transparency and fairness. L'etourneau, A et al. proposes that for ITSs need to be integrated with teacher-led instruction and require more effective, long-term research that specifically focuses on K-12 settings and addresses the ethical implications of using AI in education.

Sustainable education aims to meet learning needs for present and future generations using AI by offering personalized learning, enabling data-driven decisions, and boosting access to everyone. The evolution of AI-powered systems, particularly Intelligent Tutoring Systems including advanced machine learning and NLP to adapt to students' needs in real time. These systems are part of a trend of Technology-Enhanced Learning that includes tools like social media and VR. The integration of AI into these systems [3] can be either non-intrusive, for external data analysis, or embedded, for real-time adjustments. While this integration is powerful, with common techniques including supervised and unsupervised learning for prediction and behavior analysis, the text notes several challenges. These include a lack of infrastructure, limited teacher training, and significant concerns around privacy, bias, and the transparency of AI's decisions, a problem that Explainable AI (XAI) is trying to solve. The research trends show a focus on higher education and specific geographic regions, with growing interest in areas like student engagement and dropout prevention.

Prompted by the increasing demand for intelligent support in digital learning, this paper summarizes the state of Question Answering Systems (QAS), which are an integration of Natural Language Processing (NLP), Information Retrieval, and Information Extraction. Soares, T.G., Azhar et al. specifically focuses on the application of QAS in education, categorizing existing systems across five domains: Information Technology & Communication Science, Social Sciences, Language Sciences, Biomedical Sciences, and Other Sciences. Each domain utilizes specific techniques such as TF-IDF, neural networks and ontology to address its unique educational needs. The field

[4] also faces several limitations, including domain-specific constraints, a lack of voice interaction, and the need for better algorithms to handle complex queries. The evaluation of these systems [4] can be made by using metrics like accuracy and F1-score. Significant gaps in the current research identified are a lack of focus on curriculum alignment and pedagogical integration.

Driven by the limitations of traditional keyword-based search, which often fails to grasp context, the field of semantic search is BERT (Bidirectional Encoder Representations from Transformers), a powerful language model that understands the contextual relationships between words. SBERT (Sentence-BERT) refines this by creating meaningful sentence-level embeddings, making it a perfect fit for semantic similarity tasks within search engines. This paper surveys [5] how BERT is being integrated with search platforms like Elastic Search to enable scalable, vector-based semantic searching. It highlights a variety of related works, including SemanTelli, a meta-semantic search engine, and studies demonstrating BERT's superior performance in academic, enterprise, and medical search. The review also touches on the use of neural search engines, which employ deep learning for ranking, and how these technologies are being applied to optimize niche queries, personalize news feeds, and power sophisticated Question Answering Systems.

Sequence-to-Sequence (Seq2Seq) models are widely used for Natural Answer Generation (NAG) in chatbots have significant weaknesses, such as generating generic or inconsistent responses. These flaws stem from issues like insufficient input context, a tendency of the cross-entropy loss function to favor common words, difficulty with rare terms, and a lack of coherence over long conversations. To combat these problems, researchers have proposed a variety of enhancements. Structural modifications involve adding embeddings for persona or emotion and using hierarchical encoders to better manage dialogue history. Augmented learning strategies [6] include adopting alternative loss functions like Maximum Mutual Information (MMI) to boost response diversity and using multi-task learning or reinforcement learning to optimize for multiple objectives. Finally, complementary mechanisms like copy mechanisms and advanced attention architectures allow the model to directly use information from the input and external knowledge bases, thereby improving the relevance and specificity of its answers.

Gligore et al. [7] synthesizes 63 articles from the years 2010 to 2023 to assess the role of AI and Machine Learning in the context of adaptive e-learning systems. These systems provide the customized educational experience that they do by dynamically modifying the content and teaching the relevant material based on learner input. The study also attempts to collect all the AI/ML techniques relevant to this problem, which include supervised learning models such as SVMs and decision trees, unsupervised clustering, deep learning, and reinforcement and deep learning. They help to create person-

alized learning pathways, provide real-time feedback, detect learning styles, recommend content tailored to the learner, and make predictions about students' educational outcomes. The review cites various advantages including heightened learner engagement, improved academic performance, and optimized content delivery systems. At the same time, it also raises some of the more serious problems such as the privacy of information, the 'cold-start problem' experienced by new users, biased algorithms, and difficulty in integration.

The evolution of Intelligent Tutoring Systems (ITS) went from highly structured early systems like SCHOLAR to contemporary ones that use AI. The most significant change came from the application of AI, particularly through machine learning (ML) and natural language processing (NLP). Modern ITS applications feature ML to assess student interaction and change instructions accordingly, utilizing everything from supervised learning on a prediction level to reinforcement learning on path optimization. NLP facilitates understanding and responding to student input. The application of data integration in analytics, focusing on student behavior, and HCI frameworks aimed at creating compelling interfaces add to the systems' effectiveness. These included [8] educational programs, MATHia from Carnegie Learning, Duolingo, and DreamBox Learning, which the paper describes as successful and highlight the effectiveness of tailored teaching. It equally draws attention to the critical risks such as privacy and algorithmic concerns as well as the constant need for maintenance, technical synergies, and system outreach.

Intelligent Tutoring Systems (ITS) are AI-powered platforms designed to mimic human tutors, and they are defined by four core components like the student model, the teacher model, the domain model, and the diagnosis model. Early ITS like SQLTutor and Cognitive Tutor were crucial in establishing personalized learning, but the field continues to grapple with significant challenges. These include the difficulty of creating high-quality domain models, the high cost of development [9] and training data, and a lack of adaptability. Due to these limitations, some researchers advocate for a hybrid approach, using "Intelligent Teaching Assistants" to augment human educators rather than replacing them. The paper proposes a semi-automated assessment generation module to address one of these challenges, using a combination of rule-based methods and machine learning. The system uses Finite State Automata for generating question templates and encoder-decoder neural networks.

Recognizing the challenges posed by manual methods of software testing, automating testing processes became essential for modern software quality assurance, particularly for intricate systems, fast-paced development timelines, and agile frameworks. These systems are beneficial because they improve overall efficiency, ensure tests can be repeated, and provide wider test coverage which helps detect defects earlier. While automated testing offers numerous advantages, it also presents automation specific challenges such as high overhead

costs, the requirement of specific programming competencies, and lack of automation abilities to replace human judgement in areas such as usability testing. Hanushali [10] examines different automation frameworks as well as their Linear, Modular-Based, and Hybrid models which use multiple strategies for increased flexibility. Industrial and mobile case studies reveal automation enables up to 90% time savings in some instances, yet fails to detect subtle bugs that human tests would identify.

Taking into consideration the research conducted on the effectiveness and prospects of adaptive learning technology, it can be confidently said that it has the potential to be the next big thing in e-learning. De Castro & Perez [11] divides adaptive educational technologies into three groups: Intelligent Tutoring Systems (ITS), Adaptive Hypermedia Systems (AHS), and Intelligent Collaborative Learning Systems, focusing on their different functions. There is a distinction between intelligent and adaptive systems – the former use AI to enhance help, while the latter is based on a student model and attempts to tailor content to the learner. Most of the systems examined, including ITES, Logicando, and TANGOW, were found to have significant learning gains, and most of the studies included ITES reported a medium to large effect size (ES), especially for students with a lower prior knowledge. In a more future oriented view, the research still has great importance because it pointed out the problems of automating the content creating process, merging different data for better student modeling, and using intelligent agents for more flexible and widespread learning systems. This comprehensive report examined and analyzed the systems put forward.

Based on the historical survey of Intelligent Tutoring Systems (ITS), Alkhatlan & Kalita [12] traces their origins back to the 1970s, building upon the foundations of Computer-Assisted Instruction (CAI). Inspired by Bloom's 2-sigma study, which demonstrated the effectiveness of one-on-one tutoring, ITS development has been driven by the goal of simulating human tutors using AI. The paper reviews key general and specialized ITS surveys that have defined its architecture and explored various facets, from authoring tools to conversational interfaces. The traditional ITS architecture is composed of four core components: the domain model, student model, tutor model, and user interface. A significant portion of the research focuses on student modeling, outlining various techniques such as overlay, stereotype, and constraint-based models, as well as more advanced probabilistic approaches like Bayesian networks and fuzzy logic. The tutor model employs strategies like feedback loops and adaptive sequencing, with modern systems like AutoTutor and DeepTutor using natural language to enhance interaction.

Looking at how adaptive learning fits into running schools, it's clear it's important for where education is going, mainly because of all the changes happening in the industry. Adaptive learning is basically uses tech like data and machine learning to make lessons fit each student. It changes things up based on what they need, how fast they learn, and what they want

to get out of it and this makes learning feel more personal, works better, keeps students interested, and helps everyone, no matter how they learn. Also, teachers get information that helps them teach better. Backing these assertions are empirical observations of the effects from numerous studies, which, while confirming possibilities for rich learning experiences, also pose important questions regarding data privacy and the likelihood of potentially widening current socioeconomic disparities. The paper [13] suggests major challenges, including the requirement for strong technological infrastructure, effective teacher training, and responsible use of student data to ensure fairness. A survey-based conclusion supports the importance of adaptive learning, as students have strong agreement that it improves skill building, motivation, and preparation for future challenges.

Gomes D looks at how Intelligent Tutoring Systems (ITS) [14] have gotten better by adding Artificial Intelligence (AI). Starting in the 1970s, ITS changed from basic Computer-Assisted Instruction (CAI) by using AI to teach in a way that fits each student. The main parts of ITS are explained: a Student Module, a Teaching Module, a Knowledge Base, and a User Interface. These parts try to copy what a real tutor does. ITS is mostly used for college and IT classes. A big part of the study is about how AI is changing ITS by giving quick feedback, changing things as needed, and guessing what students will do next. AI does this by using things like Machine Learning, Natural Language Processing, and Bayesian Networks. Good examples, such as Duolingo and programs with talking robots, show how AI can be used to make learning plans and grade tests on its own. The study also talks about the different jobs AI can do in education and mentions important moral issues, like keeping data safe, avoiding unfair biases in algorithms, and making sure everyone has the same chance to use these tools.

Redmond et al. synthesizes the application of Natural Language Processing (NLP) techniques to analyze student feedback, outlining key methodologies, challenges, and emerging trends. The paper details several NLP techniques used to process textual feedback, including feature extraction methods like Bag of Words (BoW), TF-IDF, and advanced word embeddings like BERT, which can understand context. It also covers feature selection techniques to optimize model performance and topic modeling approaches [15] like LDA to uncover thematic patterns in feedback. The applications of these methods are diverse, ranging from text summarization and document categorization to sentiment annotation and the creation of knowledge graphs. However, the review also highlights significant challenges in this domain. These include dealing with domain-specific language, the difficulty of detecting sarcasm and ambiguity, and the need for specialized processing of emoticons. The issue of data imbalance in feedback types is also identified as a factor that can affect the accuracy of NLP models.

Tang et al. [16] provides a detailed review of recent

advancements and persistent challenges in Complex Question Answering (CQA) over Knowledge Bases (KBs). The primary methodologies are categorized into traditional rule-based methods, Information Retrieval (IR)-based methods, and Neural Semantic Parsing (NSP)-based methods. Rule based systems are limited by their reliance on hand-crafted templates, which makes them unscalable and poor at handling the diversity of complex questions. In contrast, IR-based methods treat QA as a semantic matching problem, and recent advances have incorporated multi-hop reasoning via memory networks and external knowledge from sources like Wikipedia, though they still struggle with interpretability and constraint-based questions. The most significant progress has come from NSP-based methods, which translate natural language into logical forms or query graphs. These approaches, including query-graph-based models like STAGG and encoder-decoder models like Seq2Seq, offer better coverage of complex semantics but are hindered by the need for annotated logical forms and the scarcity of training data.

Fitria explores how Artificial Intelligence (AI) is being integrated into teaching and learning. It outlines various tools and platforms that improve educational delivery. The review identifies several key uses of AI in education. These include virtual mentors that provide personalized feedback and resources, along with voice assistants like Google Assistant and Siri that allow hands-free content searching to improve accessibility. These platforms efficiently organize and deliver information. AI-enhanced presentation translators provide real-time subtitles that assist with language issues experienced by students in the classroom. The increasing inclusion of AI-enabled course platforms worldwide such as MOOCs and Coursera facilitate the ability for more personalized course recommendations and tracking of course progress. The paper [17] outlines many benefits of AI including improved personalization, more automated administrative tasks, however, also discusses limits of AI including decreased human interaction variable, associated costs, and dangers of cybersecurity and data privacy.

Deshmukh et al. addresses the background linking task in news articles by proposing two approaches: a weighted keyword-based BM25 retrieval and a semantic search method called IR-BERT, which combines BM25 with Sentence-BERT embeddings. Prior work relied heavily on keyword-based models like BM25 and Anserini, which lacked semantic understanding [18] and failed to capture contextual nuances. The authors bridge this gap by integrating deep learning techniques to enhance retrieval relevance and introduce a novel diversity metric to assess the variety of retrieved documents. Their IR-BERT model outperforms previous state-of-the-art methods on the TREC 2018 Washington Post dataset, demonstrating the effectiveness of semantic search in background linking.

Garg & Moschitti addresses the challenge of training answer generation (GenQA) models without relying on costly human-labeled data by proposing a novel knowledge transfer method

from answer sentence selection (AS2) models. Instead of using supervised GenQA data, the authors use a trained AS2 model to rank answer candidates, treating the top-ranked sentence as the generation target and the next k candidates as input context. They enhance this weak supervision with loss weighting based on AS2 confidence scores and score-conditioned input/output shaping. This approach [19] bridges the gap between ranking and generation tasks, enabling GenQA models to outperform both their AS2 teachers and fully supervised GenQA baselines across multiple datasets, including MS-MARCO, WikiQA, and TREC-QA.

Alhawiti explores the use of Natural Language Processing (NLP) in education and its impact on learning, assessment and content understanding by using a variety of linguistic tools, including grammar and syntax, and textual analysis. It provides a literature review of existing NLP tools, including e-rater, Text Evaluator, and Language Muse to illustrate their usefulness in supporting reading, writing, and curriculum development. Research limitations were highlighted due to the limited use of NLP tools in Arabic educational contexts and how there is a need for more intelligent systems aimed at filtering untrusted online sources. Based on a qualitative methodology using secondary data, this paper [20] illustrated how e-learning tools and assessments can be automated with the use of NLP, while enhancing the support available to teachers and students to better accommodate educational content.

Ahmed & Khan provides a comprehensive literature review on the application of Natural Language Processing (NLP) in education, organized around three core roles: assessing language (e.g., automated essay scoring and grammatical error detection), using language for instruction (e.g., dialogue-based tutoring systems), and processing language to support educational tasks (e.g., summarizing student responses or generating test questions). The methodology [21] follows an iterative research lifecycle, beginning with real-world educational needs, informed by pedagogical theory, and culminating in the design and evaluation of NLP technologies. A key research gap identified is the limited generalizability and real-time performance of existing NLP tools when applied to noisy, student-generated data, especially in formative assessment and open-ended learning environments like MOOCs. The paper highlights the need for pedagogically meaningful features, scalable systems, and extrinsic evaluations in authentic educational contexts to advance the field.

Dan, George& Loverth examines the changing role of Natural Language Processing (NLP) in education, with an emphasis on its use in communication, personalized learning, and teaching efficiency as well as its various applications. The methodology was conceptual and exploratory in nature and integrates existing technologies [22] and pedagogy to articulate a framework for how educational institutions can use NLP technology. This research has also identified critical gaps with NLP use in education, including, but not limited to, its implications for data privacy, algorithmic bias, and lack

of infrastructure to support scalability. This paper highlights the importance of ethical considerations, in terms of safety, responsibility, and equity, as well as better regulatory practices and design for inclusion, to support a vague, though important emerging role for NLP technology in education.

Medhat, A., & Ahmed, R investigates the use of Natural Language Processing (NLP) to analyze students' qualitative feedback to instructors, synthesizing 28 peer-reviewed studies across various educational contexts. The literature is categorized into six key aims: sentiment prediction, category and rating prediction, emotion analysis, opinion mining, lexicon creation, and statistical analysis. [23] Methodologically, the study follows Creswell's five-step approach, employing constant comparative analysis to extract themes related to aims, models, tools, data characteristics, and labelling practices. A major research gap identified is the limited application of NLP in non-English languages, lack of pedagogically grounded lexicons, and minimal use of machine labelling.analysis in diverse learning environments.

Almeida & Simoes explores the fundamental aspects and outcomes of personalized adaptive learning (PAL) in higher education, aggregated across 69 studies published between 2012 and 2024. It emerges from the literature that PAL presents opportunities for better academic performance and student engagement through individual learning pathways, adaptive platforms like Moodle and LearnSmart, and indicators such as pre-knowledge quizzes and learning analytics. Methodologically, the research follows the Joanna Briggs Institute framework, including systematic searching of databases, Covidence-based screening and narrative synthesis. Significant findings highlight a lack of standardization [24] in terminology, disproportionate representation of studies from developing regions, and inconsistent reporting of student engagement. While the review identifies areas for growth, these include more longitudinal studies, inclusive design of platforms, and analyses across different modalities of data to optimize scalability and effectiveness of PAL systems across diverse college environments.

Unger, Freitas & Cimiano reviews question answering (QA) systems over knowledge bases (KBs) with attention to the fact that the QA systems must be able to answer natural language queries based on structured data. The literature covering this topic is organized into a number of different key components such as semantic parsing, entity linking, relation extraction, and query generation. It is also organized according to the type of methods studied from rule-based approaches [25] to deep-learning architectures. In terms of method, the paper reviews existing QA frameworks and end-to-end benchmarks that utilize numerous datasets, including Freebase and DBpedia. A major gap in research is outlined in terms of the challenge multi-hop and complex questions pose for QA systems and the limited robustness of QA models in low-resource or narrow-domain environments. The review notes the need for interpretable models, for training models that

generalize across KBs, and for better reasoning over contextual and commonsense knowledge for further QA development.

Dwivedi & Singh provides a structured review of Question Answering (QA) systems, categorizing them into three primary approaches: linguistic, statistical, and pattern matching. The literature spans early systems like BASEBALL and ELIZA to modern hybrid models such as IBM's Watson, highlighting the evolution from domain-specific knowledge bases to web-scale, data-driven architectures. Methodologically, the paper analyzes QA systems based on their processing stages—question analysis, document retrieval, and answer extraction—while comparing techniques like NLP, machine learning classifiers, and pattern-based heuristics. A notable research gap noted within the literature [26] is there seems to be a lack of semantic understanding and scaling capacity across heterogeneous data, especially in the case of pattern-based systems. Previous authors have suggested hybrid approaches that leverage the respective strengths of the individual models to improve precision, adaptability, and user relevance across diverse QA environments.

Zhang has illustrated the application of Natural Language Processing (NLP) technology in intelligent education, as well as provided an "NLP and education" enabling model, underpinned by a double-circle mechanism that includes both coordinated internal and external factors, involving stakeholders, platforms, and infrastructure. The literature review accounted for the development of NLP in education through four key stages, from traditional machine learning approaches to large-scale models like ChatGPT, discussing the use of NLP across text classification, intelligent question and answer systems, as well as generation and translation. In terms of methodology, the conceptual modelling [27] and review of the literature were employed to assist in developing a framework for sustained transformation in education. In part, the main gap of the research is the lack of studies with systemic mechanisms, and issues associated with human resource training and development, technology adoption, and consumption of resources.

Roy et al. proposes a novel method to enhance Systematic Literature Reviews (SLRs) by combining Sentence-BERT (sBERT) for semantic similarity and Doc2Vec for document retrieval. The approach integrates text extraction, sentence embedding, cosine/TF-IDF similarity measures, and k-nearest neighbor retrieval to identify top-related documents. A review of prior models, including SPAR-4-SLR, S3BERT, and Refined SBERT, highlights the gap in scalable, automated semantic comparison tools for heterogeneous corpora. A prototype demonstrates [28] that packaging this workflow into a web interface can improve the speed, efficiency, and quality of literature discovery.

Sharma et al. outlines the development of semantic search technologies, with an emphasis on the change in context-aware systems from keyword-based information and retrieval, which has intensified in recent years with the Natural Lan-

guage Processing (NLP) capacity. This paper illustrates the journey from rule-based models to Latent Semantic Indexing into current models, such as Word embedding (Word2Vec) and the latest advancements, as in Deep Learning, including BERT. The approach has integrated Word2Vec for semantic meaning and subsequent Annoy Index as a way to run a nearly accurate nearest neighbor index that was tested using large datasets, including NEET 2024 results. It is shown that the system [29] has improved accuracy, scalability, and real-time performance over the first traditional search engines. However, the study shows areas to be improved, such as for extremely large datasets, accurately and appropriately trained data, and applying it in a variety of domains.

Masuda et al. explores semantic search systems enhanced by online integration of Natural Language Processing (NLP), focusing on improving search relevance and user experience beyond traditional keyword-based methods. The literature review covers foundational techniques such as TF-IDF and Latent Semantic Analysis, progressing to advanced models like BERT and GPT for contextual understanding. Methodologically, the study proposes a hybrid framework combining NLP-based semantic parsing with real-time query refinement and ranking algorithms. One of the major research gaps identified is the limited adaptability of current semantic search engines and systems to dynamically changing user intent in multilingual queries. The paper also highlights the need for more scalable, real-time architectures [30] and more diverse datasets that assist different user populations with evolving information needs.

Praveen Kumar et al. introduces a semantic search engine leveraging Natural Language Processing (NLP) to improve the relevance of search results by interpreting user queries beyond simple keyword matching. The literature review highlights prior efforts using ontology-based frameworks, lexical databases like WordNet, and hybrid models combining semantic and traditional search techniques. The proposed methodology [31] employs POS tagging via Stanford Parser, domain classification, and weight-based ranking of documents using noun and sentence matching, with special handling for negative expressions. Experimental evaluation using the American National Corpus demonstrates high accuracy (up to 97%) across varying document sets. However, the research identifies a gap in processing speed and scalability, suggesting future enhancements to reduce latency and support broader datasets while maintaining semantic precision.

Desai et al. identifies and evaluates the various approaches and metrics for semantic similarity, explaining the importance of these in the fields of Natural Language Processing (NLP), Artificial Intelligence (AI), and Information Retrieval (IR). The approaches are grouped into five categories: metrics-based approaches, corpus-based approaches, ontology-based approaches, relational-based approaches, and hybrid approaches, along with coverage of different techniques such as edge counting, Latent Semantic Analysis (LSA), WordNet-

based measures, and vector space models. On a methodological level, this paper [32] will discuss comparative studies and consider how approaches and techniques are evaluated based on datasets, precision, recall, and other correlation metrics. One of the research gaps highlighted is the absence of context-aware and globally-applicable measures of similarity, because most of the existing measures are limited by ambiguity, scalability, and domain-specific measures, or all of the above.

Gutierrez et al. critiques the limitations of traditional semantic similarity measures in geospatial contexts, arguing that semantic similarity alone fails to account for geographical proximity—a key factor in applications like location-based recommendation systems. The literature review spans semantic similarity models from TF-IDF and Word2Vec to BERT and BERTopic, highlighting their strengths in textual analysis but weaknesses in spatial relevance. Methodologically, the authors propose a novel approach combining topic modeling (LDA and BERTopic) with Moran's I spatial autocorrelation index to enhance embeddings with geographic indicators. Tested on Persian Wikipedia articles and real estate ads, the model significantly improves the correlation between semantic and geographical similarity. [33] The key research gap identified is the lack of integrated models that jointly consider semantic meaning and spatial distribution, especially in domain-specific datasets.

Hussain & Siddique describes a systematic mapping study of sentiment analysis focused on student feedback in educational settings from the perspective of natural language processing (NLP), machine learning (ML), and deep learning (DL) methods. The authors [34] evaluate and present 92 studies from 2015 to 2020 that were selected as the highest quality, using the PRISMA approach. They categorize the papers based on learning approach, methods (algorithms), data source, and evaluation methods. The review indicated that the field is progressing from NLP and ML-based to DL based methods, with mixed and lexicon-based methods increasingly used. They note that there is growing interest in studies like the one they undertook, but there were some notable gaps in the research, such as conducting fine-grained sentiment analysis, not detecting emotions, no benchmark datasets supported the studies, and no standardized tools and frameworks.

Wu & Yang introduces SEQ2SEQ++, a multitask learning-based sequence-to-sequence (Seq2Seq) model [35] designed to improve the quality of chatbot-generated answers by addressing key limitations in traditional Seq2Seq models—namely, generic and inconsistent responses. The literature review highlights prior efforts using attention mechanisms and multitask learning (MTL), noting that existing models often suffer from overfitting and rely on fixed task weights or unnatural binary classification. The proposed methodology integrates four novel components: a comprehensive attention mechanism (CAM), a dynamic task loss weighting scheme (DL), a multifunctional encoder (MFE), and a ternary classifier (TC). These components work in parallel to enhance answer generation,

classification, and prediction tasks. The research gap lies in the lack of holistic approaches that simultaneously tackle language model influence, answer generation overfit, and question encoder overfit. Experimental results on NarrativeQA and SQuAD datasets demonstrate that SEQ2SEQ++ significantly outperforms benchmark models in accuracy, error reduction, and response diversity.

Nayel & Elmaghraby provides a rigorous and systematic literature review (SLR) of existing Short-Text Semantic Similarity (STSS) research, in an important subarea of natural language processing for applications including popular topics such as question-answering systems, sentiment analysis, and information retrieval. The methods, overall, took the methodology to use PRISMA better into consideration, and the finished project [36] is broken down and described as planning, conducting, and reporting, taking into account strict inclusion and exclusion criteria. The authors point out that it is important to stress the major challenges involved in STSS, namely ambiguity, polysemy, sparsity, and non-uniform data. In addition, the contributors identify key knowledge gaps in that none of the models they use successfully defend the semantic differences involved in virtually identical short texts, nor provide exclusive universal quality benchmarks. Any further exploration of STSS should then include applying context information with embeddings, employing subject-specific datasets, and improving pre-processing models.

Makela provides a comprehensive survey of semantic search research, focusing on techniques that enhance traditional keyword-based search using semantic technologies or enable search over formally annotated semantic content. The literature review identifies five key research directions: augmenting keyword search, basic concept location, complex constraint queries, problem solving, and connecting path discovery. Methodologically, the paper [37] highlights common patterns such as RDF path traversal, keyword-to-concept mapping, graph pattern formulation, and the use of fuzzy logic for handling ambiguity. The research gap lies in the limited integration of reasoning-based problem solving and the underutilization of inference mechanisms due to scalability and complexity challenges. The study suggests that future work should aim to unify these methodologies into modular, interoperable systems that can support richer, context-aware semantic search experiences.

Diefenbach et al. describes a systematic review of question answering with knowledge bases (KBQA), by analyzing studies ($n=66$) published since 2015, and uses system architectures to classify the systems (semantic parsing pipelines, subgraph matching, template-based approaches, and information extraction). The studies addressed indicate a shift from rule-based systems to mostly neural-based models and that the field is shifting toward more modular and reusable systems and hybrid architectures. [38] The authors apply the PRISMA guidelines and the PICO framework to methodically review and select primary and secondary studies to ensure complete

and repeatable results. While KBQA studies are increasing in number, there are still several key research gaps: complex queries, incompleteness in the knowledge base, and improvements in entity-relation linking. Finally, the study revealed the community will benefit from more training data, more external knowledge, and stronger or more complex models such as reinforcement learning and conversational agents to improve KBQA performance.

[39] Aleksander examines the use of Natural Language Processing (NLP) with Adaptive Learning Technology (ALT) to facilitate Self-Regulated Learning (SRL) in educational contexts. In the review of literature, there is increasing evidence for the value of ALT in personalizing education and assisting with SRL, as well as demonstrating the ability of NLP to provide immediate contextual feedback. Prior research has shown that ALT can scaffold cognitive and metacognitive processes, but often lacks pedagogical grounding and qualitative insights. Methodologically, the study employed a one-year co-design process with a Swiss high school, combining qualitative interviews, surveys, and NLP techniques such as opinion mining, part-of-speech tagging, and sentiment analysis to refine a digital tool called studybuddy. The research gap lies in the limited use of NLP-informed ALT systems that are pedagogically grounded and responsive to learners' emotional and motivational states.

Fang et al. describes a unique architecture for task-oriented dialogue systems that combines the in-context learning capabilities of large language models (LLMs) with deterministic operation of business logic to address previous drawbacks of intent-based natural language understanding (NLU) systems. The literature assess also discusses previous solutions, like end-to-end learning, dialogue transformers, and dataflow-based models, and the features that hindered their scalability, data efficiency, and potential to utilize conversational aspects. The method uses command generation in a domain-specific language (DSL) to replaced intent classification, thereby enabling context-aware dialogue understanding without requiring training data annotated with intents. [40] It introduces components like dialogue stacks, conversation repair patterns, and flow-based task definitions to manage complex interactions. The research gap lies in the rigidity and scalability issues of intent-based systems, and the lack of modular, explainable, and developer-friendly frameworks for industrial-scale assistants.

Shameem & Kumar [41] investigates the role of Quality Assurance (QA) automation in supporting waste reduction in software development teams, primarily through productivity improvements and management of defects and resources. The literature review documented existing literature on agile frameworks, lean methods, and automated testing frameworks. Although literature exists on QA automation, little research, if not no research, focus on the role QA automation plays in waste elimination. The methodology for this study combined qualitative and quantitative research methods, and included case studies distilled from interviews and surveys of industry

practitioners regarding the effectiveness of QA automation technologies and practices. The research presents a gap in the literature linking QA automation directly with waste metrics related to rework, idle time, and excess communication overhead.

Shameem & Kumar [42] examined the effects of the integration of an adaptive learning (AL) tool, CogBooks®, into a university statistics course taught in both an online (FIT) and traditional face-to-face format. The literature review highlighted the expanding need for personalized education and the new role of AL to increase engagement, improve outcomes, and increase satisfaction. Prior research has shown mixed results with some research showing improved learning outcomes while other recent research described increased satisfaction without having a learning outcome improvement. Methodologically the analysis applied a comparative quantitative design over the course of three semesters, comparing exam scores, final grades and student course satisfaction surveys using statistical tests including Shapiro-Wilk, t-tests and Mann-Whitney U Tests. The significant research gap addressed in this study was the absence of empirical evidence of the effectiveness of AL in a higher education context, and limited research into the impacts of AL along both face-to-face and online delivery modes.

Fryer & Bovee [43] surveys the emerging role of Virtual Teaching Assistants (VTAs) in Chinese secondary education, particularly in response to the Double Reduction Policy aimed at reducing academic pressure and tutoring costs. The literature review highlights VTAs' potential in automating grading, generating personalized student performance reports, and supporting exam preparation, drawing on examples like Jill Watson and other AI-enabled tutoring systems. The approach taken is an in-depth review of the literature combined with qualitative interviews with 14 qualified representatives from academia, industry, and education. The authors look at the pedagogical, technical, and ethical aspects of the deployment of VTAs. The main research gap is the lack of empirical evidence for the impact of VTAs on improved student well-being and teacher workload, and the concerns surrounding emotional intelligence, data bias, and transparency regarding ethical practice. They recommend, as an alternative to potential blind adoption without stakeholder education, iterative development using a hybrid trial-based implementation model, so VTAs are positioned to add to teaching practices, rather than replace the human part of the teaching process.

III. METHODOLOGY

A. Dataset

The dataset used in this project was generated by collecting questions across multiple domains of Computer Science, including C Programming, Data Structures, Operating Systems, Computer Networks, and Database Management Systems. Each entry in the dataset consists of a student-style query and an expert-provided answer, along with a unique identifier and

a keyword that reflects the type of question. The final dataset contains 2463 question–answer pairs.

Unlike generic question-answer datasets, this one has been curated to reflect real classroom doubts. For instance, simple factual questions such as “Who developed the C language?” coexist with conceptual questions like “Why is C called a middle-level language?” and definitional queries such as “What is a token in C?”. The presence of varied question types ensures that the system can handle different learning contexts.

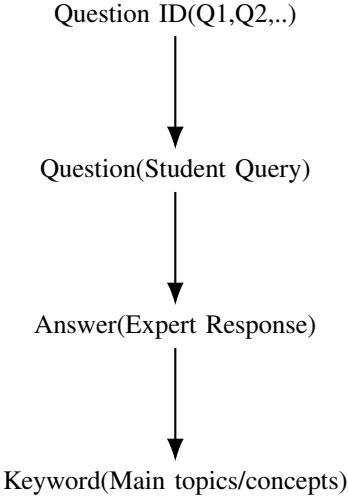


Fig. 1: Sample Dataset

B. Data Preprocessing

Data preprocessing is a crucial stage in the development of the automated doubt resolution system, as it transforms raw natural language input into structured forms suitable for computational analysis. Since the dataset contains 2463 question–answer pairs collected across various domains of Computer Science, the text naturally exhibits inconsistencies such as variations in capitalization, redundant spaces, punctuation, and different grammatical forms of the same concept. If left untreated, these irregularities would reduce the ability of machine learning models to capture meaning and retrieve the correct answers. Therefore, a comprehensive preprocessing pipeline was implemented to standardize the data, improve semantic consistency, and produce robust vector representations for retrieval.

C. Text Cleaning and normalization

The first step in preprocessing was text cleaning. All questions and answers were converted into lowercase to maintain uniformity and prevent the model from misinterpreting words like Array and array as distinct tokens. Unnecessary punctuation marks, digits, and special characters were removed to

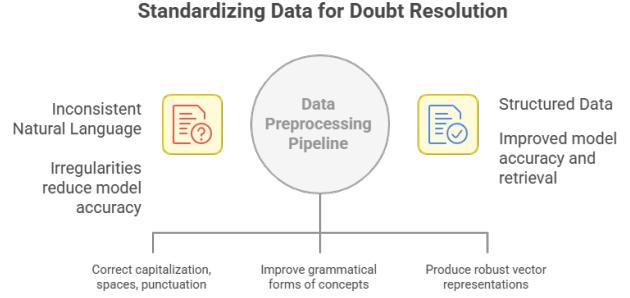


Fig. 2: Data preprocessing

reduce noise, while contractions were normalized so that isn't became is not. By carrying out these operations, the textual data was stripped down to its essential content, free of surface-level inconsistencies.

D. BERT WordPiece Tokenization

Once cleaned, the queries underwent tokenization. In this project, tokenization was carried out using the BERT Word-Piece tokenizer, which does not simply break sentences at spaces but further decomposes rare or unknown words into subword units. For instance, the word programming may be split into program and ming. This approach allows the system to handle a wider vocabulary by learning subword patterns instead of memorizing every complete word. Special tokens such as [CLS] (Classification) and [SEP] (Separator) were also inserted at the beginning and end of each input sequence to align with BERT's architecture, enabling the model to interpret the input consistently.

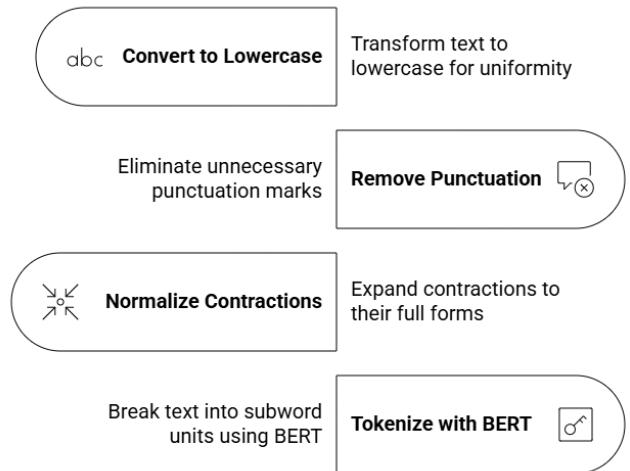


Fig. 3: Text Preprocessing and Tokenization

question_id	question	answer	keyword
1	Who developed the C language?	Dennis Ritchie developed the C programming language in 1972 at Bell Labs. He created, developed, language, when, C	
2	Q1	Why is C called a "middle-level" language because it combines the features of both low-level (like assembly, machine code) and high-level (like C, C++, Java) languages.	
3	Q2	Why is C called a "middle-level" language because it combines the features of both low-level (like assembly, machine code) and high-level (like C, C++, Java) languages.	
4	Q3	What's the basic structure of a typical C program includes preprocessor directives, a main function, variable declarations, basic structures, program logic, etc.	
5	Q4	What is a token in C?	
6	Q5	A token is the smallest individual unit of a C program. This includes keywords, identifiers, operators, and literals.	
7	Q6	What are identifiers?	
8	Q7	An identifier is a name given to a variable, function, or other user-defined term.	
9	Q8	What is the purpose of the #include directive?	
10	Q9	The #include directive instructs the C preprocessor to insert the content of another file, #include, preprocessor, directive	
11	Q10	What is the difference between A compiler translates the entire source code into machine code at once, generating, difference, compiler, interpreter	

E. Linguistic Filtering: Stop Word Removal and Lemmatization

Even after tokenization, many words in a query are semantically unimportant. For example, in the question “What is the definition of an array in C?”, words such as the and of do not contribute meaning to the query. These words, known as stop words, were removed so that the system could focus on more meaningful tokens such as definition, array, and C. In addition, lemmatization was applied to reduce different word forms to their dictionary base form. Thus, programming, programs, and programmed were all converted into program. This step ensured that questions expressed in slightly different grammatical forms could still be recognized as semantically related, which is particularly important in an educational dataset.

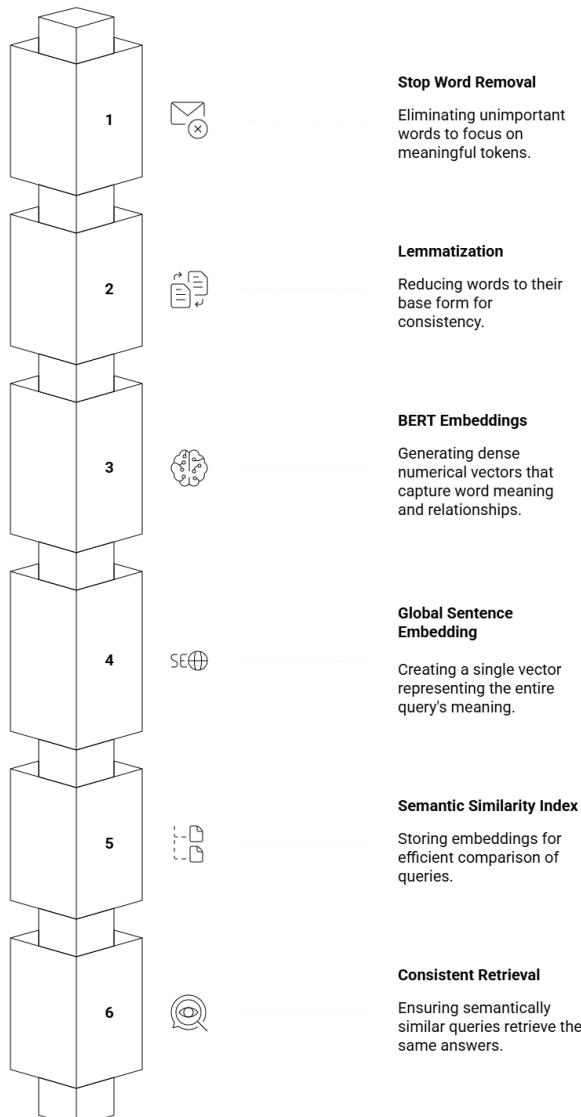


Fig. 4: Building a Semantic Search system

F. Vectorization and Semantic Embedding

After cleaning, tokenizing, and normalizing, the preprocessed text was ready for vectorization. Traditional approaches such as Bag of Words or TF-IDF were deemed insufficient because they treat words as independent units and fail to capture context. Instead, the project employed BERT embeddings, which provide dense numerical vectors that encode both the meaning of individual words and the relationships among them. When a query is passed through BERT, the model produces a global sentence embedding derived from the [CLS] token. These semantic embeddings are then stored in a similarity index, allowing the system to compare new queries with existing ones based on semantic closeness rather than superficial keyword overlap. As a result, semantically similar queries such as “What is meant by a token in C?” and “Explain C tokens” are mapped to nearly identical embeddings, enabling the system to retrieve the correct answers consistently.

G. System Architecture

The automated doubt resolution system is designed as a retrieval-based architecture that leverages natural language processing and semantic similarity techniques to provide relevant answers to student queries. The architecture is built on the principle that the meaning of a question can be captured through dense vector embeddings, which are then compared with embeddings of pre-stored questions and answers in the dataset. By adopting a retrieval-based approach, the system ensures that answers are factually accurate and consistent, since they are drawn directly from a curated knowledge base rather than being generated freely by the model.

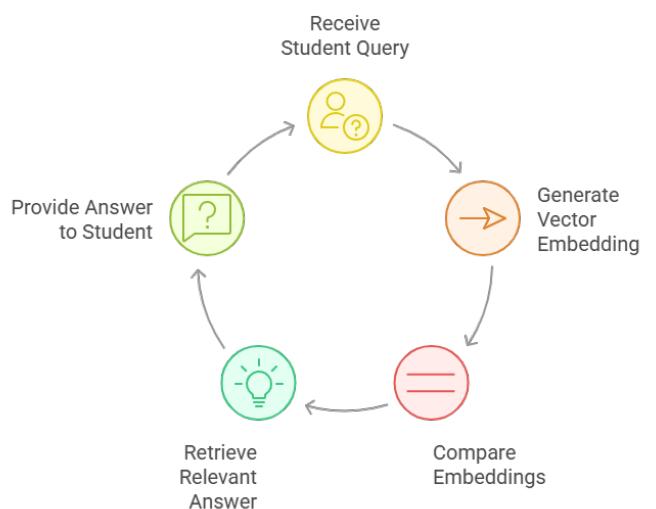


Fig. 5: Automated Doubt Resolution Cycle

H. Core Operational Pipeline

At its core, the system follows a pipeline that begins with the student entering a question in natural language. This query then passes through the preprocessing module, where it is

cleaned, tokenized, normalized, and transformed into a dense vector representation using the BERT encoder. In parallel, all questions and answers in the dataset are preprocessed and stored in an embedding index. Once the query embedding is obtained, the similarity matching module compares it with the indexed embeddings using similarity measures such as cosine similarity or approximate nearest neighbor search with FAISS (Facebook AI Similarity Search). The most relevant answer is then retrieved and presented to the student.

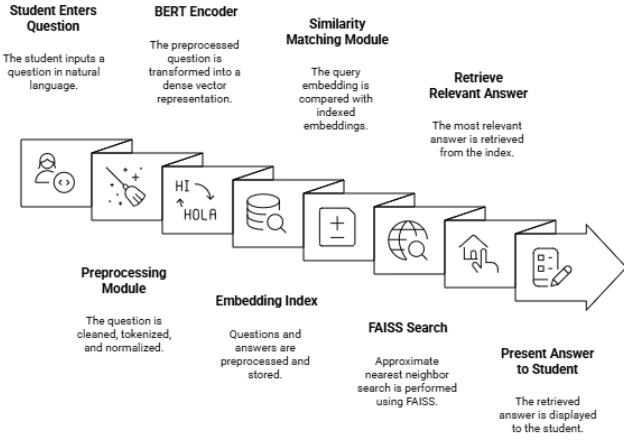


Fig. 6: Question Answering System Pipeline

I. Modular Components and Functions

The architecture is deliberately modular so that each stage performs a specialized function while contributing to the overall efficiency of the system. The Input Module is responsible for handling student queries in raw text format. The Preprocessing Module ensures the query is normalized into a standardized form. The Embedding Layer, powered by BERT, generates contextual representations that capture the deeper semantics of the query.

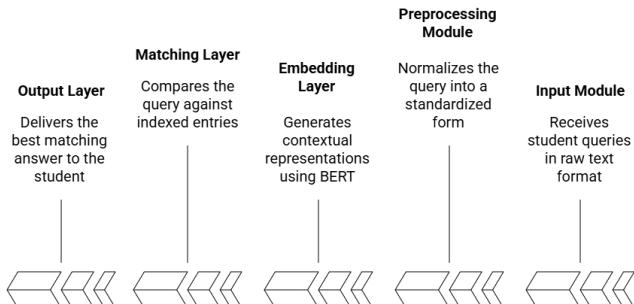


Fig. 7: Modular Components and Functions

The Matching Layer is where the retrieval logic takes place, comparing the query against thousands of indexed entries. The Output Layer delivers the best matching answer back to the

student. This modular approach ensures that the system can be extended or modified in the future, for instance, by replacing BERT with a newer language model or by expanding the dataset.

J. Semantic Similarity

A significant advantage of this architecture is its ability to handle semantic variation in student queries. Because the system uses BERT embeddings, queries that differ syntactically (e.g., “What is a token in C?” vs. “Explain the tokens used in C language.”) are mapped to nearby positions in the vector space, allowing the similarity matching module to retrieve the same answer.

To further improve efficiency, the system employs indexing techniques. The embeddings of all stored questions are indexed using FAISS, which allows sub-millisecond retrieval of the most similar entries. This ensures that the system can scale to handle much larger datasets.

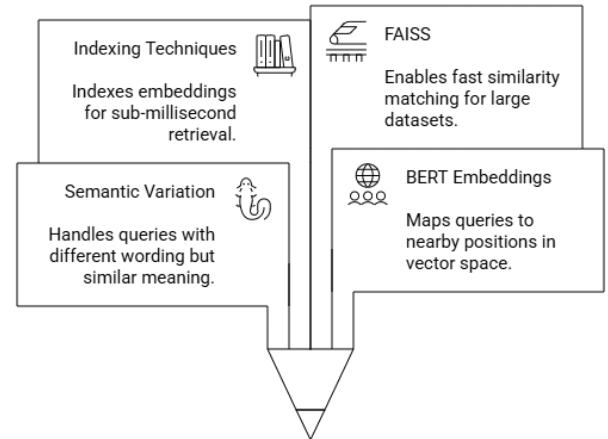


Fig. 8: Pathways to Rapid Information Access

K. Top-k Retrieval and Visualization

Another notable feature of the system architecture is the use of top-k retrieval. Instead of returning only a single answer, the system can retrieve the top k most relevant matches ranked by similarity score. This feature is particularly useful in cases where multiple answers are semantically close to the query or where the student query is ambiguous, increasing reliability and providing alternative explanations. The architecture presents a more detailed workflow, showing the interaction of each component at a granular level.

IV. CONCLUSION

The examined literature presents the progress in semantic search, natural language processing (NLP), adaptive learning, question answering (QA) systems, and sentiment analysis, which can potentially be the basis of an automated doubt resolution system. Research on semantic similarity, semantic search engines, and dialogue systems also demonstrates how, in order to provide more accurate and relevant answers,

both semantic and syntactic analysis should be treated as a married pursuit. Research pertaining to adaptive learning tools and virtual teaching assistants highlights the positive impact that personalization and awareness of context can provide to student engagement and learning. There has also been an advancement in QA over knowledge bases that has found the capacity for scalable and accurate retrieval of information. Also note that there has been sentiment analysis of user feedback, functioning to aid in the cyclical improvement of the system. Collectively, the published works suggest that an automated doubt resolution system may provide better relevance, accuracy, and usability if it combines diverse systems that can respond to the needs of language in NLP, contextualize understanding domain-led semantic knowledge, adaptive learning, and overall, in real-time dialogue administration.

REFERENCES

- [1] Scotti, V., Sbattella, L., & Tedesco, R. C. (2023, October 6). A primer on Seq2Seq models for generative chatbots. *ACM Computing Surveys*, 56(3), Article 75:1–75:58. <https://doi.org/10.1145/360428>
- [2] Létourneau, A., Deslandes Martineau, M., Charland, P., Karan, J. A., Boasen, J., & Léger, P. M. (2025, May 14). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*, 10, Article 29. <https://doi.org/10.1038/s41539-025-00320-7>
- [3] Lin, C.-C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments*, 10, Article 41. <https://doi.org/10.1186/s40561-023-00260-y>
- [4] Soares, T.G., Azhari, A., Rokhman, N., & Winarko, E. (2023). Survey categorizing paper on education question answering systems. In S.-I. Ao, O. Castillo, H. Katagiri, A. Chan, & M.A. Amouzegar (Eds.), *Transactions on Engineering Technologies: IMECS2021* (pp.77–91). Springer. https://doi.org/10.1007/978-981-19-7138-9_7
- [5] Ladanavar, S. M., Kamble, R., Goudar, R. H., Kaliwal, R.B., Rathod, V., Deshpande, S.L., Dhananjaya, G.M., & Kulkarni, A. (2024). Enhancing user query comprehension and contextual relevance with a semantic search engine using BERT and ElasticSearch. *EAI Endorsed Transactions on Internet of Things*, 10(1). <https://doi.org/10.4108/eetiot.6993>
- [6] Palasundram, K., Mohd Sharef, N., Kasmiran, K. A., & Azman, A. (2020). Enhancements to the sequence-to-sequence-based natural answer generation models. *IEEE Access*, 8, 45738–45752. <https://doi.org/10.1109/ACCESS.2020.2978551>
- [7] Gligoreia, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12), 1216. <https://doi.org/10.3390/educsci13121216>
- [8] Marouf, A., Al-Dahdooh, R., Abu Ghali, M. J., Mahdi, A. O., Abunasser, B. S., & Abu-Naser, S. S. (2024). Enhancing education with artificial intelligence: The role of intelligent tutoring systems. *International Journal of Engineering and Information Systems (IJEAIS)*, 8(8), 10–16. <https://doi.org/10.4108/eetiot.6993>
- [9] Kovács, L., Csépányi Fürjes, L., & Ahmed, G. H. A. (2022). Automated assessment generation in intelligent tutoring systems. In *Lecture Notes in Networks and Systems* (Vol. 383, pp. 808–819). Springer. https://doi.org/10.1007/978-3-030-93817-8_72
- [10] hanushali, A. (2023). Impact of automation on quality assurance testing: A comparative analysis of manual vs. automated QA processes. *International Journal of Advances in Scientific Research and Engineering*, 26*(IV), 39. <https://www.researchgate.net/publication/375342615>
- [11] De Castro, J. P., & Pérez, M. Á. (2008). An analysis of the research on adaptive learning: The next generation of e-learning. *WSEAS Transactions on Information Science and Applications*, 5(6), 866–876. <https://doi.org/10.5555/1467049.1467050>
- [12] Alkhatlan, A., & Kalita, J. (2019). Intelligent tutoring systems: A comprehensive historical survey with recent developments. *International Journal of Computer Applications*, 181(43), 1–20. <https://doi.org/10.5120/ijca2019918451>
- [13] Nazmi, R., Ardiyanto, J., Anshori, M. I., Siswanto, E., & Wirawan, R. (2023). Adaptive learning in the future of educational management adapts to student needs. *al-fikrah: Jurnal Manajemen Pendidikan*, 11(2), 272–283. <https://doi.org/10.31958/jaf.v1i2.10552>
- [14] Gomes, D. (2024). Intelligent tutoring system: A comprehensive study of advancements in intelligent tutoring systems through artificial intelligence education platforms. In *Improving Student Assessment With Emerging AI Tools* (pp. 213–244). IGI Global. <https://doi.org/10.4018/979-8-3693-6170-2.ch008>
- [15] Shaik, T., Dann, C., Tao, X., Li, Y., McDonald, J., Redmond, P., & Galligan, L. (2023). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *IEEE Access*, 10, 56720–56739. <https://doi.org/10.1109/ACCESS.2022.3185241>
- [16] Fu, B., Qiu, Y., Tang, C., Li, Y., Yu, H., & Sun, J. (2020). A survey on complex question answering over knowledge base: Recent advances and challenges. *arXiv*. <https://doi.org/10.48550/arXiv.2007.13069>
- [17] Fitria, T. N. (2021). Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. *Prosiding Seminar Nasional & Call for Paper STIE AAS*, 1, 134–139. Retrieved from <https://www.researchgate.net/publication/357447234>
- [18] Deshmukh, A. A., & Sethi, U. (2020, July 24). IR-BERT: Leveraging BERT for semantic search in background linking for news articles. *Association for Computing Machinery*. <https://doi.org/10.1145/1122445.1122456>
- [19] Garg, S., Vu, T., & Moschitti, A. (2020). Knowledge transfer from answer ranking to answer generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 326–337). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.30>
- [20] Alhwiti, K. M. (2014). Natural language processing and its use in education. *International Journal of Advanced Computer Science and Applications*, 5(12), 72–76. <https://doi.org/10.14569/IJACSA.2014.051211>
- [21] Ahmed, T., & Khan, M. Z. (2019). Natural language processing for enhancing teaching and learning. *International Journal of Advanced Computer Science and Applications*, 10(12), 80–85. <https://doi.org/10.14569/IJACSA.2019.0101212>
- [22] Dan, V., George, C., & Loverthtin. (2024, May). Natural Language Processing in education : Enhancing communication and understanding <https://www.researchgate.net/publication/380606077>
- [23] Medhat, A., & Ahmed, R. (2019). Natural language processing of student's feedback. *International Journal of Advanced Computer Science and Applications*, 10(5), 1–7. <https://doi.org/10.14569/IJACSA.2019.0100501>
- [24] Almeida, F., & Simoes, J. (2019). Personalized adaptive learning in higher education: A scoping review. *International Journal of Educational Technology in Higher Education*, 16(1), 1–20. <https://doi.org/10.1186/s41239-019-0171-8>
- [25] Unger, C., Freitas, A., & Cimiano, P. (2014). Question answering over linked data: What is difficult to answer? *Web Semantics: Science, Services and Agents on the World Wide Web*, 29, 82–98. <https://doi.org/10.1016/j.websem.2014.04.001>
- [26] Dwivedi, S. K., & Singh, V. (2013). Research and reviews in question answering system. *Procedia Technology*, 10, 417–424. <https://doi.org/10.1016/j.protcy.2013.12.378>
- [27] Zhang, Y. (2024). Research on the mechanism of intelligent education enabled by natural language processing technology. In P. Qi & Z. Chen (Eds.), *Proceedings of the 2023 International Conference on Big Data and Informatization Education (ICBDIE 2023)* (pp. 257–267). Atlantis Press. https://doi.org/10.2991/978-94-6463-238-5_5
- [28] Roy, A., & Ghosh, S. (2023, November). Semantic coupling of scientific literature using sbBERT: An enhanced model for systematic literature review. In *Digital Libraries: Sustainable Development in Education – Theme 4: Technology and Innovation in Digital Libraries* (pp. 728–739). [Conference paper]. <https://www.researchgate.net/publication/377308246>
- [29] Duhan, A., Singhal, A., Sharma, S., Neeraj, & Arti, M.K. (2024, December 9). Semantic search and recommendation algorithm: Advanced techniques for efficient real-time data retrieval and analysis – A comprehensive approach to modern data systems and methodologies [Preprint]. Netaji Subhas University of Technology. <https://doi.org/10.48550/arXiv.2412.06649>
- [30] Masuda, K., Matsuzaki, T., & Tsujii, J. (2011). Semantic search based on the online integration of NLP techniques.

- Procedia - Social and Behavioral Sciences, 27, 281–290. <https://doi.org/10.1016/j.sbspro.2011.10.609>
- [31] Pandiarajan, S., Yazhmozhi, V. M., & Praveen Kumar, P. (2015). Semantic search engine using natural language processing. In H. A. Sulaiman, M. A. Othman, Y. A. Rahim, & N. C. Pee (Eds.), Advanced computer and communication engineering technology (pp. 561–571). Springer. <https://doi.org/10.1007/978-3-319-07674-4-53>
- [32] Desai, A., & Ganatra, A. (2017). Semantic similarity: A review of approaches and methods. International Journal of Computer Applications, 161(11), 41–46. <https://doi.org/10.5120/ijca2017913230>
- [33] Gutiérrez, Y., Tomás, D., García-Serrano, A., & Gómez, J. M. (2016). Semantic similarity is not enough: A novel text similarity measure combining semantic and syntactic information. Information Processing & Management, 52(6), 1182–1197. <https://doi.org/10.1016/j.ipm.2016.05.009>
- [34] Hussain, M., & Siddique, M. A. (2023). Sentiment analysis of student feedback with NLP and deep learning. International Journal of Advanced Computer Science and Applications, 14(6), 522–528. <https://doi.org/10.14569/IJACSA.2023.0140662>
- [35] Wu, Y., & Yang, Z. (2020). SEQ2SEQ++: A multitasking-based seq2seq model with multiple pretrained models for text summarization. Journal of Physics: Conference Series, 1646, 012014. <https://doi.org/10.1088/1742-6596/1646/1/012014>
- [36] Nayel, H., & Elmaghriby, A. (2020). Short-text semantic similarity (STSS) techniques, challenges, and future research directions. IEEE Transactions on Knowledge and Data Engineering, 34(3), 1086–1101. <https://doi.org/10.1109/TKDE.2020.2964586>
- [37] Mäkelä, E. (2008, July). Survey of semantic search research. In Proceedings of the 1st International Conference on Web Intelligence, Mining and Semantics (WIMS). <https://www.researchgate.net/publication/225070162>
- [38] Diefenbach, D., Both, A., Singh, K., & Maret, P. (2018). A systematic review of question answering over knowledge bases. Semantic Web, 11(2), 241–277. <https://doi.org/10.3233/SW-180301>
- [39] Aleksander, I. (1996). Taking adaptive learning in educational settings: A neural network perspective. Computers & Education, 26(1–3), 1–7. [https://doi.org/10.1016/0360-1315\(96\)00003-3](https://doi.org/10.1016/0360-1315(96)00003-3)
- [40] Fang, H., Guo, F., Wang, L., Yang, Z., Li, X., Duan, N., & Zhou, M. (2021). A survey of task-oriented dialogue systems. Science China Technological Sciences, 64(9), 2011–2035. <https://doi.org/10.1007/s11431-020-1744-2>
- [41] Shameem, M., & Kumar, R. (2021). The role of QA automation in eliminating waste in project teams. Journal of Physics: Conference Series, 1840(1), 012029. <https://doi.org/10.1088/1742-6596/1840/1/012029>
- [42] Kerr, D., Rynearson, K., & Kerr, A. (2006). Using an adaptive learning tool to improve calculus students' performance. Mathematics and Computer Education, 40(3), 213–226. <https://doi.org/10.1186/s40561-024-00292-y>
- [43] Fryer, L. K., & Bovee, H. N. (2021). Virtual teaching assistants: A systematic review. Computers and Education: Artificial Intelligence, 2, 100020. <https://doi.org/10.1016/j.caai.2021.100020>