ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Knowledge Management (Km) Information System for Laguna State Polytechnic University - Santa Cruz Campus

Badillo, Jerahmeel A , Calapao, Jan Reinnen S , Rebong, Dexter D , Villarica, Mia V

College of Computer Studies Laguna State Polytechnic University Laguna, Philippines

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150500067

Received: 01 May 2026; Accepted: 07 May 2026; Published: 01 June 2026

ABSTRACT

Documents in any institution have been outnumbered and scattered in various repositories, hindering efficient information retrieval and threatening institutional memory. This study developed a web-based Knowledge Management Information System for Laguna State Polytechnic University - Sta. Cruz Campus that consolidates core institutional documents such as research outputs, policies, course materials and extension documents into a centralized, secure, and searchable platform. Employing a Developmental and Experimental Research Design, KMIS engineered four core intelligent components: Role-Based Access Control centralized repository, Retrieval Augmented Generation pipeline that uses vector embeddings to utilize semantic search and LLM-generated summaries from natural language queries. Lastly, the automated QPRO analysis engine. The developed KRA text classification model achieved an overall accuracy of 98% and automatically categorized institutional reports against the 22 Key Result Areas from the 2025-2029 LSPU Strategic Plan.

Keywords: Knowledge Management, Centralized Repository, Role-based Access Control, Semantic Search, Retrieval Augmented Generation, LLM, Transformers

INTRODUCTION

Knowledge can be considered as one of the most important assets of universities because it is used as a basis for teaching-learning activities, research and administration. According to the World Bank, Higher Education Institutions are able to manage information through digital systems which allows them to organize their services effectively and improve decision-making. Knowledge Management Information Systems at Laguna State Polytechnic University – Sta. Cruz Main Campus can help in gathering files and practices scattered in different areas to a secured platform with searchable functions that can ease every day transactions. Knowledge at LSPU – Sta. Cruz Main Campus is fragmented. It is located per office, each individual's google drive and some are kept in different systems. This makes it difficult to locate, reuse and manage vital documents.

Currently knowledge at the university is distributed among offices, personal google drives and other platforms making it difficult to identify, reuse and manage key papers. This results in duplicate effort, out-of-date references, and loss of institutional memory with human or institutional employee changes. One approach to address difficulties is a campus-wide knowledge management information system that provides a central repository for institutional files and reports such as research outputs, syllabi, policies, and extension reports with explicit access controls and standard forms.

Artificial intelligence can improve the efficiency of knowledge management by offering more precise search results and useful tags for content (Taherdoost, 2023). Intelligent search allows university users to ask natural questions and receive relevant results, while automated tagging helps to provide consistent metadata without burdening contributors. In practice, intelligent search allows users in universities to ask queries in natural language and get relevant answers, while automated tagging ensures consistent metadata without burdening contributors. These two aspects are realistic and valuable to be focused in the first place by LSPU – Sta. Cruz Main Campus and not a commitment for immediate enhancements, but for future enhancements with analytics and chatbots.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

The ideal knowledge management information system for the campus should be accessible and secure and role mapped to the different roles like administrators, faculty, personnel and students. It should have simple upload work-flows, standard tags and reliable retrieval so knowledge flows easily between people and units. Such infrastructure allows universities to retain institutional memory, speed up academic and administrative processes and raise the profile of research and community outcomes. In the World Bank's view, the digital transformation in Philippine higher education is about building shared platforms and data-driven processes that will improve teaching, research and governance. These directions are good background for a KMIS which enhances knowledge consolidation, discoverability and responsible access at LSPU – Sta. Cruz Main Campus. Further details on the performance metrics are given in the following sections. What we need now is a good working central system for the regular chores. This project is about the design and implementation of a KMIS for Laguna State Polytechnic University – Sta.Cruz Main Campus. The initial scope was a central repository, role-based access, intelligent search and auto tagging.

Research Objectives

The purpose of this study is to create and develop a web-based Knowledge Management Information System (KMIS) for Laguna State Polytechnic University – Sta. Cruz Main Campus that creates institutional knowledge into a single, secure, searchable platform with role-based access and intelligent search to improve everyday tasks such as teaching, research and administration. Specifically, this study aimed to do the following:

1. To design and implement a centralized repository with role-based access controls for core assets or institutional documents such as research outputs, course materials, campus policies, and extension documents, ensuring secure storage and consistent organization across units.

2. To implement a semantic search functionality capable of processing natural language queries to optimize information discovery and improve search precision while streamlining data retrieval workflows.

3. To develop an automated quarterly progress report on objectives analysis engine that intelligently evaluates strategic plan progress through AI-powered document analysis, providing real-time insights into key result areas and key performance indicator achievement across organizational units.

4. To design and train a text classification model for the automated categorization of institutional accomplishments into specific key result areas (KRAs) and to evaluate the performance of the trained model using standard metrics such as accuracy, precision, recall, and F1-score.

RELATED LITERATURE

The accumulated studies have been analyzed by the researchers and have served as a guide to concepts, methods, strategies, and techniques in order to achieve the desired goals for the implementation of a centralized Knowledge Management Information System.

Knowledge management aims to develop a structured process for creating, capturing, and sharing institutional information or assets. Alvarenga et al. (2020) defined digital transformation as an essential shift that changes how knowledge is produced and stored in higher education which makes it more efficient and relevant. In fact, Putri et al. (2023) emphasized that without a centralized system, institutional knowledge remains a personal advantage of individuals rather than a shared asset of the university, leading to documentation lacks and duplication of effort.

In relation to the technical implementation, Che et al. (2024) highlighted the use of Retrieval-Augmented Generation (RAG) and hierarchical context augmentation as powerful methods for handling huge number of scientific papers. This approach allows the system to capture institutional knowledge such as the relationship between abstracts which is more effective than traditional keyword-based searches. Similarly, Tinh and Phuong (2025) highlighted heading-aware chunking as a way to enhance context retrieval by ensuring that the system understands the contained headings and directory hierarchies of university files.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Moreover, the application of Large Language Models has revolutionized document classification and information retrieval. Xu (2024) shows that models like LLaMA2 outperform typical machine learning classifiers in terms of precision and recall for academic document classification. Wang and Pei (2024) researched LLM-based query expansion for increasing the accuracy of user inquiries so that the system could give accurate answers even if the query terms of the user are incomplete.

Finally, the design of the system interface is important to its acceptance and success. According to Kurniawan et al. (2024), the platform addresses the real business needs of a higher education institution using the User-Centered Design (UCD) and evaluation by System Usability Scale (SUS). The proposed KMIS of the LSPU – Santa Cruz Campus integrates these sophisticated AI-enabled retrieval techniques in a user-centered design paradigm to establish a repository of institutional knowledge that is robust, effective and highly accessible.

Digital transformation claims adaptive knowledge management in higher education institutions (HEIs) to sustain relevance among fast technological and informational shifts. Alvarenga et al. (2020) emphasized that universities must properly manage knowledge resources to promote intellectual alignment with institutional visions and strategies. Mehta (2021) described knowledge management as essential for collecting, analyzing, and sharing data to generate insights, strategies, and innovations among students and personnel where knowledge management handles organizational output through best practices and efficient processes. Galgotia & Lakshmi (2022) highlighted its role in enabling quick responses to challenges, better decision-making, skill enhancement, reduced redundancies, cost savings, and error minimization via targeted policies and faster information access. Putri et al. (2023) identified issues like poor documentation, uncoordinated updates, trust on informal verbal sharing, duplicated efforts in operations, not using of online guides, and human errors producing misleading data.

Academic repositories are improving knowledge access and storage in higher education by using structured metadata and open data principles. Mosha & Ngulube (2023) literature study 2003-2023 through Scopus, Web of Science, Google Scholar, Boolean searches. Metadata standards highlight repositories in research data sharing since open data paradigm. Dube (2025) support the whole research data lifecycle, in accordance with the FAIR principles of discoverability, reuse and collaboration. Faculty and institutional repositories gather different resources in one place, making them more efficient. Faculty repositories, as observed by Zibani, Rajkoomar, & Naicker (2021), can be used to curate teaching, learning, and research information and thereby save search time compared to using the web to locate scattered web resources. Ferreras et al. (2013) pointed out the importance of openness, rich metadata and interoperability in institutional repositories for information access.

Effective KM requires knowledge repositories with technical features for sharing and discovery. Fadhlan & Sensuse (2022) recommend the use of ontological tagging for organizational knowledge bases that may be used by LSPU students and faculty because data catalogs, strong search, APIs and visualization are needed.

RESEARCH METHODOLOGY

Research Design

The suggested Knowledge Management Information System has been carefully designed, developed and deployed utilizing the developmental research design. The approach was to evaluate the existing information management processes at LSPU, identify system needs, design the system architecture, and design core features such as centralized knowledge repository, AI-based search, automatic tagging, and recommendation tools. Developmental design included system development, integration of machine learning components, and continuous improvement based on usability and usefulness. This approach is ideal (Clark-Plaskie et al., 2022) as the primary purpose of the project is to create a usable institution-specific KM Information System that promotes knowledge storage, retrieval, and sharing in the university. In addition, an experimental study design was used to evaluate the performance of the constructed system. The second part of the study is controlled testing. In this part the accuracy of the semantic search function, the relevance of the document insights and suggestions, and the usability and reliability of the system are assessed. The team examined the system outputs and the user interactions to assess if the new KMIS features actually enhance information accessibility and user

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

experience in the academic setting of the LSPU.

According to Clark-Plaskie et al. (2022), developmental research designs are effective for studies focused on system creation and improvement because they allow researchers to observe changes and enhancements throughout the development process. Meanwhile, Zubair (2022) explained that experimental research includes controlled testing of variables to determine their effects on results. By relating the concepts above, combining both developmental and experimental research designs ensured that the proposed KMIS was not only should be systematically evaluated in terms of functionality, performance, and user satisfaction.

Applied Concepts and Techniques

The study advanced on the principles and methods that were used in developing the knowledge management information system so as to solve the challenges connected with disorganized information storage and retrieval, the following concepts and methodologies were used:

Knowledge Management (KM)

Knowledge management is the basic concept used in this study. Knowledge management is a systematic approach to obtaining, storing, organizing, sharing and retrieving knowledge inside an institution. It is composed of academic materials, research findings, administrative documents, and institutional policies in the setting of Laguna State Polytechnic University. The KMIS is designed to serve core KM tasks, including knowledge creation, storage, sharing and reuse, through a centralized digital repository.

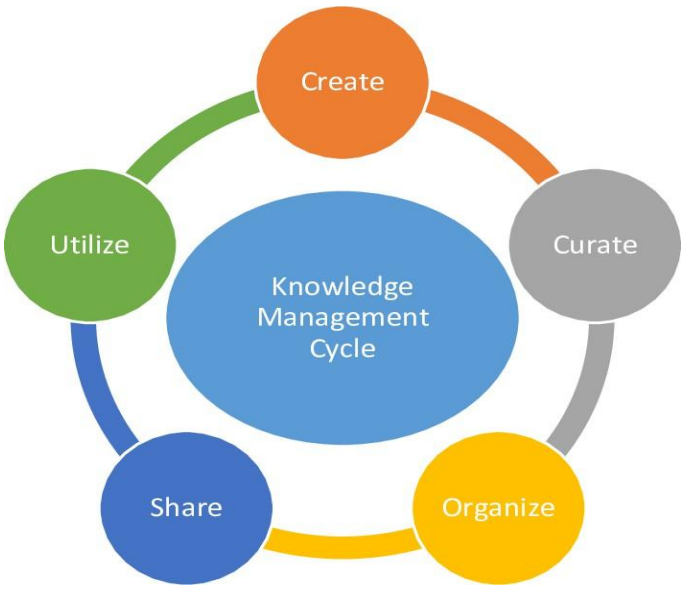


Figure 4. Knowledge Management Cycle (Manjunatha, 2023)

Figure 4 depicts the cyclic process of the framework for institutional knowledge handling in Laguna State Polytechnic University The diagram shows four interconnected phases which are capture, store, organize and share which are placed in a continuous loop to emphasize the iterative nature of information management. They draw the dynamic flow with arrows from stage to stage. Knowledge is obtained from institutional documents, such as strategic plans and operational reports, stored in centralized repositories, structured to enable easy access and retrieval, and shared with stakeholders to guide decision making and align performance.

Machine Learning

Machine learning is a way for computer systems to find patterns from data and make predictions without being explicitly programmed. This study applies machine learning to categorize institutional documents into the 22 Key Result Areas (KRAs) in LSPU's Strategic Plan 2025–2029, converting unstructured strategic documents and QPRO reports into structured, actionable insights for performance monitoring and alignment.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

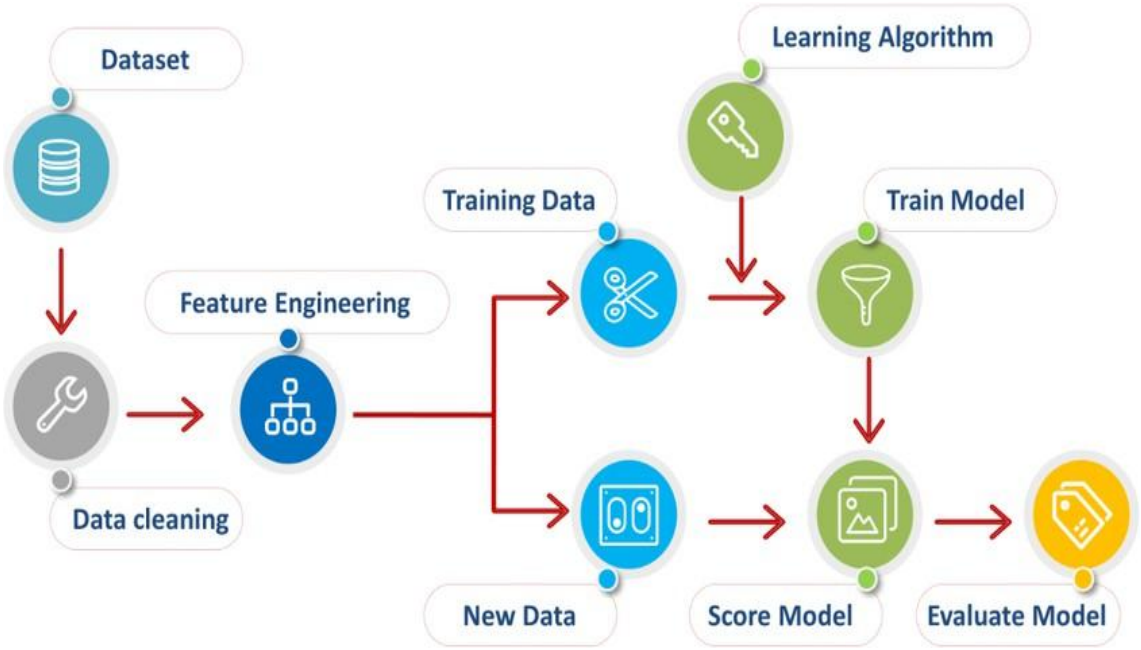


Figure 5. Machine learning process diagram (Al-Saad et al. 2023)

Figure 5 shows the operation flow of ML algorithms used in this study. The procedure starts with a dataset of institutional papers such as the QPRO reports, the filled PDO forms and the LSPU Strategic Plan. That leads into data cleansing to eliminate inconsistencies, then feature engineering to provide structured inputs.

Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) is a technique to augment large language models with real-time retrieval from an external knowledge source. It generates contextually correct responses. It uses live documents for the domain. This study uses RAG to enable the institutional knowledge repository to provide semantic search for all kinds of submitted documents such as PDF, Word files, spreadsheets and reports regardless of their format or structure.

Once documents are submitted to the centralized repository, documents will be processed automatically including text extraction, generation of semantic embeddings with transformer models, and indexing in a vector database for efficient similarity searching. When the user poses a query, the system will perform a retrieval step where chunks of documents that are most semantically similar will be identified and ranked based on the cosine similarity of the embeddings. The retrieved segments are then used as context for the generative model to synthesize accurate answers, citing the location of the sources.

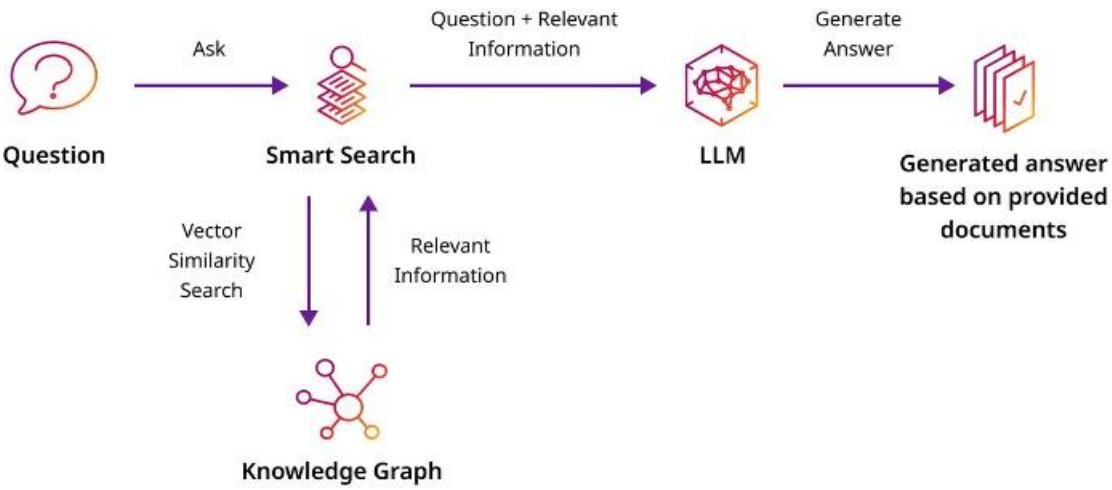


Figure 6. Retrieval Augmented Generation Diagram

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Figure 6 depicted the design of the Retrieval Augmented Generation pipeline implemented in the institutional knowledge management information system. Smart Search occurs on user query with vector similarity matching on embeddings in knowledge graph to retrieve relevant information from uploaded files. This context is loaded with the query and sent to a Large Language Model which provides answer output based on the retrieved provided documents.

Tokenization

Tokenization is the first step in preprocessing in natural language processing. It breaks down unstructured texts of institutional documents into separate pieces or tokens that a machine can interpret. In this work, we used the distilbert-base-uncased tokenizer in Hugging Face Transformers to tokenize the extracted text from the LSPU Strategic Plan 2025–2029 and QPRO reports into subword tokens that preserve the semantic sense of the text but enable out-of-vocabulary phrases to be processed by WordPiece algorithms.

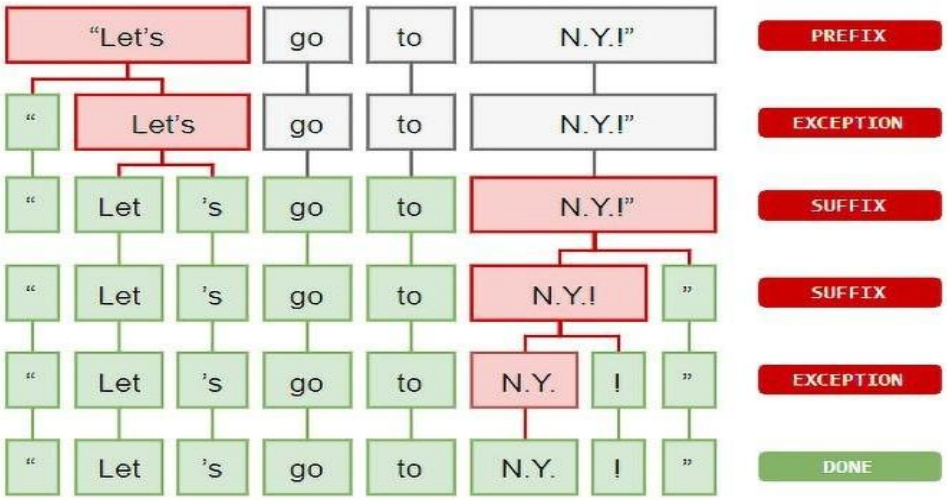


Figure 7. Tokenization Method (Hashem et al., 2021)

Figure 7 provides an example of the tokenization approach that divides a phrase, a sentence, a paragraph or a full text document into smaller units such as single words or phrases. Each little component is called a token (Hashem et al., 2021).

Large Language Model

Large Language Models (LLMs) are transformer-based neural networks trained on enormous text structure, offering improved capabilities in natural language processing, creation and reasoning. In this work we show how LLMs can be used to improve semantic search, by improving the user query, generating semantically correct answers from retrieved chunks of documents and generating prescriptive analysis from QPRO reports. It can identify the Key Result Areas (KRA's) and performance gaps between the actual operational data and the strategic targets and can generate actionable recommendations e.g. re-allocation of resources or priority interventions for planning officers. In other words, it can translate raw institutional documents into strategic decision support.

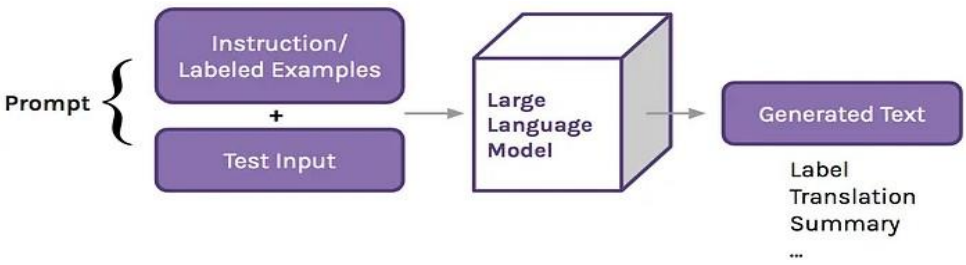


Figure 8. How Large Language Model Work

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

The main inference method of the LLM in the semantic search engine is illustrated in Figure 8. The instruction prompt with labeled examples, and test input are created and passed to the Large Language Model. This structured input is then passed to the model to generate text. The resulting text is then translated into labels to build the final outputs such as KRA classifications or prescriptive insights.

Algorithm Analysis

The proponents used techniques for the analysis of the algorithms that may help the researchers to determine the most suitable for the development of AI based system for institutional document classification and Key Result Area (KRA) mapping in the Knowledge Management Information System (KMIS) application. To mitigate this, researchers explored a number of transformer models for natural language processing (NLP) cited in the literature, particularly those related to text classification and knowledge organization. The researchers hope that by testing and comparing various models, they would be able to determine which algorithm offers the best accuracy, speed and reliability for sorting university assets into their respective KRAs.

DistilBERT

DistilBERT is a distilled version of BERT base which is a compact, fast, inexpensive and light Transformer type. It is designed to be 40% smaller than a BERT model, while keeping the same language understanding and running 60% faster. DistilBERT employs a method known as knowledge distillation in which a smaller model is trained to replicate the behaviors of a larger model. It makes it viable for settings where the computational capacity is limited but the high-quality text classification is still needed.

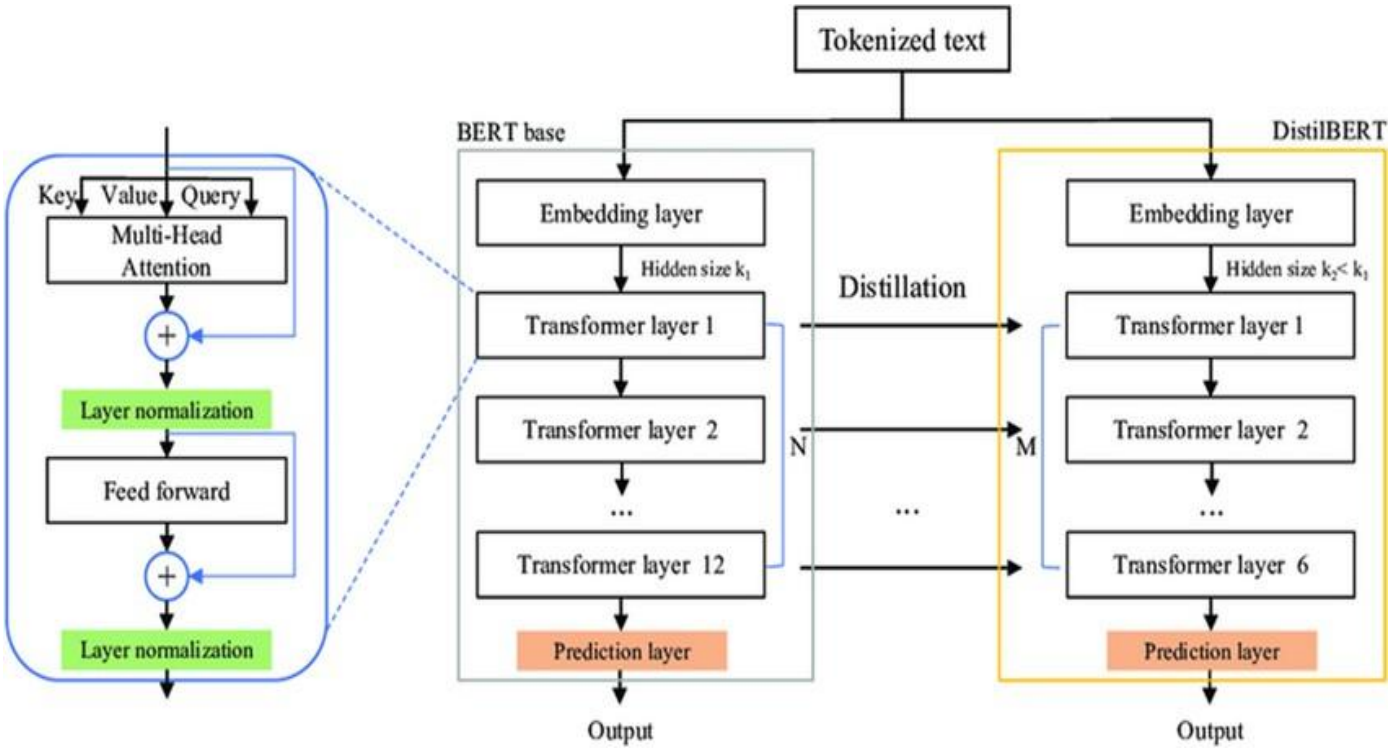


Figure 9. DistilBERT Architecture

The architecture of the DistilBERT used for KRA text classification is presented in Figure 9. The model takes as input tokenized text from QPRO reports and segments of the strategic plan. BERT Embedding Layer generates contextual embedding. These are then fed through 6 Transformer Layers (diluted from the original 12 of BERT) each consisting of multi-head attention mechanisms to capture semantic relationships as well as feed-forward networks to apply non-linear transformations. Layer Normalization which stabilizes training. Multi-Head Attention, which allows you to attend to different parts of the text at the same time. The final output layer generates KRA probability distributions to facilitate appropriate transformation from operational data to strategic objectives.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

DeBERTa-V3

DeBERTa-V3 (Decoding-enhanced BERT with Disentangled Attention Version 3) is a major improvement of the original Transformer models, combining the pre-training methodology of ELECTRA with a disentangled attention mechanism. While the common BERT models merge content and position information into a single vector, DeBERTa disjoins the two, allowing the model to investigate how the meaning of administrative phrases changes according to their relative distance to one another. Specifically, V3 uses Gradient-Disentangled Embedding Sharing (GDES) to improve the training efficiency and learn more discriminative features for each category.

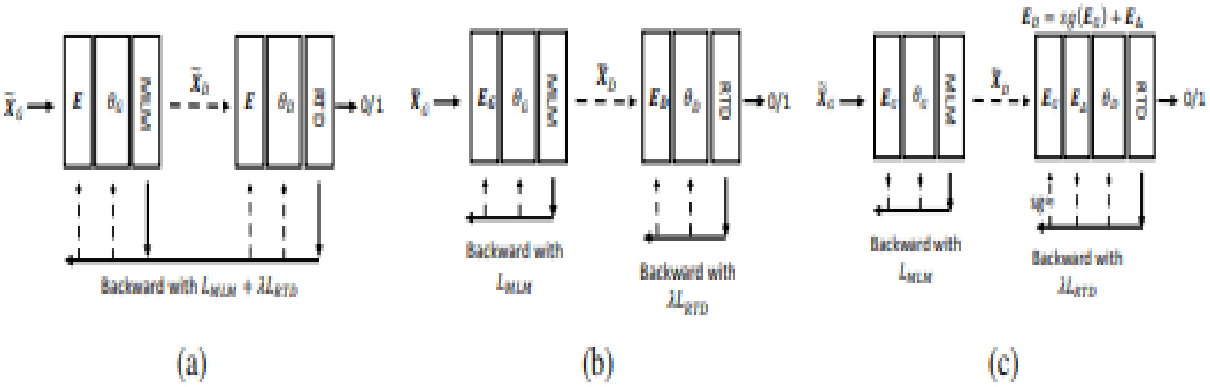


Figure 10. DeBERTa-V3 Disentangled Attention and RTD Mechanism

The main architecture, as shown in Figure 10, employs a Replaced Token Detection (RTD) objective rather than standard masking. Here, a generator model replaces some words with plausible fakes, and the DeBERTa discriminator has to detect the true words. This sensitizes the model to indirect changes in organizational language. In addition, the disentangled attention mechanism computes four different interaction scores, namely content-to-content, content-to-position, position-to-content and position-to-position. By separately considering these interactions, the KMIS AI is able to better discriminate between a “Research Report” (KRA 5) and a “Financial Report for Research” (KRA 22) and reach the level of accuracy and precision required for institutional data management.

MiniLM

MiniLM (Deep Self-Attention Distillation) is a highly efficient lightweight pre-trained model that focuses on distilling the "self-attention" distributions of deep Transformer networks. Unlike other models that distill the output of layers, MiniLM focuses on the relations between tokens. This allows the model to capture the complex structure of a sentence even with a much smaller number of parameters, making it one of the best choices for high-speed automated tagging and real-time search functionality in information systems (Wang et al., 2020).

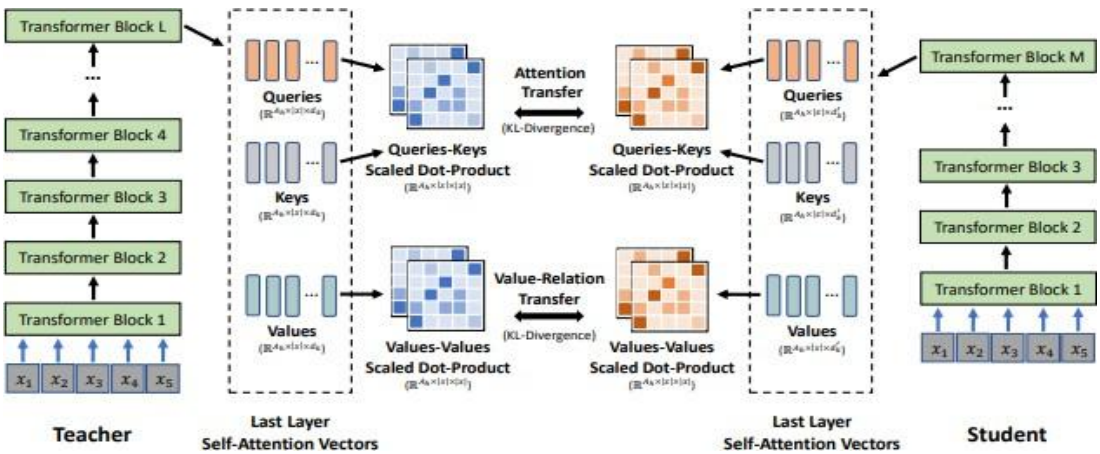


Figure 11. MiniLM Self-Attention Distillation

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Figure 11 shows how MiniLM processes text by distilling the "Self-Attention" maps from the teacher model. The researchers apply a deep self-attention distillation strategy, training the student model to mimic the attention values and the “Value-Relation” between tokens. This guarantees that the model preserves the most important linguistic features needed for the classification of the 22 KRAs of the university. The main merit of this model is its capability of working on a high level and being substantially smaller than the standard models. This helps the KMIS to stay responsive with the increase of the university's document database.

Data Collection Methods

Data collection methods are the process of collecting and evaluating information or data from multiple sources to find answers to research problems, answer questions, evaluate outcomes, and forecast trends and probabilities (Pulkit, 2025). In addition, data collection is the methodological process of gathering information about a specific subject. It is crucial to ensure the data is complete during the collection phase and that it is collected legally and ethically (Cote, 2021).

The researchers’ study entitled “Knowledge Management Information System (KMIS) for Laguna State Polytechnic University- Santa Cruz Campus” is a collection of institutional papers and knowledge assets that were exclusively obtained from the university departments. Such resources as accomplishment reports, research outputs, administrative files were gathered by systematic collecting throughout LSPU units such as CSS, COE, OSAS, HRMU etc. The Researchers perform formal data gathering by personally visiting various offices of LSPU, the College of Computer Studies and administrative facilities to ask for assistance and to validate the relevance of the dataset for KMIS development. They also worked with faculty specialists and department heads internally through official channels to provide access to complete records needed for training, testing and implementation of AI models for automated tagging, search and knowledge sharing.

Data Model Generation

In this study, the researchers include data collection, pre-processing, splitting and balancing, testing and validation, data evaluation, model training and model evaluation, model selection, and integration. Hence, it replicates within the scope of the data information that was being generated.

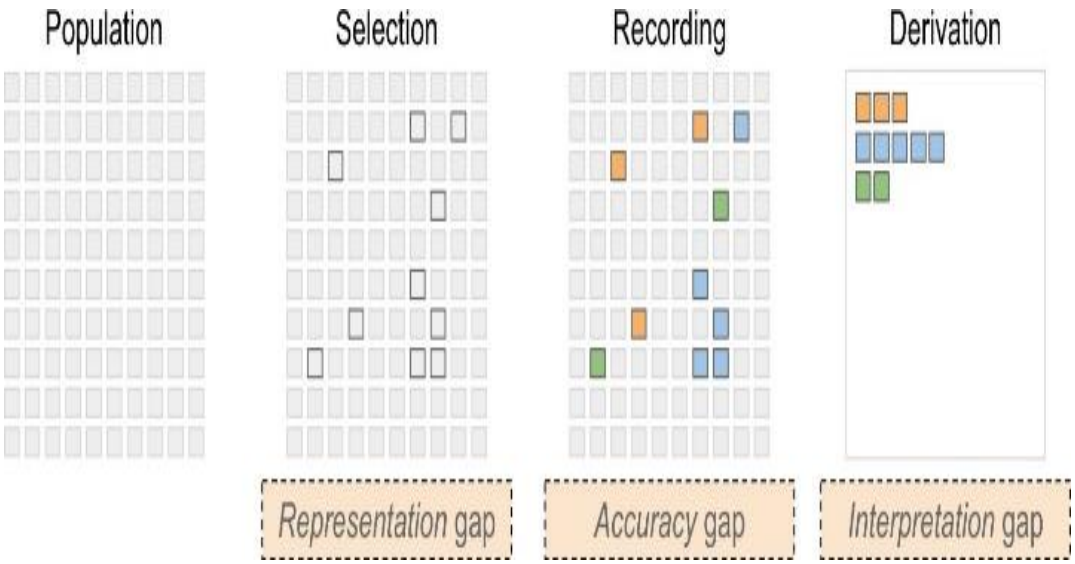


Figure 12. Data generation process

Figure 12 states that the data generation process involves four key stages, population, selection, recording, and derivation. Each of which can introduce specific types of gaps or biases. The representation gap occurs during the selection phase when only a subset of the population is chosen, which may not accurately reflect the whole. The accuracy gap happens in the recording phase due inconsistencies in the captured data. The interpretation gap happens during derivation where incorrect conclusions may be drawn from incomplete or biased data.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Data Preparation

Data preparation is important to ensure that the institutional documents obtained are ready for analysis and matched with the study's objectives. The data preparation in this research was concentrated in the refining of the Quarterly Physical Report of Operation (QPRO), and the LSPU Strategic Plan 2025-2029 to generate accurate and reliable information for evaluation according to the needs of the study.

QPRO documents of each campus unit were collected, classified and formatted in a similar manner and then labelled as the test dataset. This is to see if the operational performance of the university meets the requirements and performance targets set in the strategic plan. The reports were checked for completeness, consistency of the reporting period and comparability of the indicators in order to leave only certified entries for further research.

The LSPU Strategic Plan 2025-2029 is subject to data cleaning to eliminate narrative portions and other items not directly related to the assessment of institutional performance. Critical components, i.e. Key Result Areas (KRAs) and their related Key Performance Indicators (KPIs) were the only components which were kept and encoded in a structured fashion to be the major reference framework. These KRAs and KPIs were further linked to the respective indicators of the QPRO dataset. This allowed for a logical and objective comparison between the actual operational data and the strategic performance requirements of the university.

Data Pre-processin*g*

Data preprocessing in this study was performed on the LSPU Strategic Plan 2025–2029, which was provided as a PDF document. The goal of this stage was to convert the unstructured textual content into a machine-readable format suitable for input to the trained chosen model through a series of extraction, cleaning, balancing, and tokenization procedures.

For data extraction and labeling, the text was first broken down from the PDF using a Python-based PDF processing library. To partner each text segment with its corresponding Key Result Area (KRA), a rule-based labeling strategy was applied. Specifically, the document was read consecutively and whenever a KRA marker was encountered, all following lines were defined with that KRA label until the next KRA marker appeared, resulting to a labeled corpus organized by KRA.

The text that was extracted was then cleaned to remove noise from the PDF structure and formatting. Headers, footers, labels and administrative phrases that did not contain any useful information were removed via stop-phrases and regular-expression matching. Bullet points, numbering stylization, and other special characters were removed. Extremely small chunks of text were removed. Identical duplicate rows were dropped to prevent data leakage in model training.

An inspection of the labeled corpus showed remarkable imbalance across the 22 key result areas (KRA) with some areas being excessively relative to others. To reduce this, random oversampling was used which the KRA with the highest sample count was taken as the reference and instances from minority KRAs were repeatedly sampled with replacement until all KRAs attained an equal number of examples. This procedure yielded a fully balanced dataset with uniform representation for each KRA class.

Data Splitting and Balancing

Initial observations revealed class imbalance across the 22 KRAs. Random Oversampling accomplished this by identifying the KRA with maximum samples and up-sampling minority classes with replacement until all achieved equal representation. The balanced dataset was then structured into training (80%) and testing (20%) subsets to maintain proportional KRA distribution across splits, ensuring robust model evaluation.

Data Testing and Validation

The test dataset, which is 20 % of the balanced and tokenized corpus, was set aside for the final model evaluation to achieve an unbiased performance evaluation. To address the multi-class (22 KRAs) balancing, the DistilBERT

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

model was evaluated with accuracy, precision, recall, F1-score, and macro-averaged scores on the lasted subsets from the training split. We then tuned the hyper-parameters and detected over-fitting on the training set via cross-validation with early stopping.

Confusion matrix was built to perform detailed per-class performance evaluation and identify misclassifications among related KRAs by comparing the model predictions on the test set with the actual KRA labels. Results were evaluated for statistical significance using bootstrap resampling of test predictions, demonstrating resilience across the data set. All validation methods are standard procedures for NLP categorization, so we can be confident of the insights gained with respect to the degree to which the model aligns extracted text segments to strategic plan objectives.

Data Evaluation

After testing and validation, the data is assessed to determine the accuracy and reliability of the constructed dataset. The evaluation process is conducted using analytical techniques to assess the dataset's capability to support decision making for predictive purposes, missing value analysis, outlier detection and data consistency with the original LSPU Strategic Plan and QPRO documents. The assessment uses analytical tools that consider the ability of the dataset to predict decision making, including missing values checks, outlier detection and consistency check against the original LSPU Strategic Plan. Descriptive statistics, distribution plots and cross-referencing techniques were used to assess the data quality to generate reliable insights avoiding confusing patterns The data format and sources were reviewed to confirm the quality and consistency of the information used in the analysis.

Model Integration

The researchers did not directly integrate the model to the system, but first took steps to verify the reliability of the model. They wanted to see if the model was good enough. Model selection stage: it was tested thoroughly for performance before it was used as a model. The measures were the accuracy, precision and overall effectiveness of the selected algorithms. And if there were any problems then they would be fixed here so the model would work in the application and give results.

RESULTS AND DISCUSSION

This chapter presents the comprehensive results and detailed discussion relating to the development and evaluation of the Knowledge Management Information System (KMIS) for Laguna State Polytechnic University – Santa Cruz Campus.

Research Objective 1: To design and implement a centralized repository with role-based access controls for core assets or institutional documents such as research outputs, course materials, campus policies, and extension documents, ensuring secure storage and consistent organization across units.

Researchers have developed a centralized institutional document repository providing a single document model to store institutional core assets of documents with standard metadata and versioning, and role-based access controls. The repository is organized at all institutional units with a dedicated unit entity and explicit unit-document linkage. This allows structured grouping and retrieval while having a single source of truth. Data are stored in a secure environment with controlled access via role-based access control with defined institutional roles.

In addition to global RBAC, the system also supports unit scope (READ/WRITE/ADMIN) and document level permissions, thus providing fine-grained governance over sensitive institutional files. The repository supports accountability and auditability by persistently logging document views and downloads to reinforce data governance requirements anticipated in an institutional knowledge management environment.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

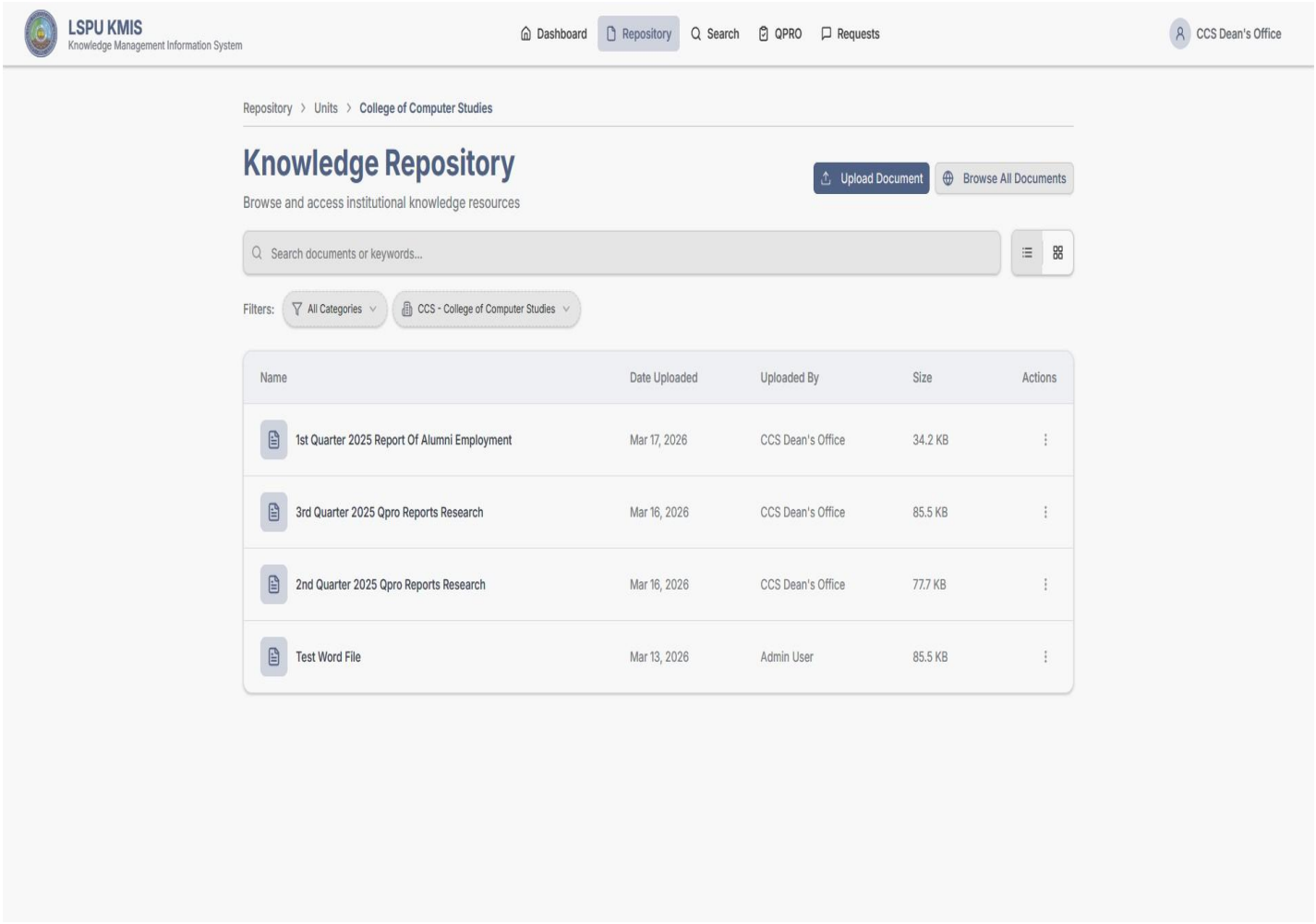


Figure 17. Screenshot of the system's repository UI

The visual implementation of the KMIS document repository is shown in Figure 17. Respond to Research Question 1: KMIS document repository as common and secure platform for institutional documents. UI is well balanced with a database-backed structure with institutional assets such as completed PDO forms and their associated metadata and versioning. The repository supports Role Based Access Control (RBAC) for assignment of permissions at the unit scope (READ/WRITE/ADMIN) and permissions at the document level. The system will generate a standard file list with standard metadata elements like document type, unit of origin and version number, the researchers said. The database allows version control and traceability assuring the user that he/she is looking at research outputs, course materials or administrative policies. This is important standardization to make auditability and compliance possible. It is consistent information and UI is clear with upload workflows where required metadata can be added.

Research Objective 2: To implement a semantic search functionality capable of processing natural language queries to optimize information discovery and improve search precision while streamlining data retrieval workflows.

The researchers developed a semantic search that can take natural language queries. Researchers built a potent Retrieval-Augmented Generation (RAG) pipeline tailored for extracting institutional knowledge. This pipeline allows end-users to submit natural-language queries through a chat-style endpoint that uses LLM augmentation to convert retrieved materials into accurate, human-readable summaries. The system adopts semantic processing for high precision in search. Chunking is applied to documents, and these are converted to vector embeddings. Retrieval is also optimized for speed with a managed vector index for efficient nearest neighbor queries. This system supports efficient workflows such as ad-hoc searching of attached files by a temporary session chat search path and continuous institutional search by persistent document embeddings.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

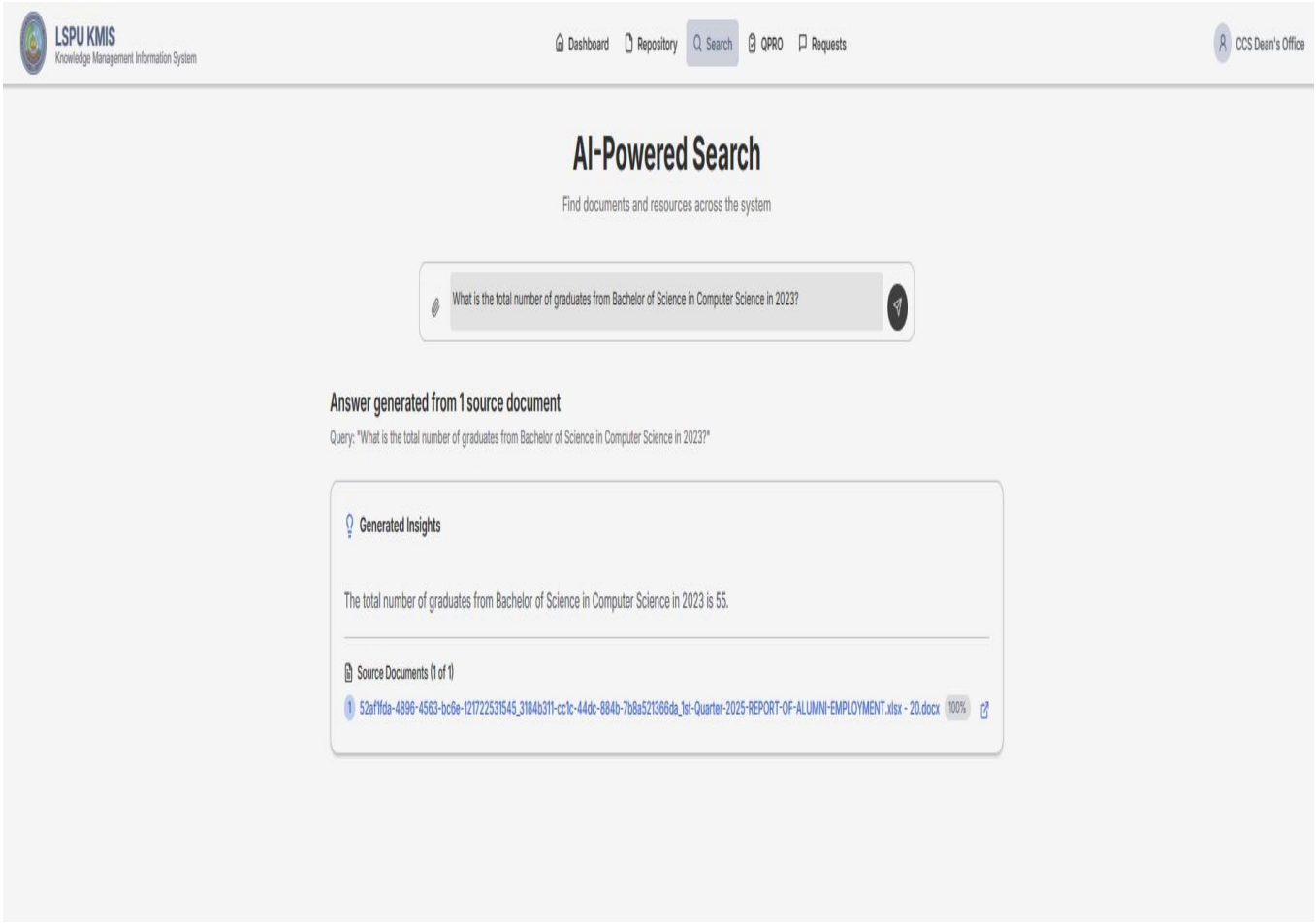


Figure 21. AI-Powered Search Query Output

Figure 21 shows that the system successfully implements semantic search and the Retrieval-Augmented Generation (RAG) pipeline. The interface tests the system's ability to enhance data retrieval processes by displaying the source documents underlying the answer, which significantly enhances search accuracy, optimizes information search and provides evidence-based responses to increase user confidence in the system's ability to retrieve knowledge.

Research Objective 3: To develop an automated quarterly progress report on objectives analysis engine that intelligently evaluates strategic plan progress through AI-powered document analysis, providing real-time insights into key result areas and key performance indicator achievement across organizational units.

The researchers built an end-to-end pipeline for QPRO analyses. This pipeline allows units to submit Quarterly Progress Report of Operation (QPRO) documents which immediately invoke a server-side Analysis Engine. The engine leverages embeddings and vector search to semantically identify the best matching document evidence to the strategic plan. Then it asks an integrated LLM to generate structured outputs of analysis such as alignment narratives, opportunities, gaps, recommendations, and an achievement score calculated for AI-powered interpretation.

The outputs generated by the LLM are verified and stored in the QPRO Analysis database model, connected to the original document. Most importantly, the system automatically extracts activities from the QPRO reports and maps them to Key Result Area/Key Performance Indicator identifiers. The system stores and processes analyses to provide strategic insights in real time. Background processing results in timely performance. The final results are presented through KPI/KRA dashboards and detail views which display achievement percentages and status badges.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

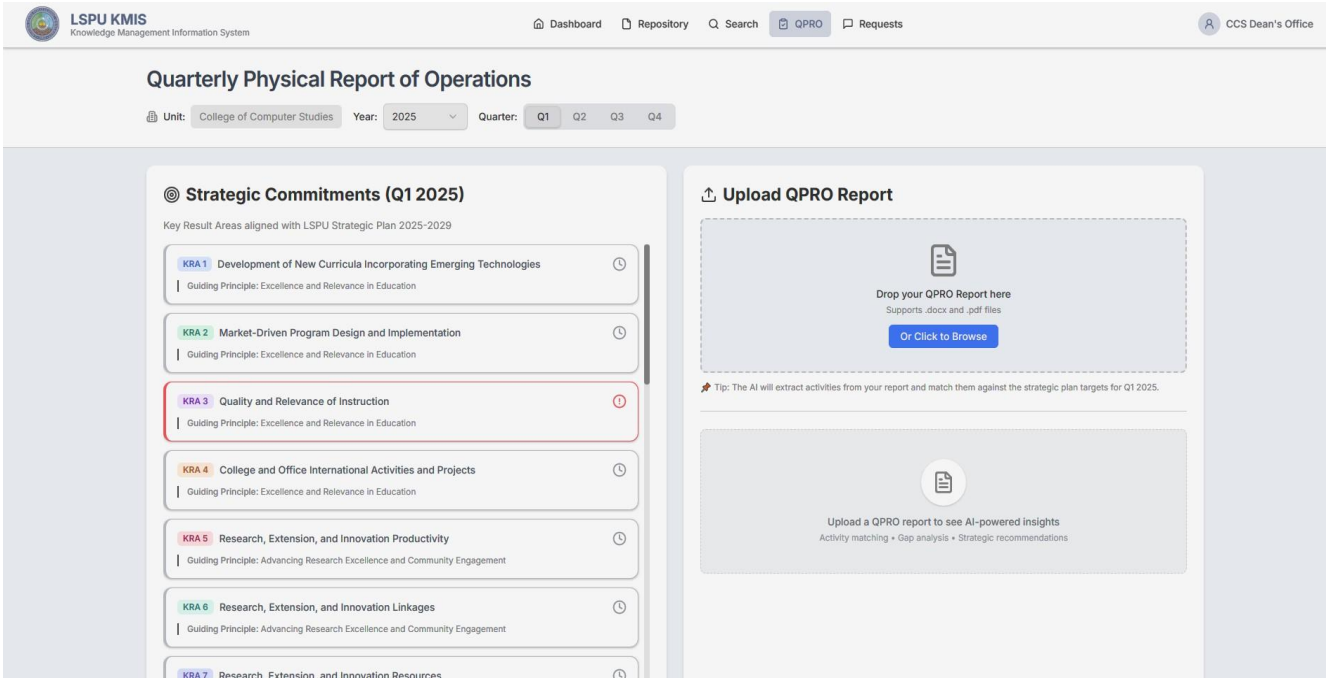


Figure 22. Submission Section of Reports

Figure 22 shows the upload flow of the Quarterly Progress Report of Operation by institutional units. On submission of the document, the server-side Analysis Engine is invoked immediately to initiate the end-to-end process of ingestion of the document, extraction and semantic analysis of the document content using embeddings and Large Language Models, and mapping of activities to Key Result Areas (KRAs) and Key Performance Indicators (KPIs).

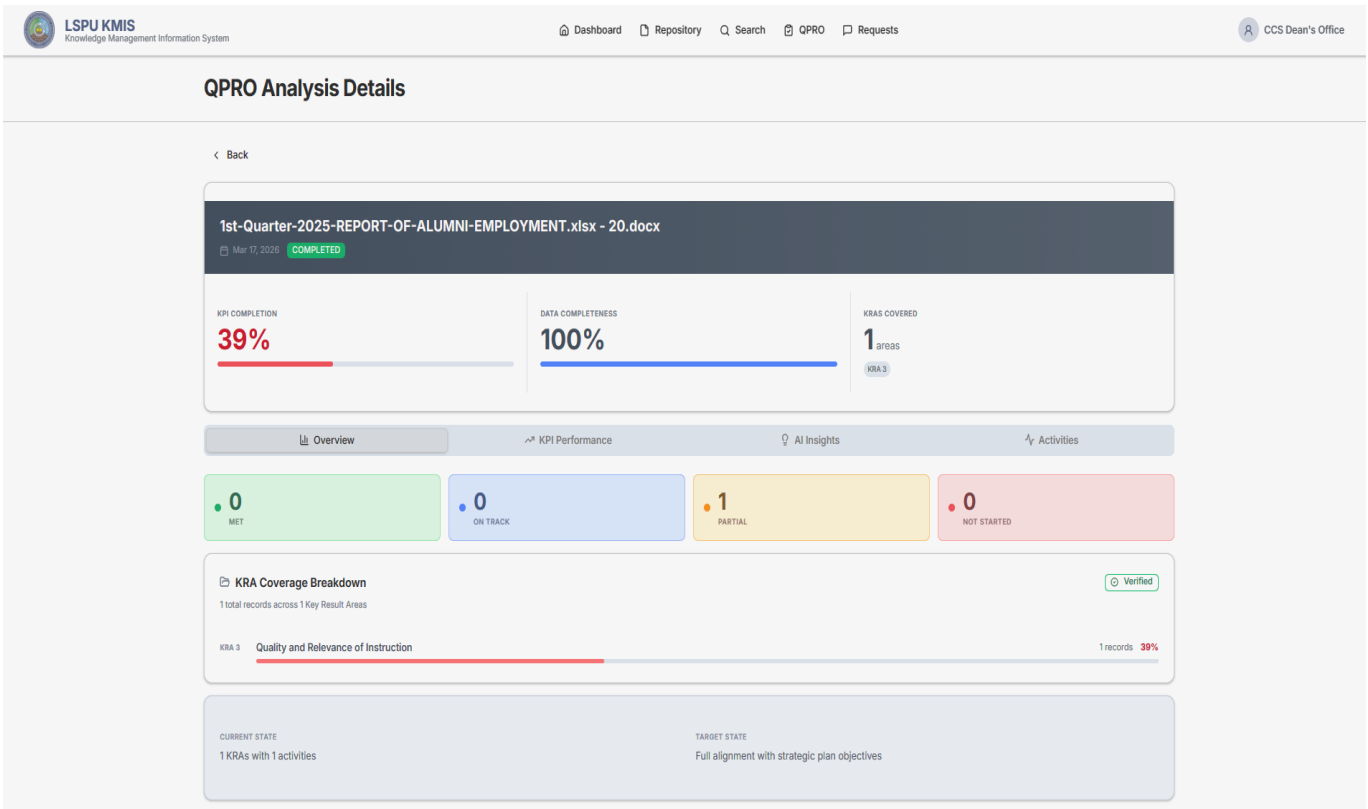


Figure 23. Overview of the Analyses

Figure 23 illustrates the overview section showing the performance tracking dashboard that yields specific insights regarding the institution's strategic alignment and operational performance.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

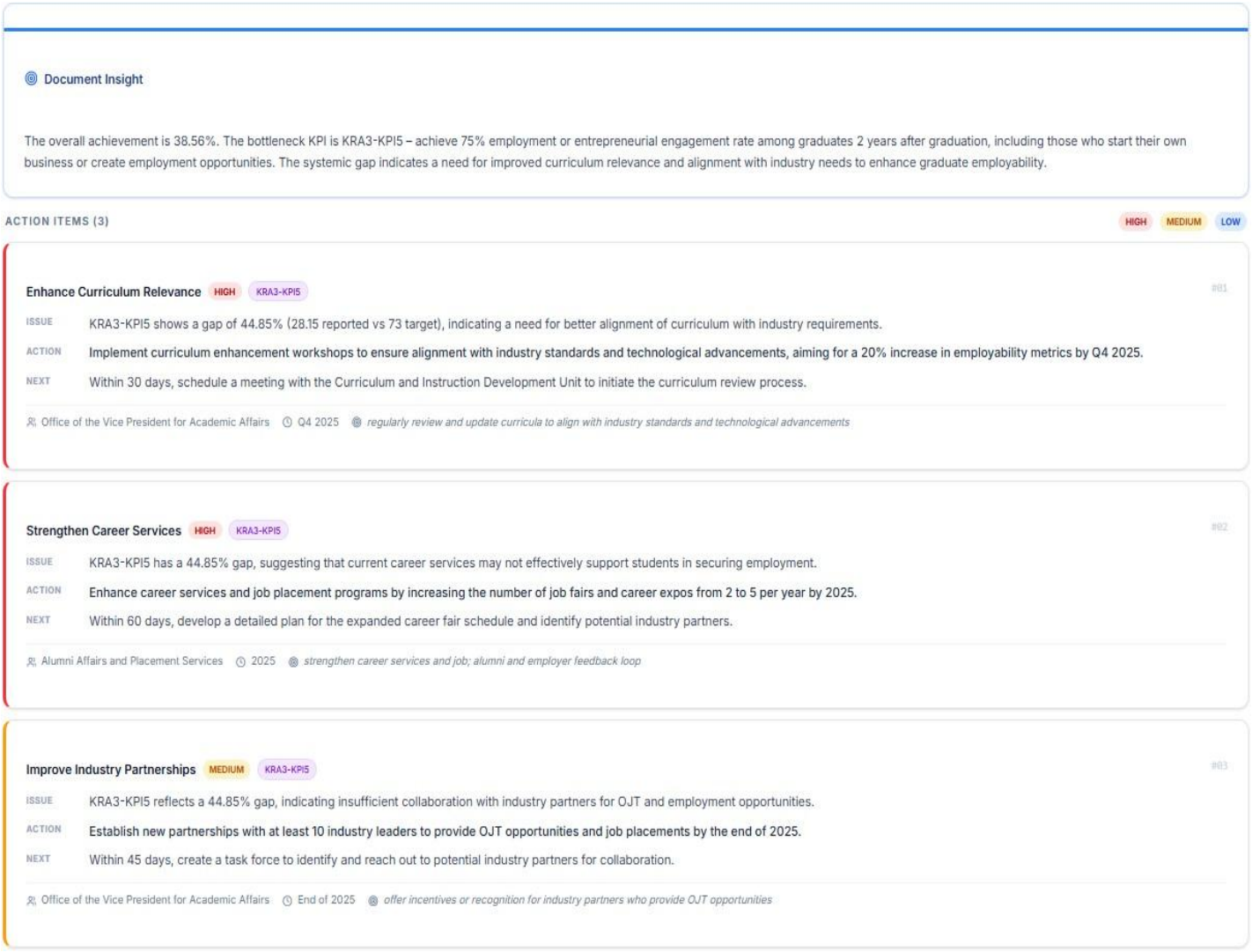


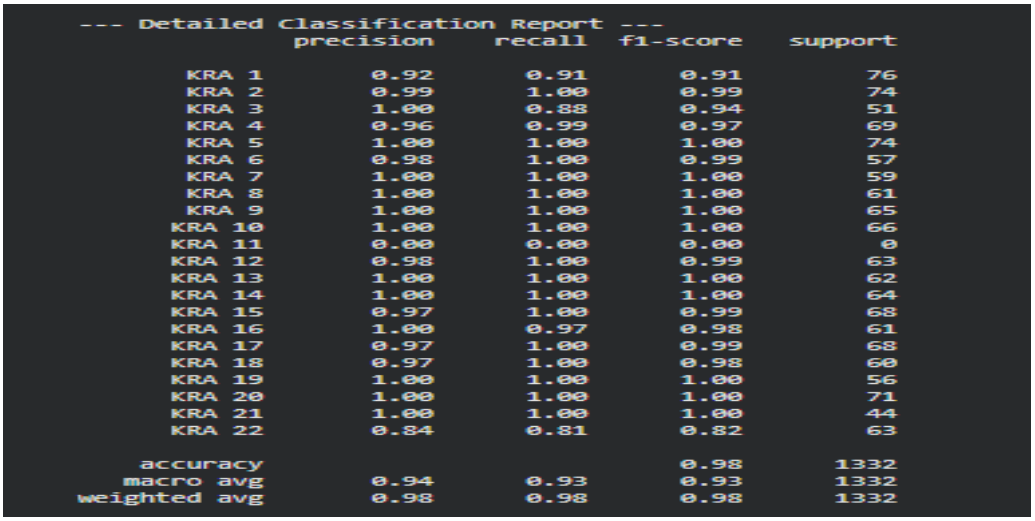
Figure 24. AI Generated Insights

Figure 24 depicts the AI Insights module of the Knowledge Management Information System (KMIS). The diagnostic assessment and the recommended strategic interventions are discussed based on the reviewed data of institutional performance. This automated output transforms raw quarterly reports into real-time data-driven guidance that enables administrators to quickly track KRA achievement and make targeted decisions on priority interventions or resource allocation.

Research Objective 4: To design and train a text classification model for the automated categorization of institutional accomplishments into specific key result areas (KRAs) and to evaluate the performance of the trained model using standard metrics such as accuracy, precision, recall, and F1-score.

In developing and training the text classifier model that classifies institutional achievements into specific KRAs, the researchers systematically developed the transformer-based classifier, which involved a complete process of data preprocessing to extract, clean and oversample the LSPU Strategic Plan 2025-2029, in order to produce a well-balanced dataset of 22 different classes of KRA, as well as to measure the performance of the trained model by means of accuracy, precision, recall and F1-score. For this purpose, the researchers selected the DistilBERT model because of its optimal performance with regards to the trade-off between speed of operation and language comprehension capability. The model was trained to map QPRO text segments to the relevant strategic objectives. In the evaluation phase, the model was put through strict tests to measure accuracy, precision, recall and F1-score, in addition to a confusion matrix test to validate the reliability of the model to automatically segment and classify the operational data, thus forming the core component of the prescriptive analytics engine of the system.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026



```
--- Detailed Classification Report ---
              precision    recall  f1-score   support

       KRA 1       0.92      0.91      0.91        76
       KRA 2       0.99      1.00      0.99        74
       KRA 3       1.00      0.88      0.94        51
       KRA 4       0.96      0.99      0.97        69
       KRA 5       1.00      1.00      1.00        74
       KRA 6       0.98      1.00      0.99        57
       KRA 7       1.00      1.00      1.00        59
       KRA 8       1.00      1.00      1.00        61
       KRA 9       1.00      1.00      1.00        65
      KRA 10       1.00      1.00      1.00        66
      KRA 11       0.00      0.00      0.00         0
      KRA 12       0.98      1.00      0.99        63
      KRA 13       1.00      1.00      1.00        62
      KRA 14       1.00      1.00      1.00        64
      KRA 15       0.97      1.00      0.99        68
      KRA 16       1.00      0.97      0.98        61
      KRA 17       0.97      1.00      0.99        68
      KRA 18       0.97      1.00      0.98        60
      KRA 19       1.00      1.00      1.00        56
      KRA 20       1.00      1.00      1.00        71
      KRA 21       1.00      1.00      1.00        44
      KRA 22       0.84      0.81      0.82        63

    accuracy                           0.98      1332
   macro avg       0.94      0.93      0.93      1332
weighted avg       0.98      0.98      0.98      1332
```

Figure 28. Model's Performance Metrics

The performance metrics such as precision, recall and f1-score validate the effectiveness of the model to automatically segment and recognize institutional reports into 22 Key Result Areas (KRAs) as shown in Figure 28.

Table 10. Overall System Accuracy

Metric	Value
Total Documents Evaluated (N)	25
Total Outputs Rated "PASS"	20
Total Outputs Rated "FAIL"	5
Overall Model Accuracy (%)	80%

Table 10 presents the results of the Ground Truth Verification phase to assess the performance of the AI-powered analysis engine of the Knowledge Management (KM) Information System. Twenty-five completed PDO documents were scored by PASS/FAIL validation and accuracy scoring. Of the 25 documents rated, 20 were rated as "PASS" and 5 were rated as "FAIL", with an overall model accuracy of 80%. The system is considered "Accurate" by the interpretation scale used in the study, indicating that the model was able to properly extract, interpret and analyze the majority of the given PDO documents.

The evaluation results revealed that the majority of the identified errors were categorized as "Data Extraction Errors". These errors occurred when the system incorrectly identified some information from the documents due to inconsistencies in formatting or unclear text structures. Despite these limitations, the system still able to maintain a satisfactory level of overall performance during the evaluation process.

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This study titled "KNOWLEDGE MANAGEMENT (KM) INFORMATION SYSTEM FOR LAGUNA STATE POLYTECHNIC UNIVERSITY - SANTA CRUZ CAMPUS" which intends to develop a web-based Knowledge Management Information System (KMIS) for Laguna State Polytechnic University – Sta. Cruz Main Campus that unites institutional documents into a single, secure, and searchable platform, utilizing role-based access to documents, intelligent or semantic search to improve retrieval of documents, efficient research, and administration. The researchers implemented a secure repository with role-based access control features for core institutional documents and applied semantic search functionality using natural language queries to optimize

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

discovery. The researchers also developed an automated quarterly progress report of operation (QPRO) analysis engine to evaluate strategic plan progress against key result areas (KRA) and key performance indicators (KPI). Lastly, the researchers design and train a KRA classification model to automatically categorize accomplishments into 22 specific KRAs and evaluate its performance using standard metrics.

Conclusion

Based on the findings, objectives and confirmed results from system development and evaluation from the study, the following conclusions are established:

1. The study successfully developed a unified institutional document repository that utilizes Microsoft Azure's blob storage. This centralized platform is secured because of the implemented RBAC (role-based access control) and a mechanism to request permission for cross-unit access.

2. The developed system has produced a highly precise semantic search capability that was achieved through the implementation of Retrieval-Augmented Generation (RAG) pipeline. It allows users to submit natural language queries and uses an integrated Large Language Model (LLM) to convert retrieved document segments into accurate, human-friendly summaries grounded in institutional documents that are uploaded in the repository.

3. The system did a good job at giving prescriptive analysis through the reports that are submitted by every unit. It is with the help of the automated QPRO analysis engine that was developed. Upon document submission, this engine uses vector search to semantically match extracted operational evidence against the strategic plan. The outputs include performance gaps and actionable recommendations that are displayed in dashboards fulfilling the objective to provide intelligent, real-time insights into strategic plan progress.

4. The study designed a KRA text classification model to automatically recognize the institutional accomplishments into the 22 specific Key Result Area (KRA) of the 2025-2029 LSPU Strategic Plan. The model is rigorously evaluated using standard metrics such as precision, recall and f1-score to confirm the model's reliability.

Recommendations

Based on the conclusion drawn from the findings of the study, the following recommendations are offered to improve and address the identified challenges.

1. Use an excellent Optical character Recognition (OCR) for extracting tables from PDF files as the current system supports only DOCX files.

2. They could include a conversational AI assistant or chatbot into future research. This improvement noted as a future upgrade in the scope and limitations will allow users to query the knowledge base using natural language that provides a more intuitive and context-aware experience than the current chat-style search.

LITERATURE CITED

1. AN APPROACH TO SEMANTIC EDUCATIONAL CONTENT MINING USING NLP. (2022, July 19). Proceedings of the 16th International Conference on E-Learning (EL 2022). 16th International Conference on e-Learning. https://doi.org/10.33965/EL2022_202203L0 05
2. Anshari, M., Syafrudin, M., Tan, A., Fitriyani, N. L., & Alas, Y. (2023). Optimisation of Knowledge Management (KM) with Machine Learning (ML) Enabled. Information, 14(1), 35.
3. https://doi.org/10.3390/info14010035 Antonio de Carvalho Junior, M., & Bandiera-Paiva, P. (2021). Implications of loosened Role-based Access Control session control implementation for the enforcement of Dynamic Mutually Exclusive Roles properties on Health Information Systems. Informatics in

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Medicine Unlocked, 27, 100780. https://doi.org/10.1016/j.imu.2021.100780 Che,
4. T.-Y., Mao, X.-L., Lan, T., & Huang, H. (2024). A Hierarchical Context Augmentation Method to Improve Retrieval-Augmented LLMs on Scientific Papers. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 243–254. https://doi.org/10.1145/3637528.3671847
5. Dube, T. V. (2025). Research data management in academic libraries: Institutional repositories as a reservoir for research data. Library Management, 46(5), 319–331. https://doi.org/10.1108/LM-06-2024-0070
6. Eguchi, M., & Kyle, K. (2024). Building custom NLP tools to annotate discourse-functional features for second language writing research: A tutorial. Research Methods in Applied Linguistics, 3(3), 100153.
7. Faruqui, S. H. A., Tasnim, N., Basith, I. I., Obeidat, S., & Yildiz, F. (2024). Integrating A.I. in Higher Education: Protocol for a Pilot Study with "SAMCares: An Adaptive Learning Hub" (No. arXiv:2405.00330).
8. Galgotia, D., & Lakshmi, N. (2022). Implementation of Knowledge Management in Higher Education: A Comparative Study of Private and Government Universities in India and Abroad. Frontiers in Psychology, 13, 944153. https://doi.org/10.3389/fpsyg.2022.944153
9. Hafeez, S., Shahzad, K., Helo, P., & Mubarak, M. F. (2025a). Knowledge management and SMEs' digital transformation: A systematic literature review and future research agenda. Journal of Innovation & Knowledge, 10(3), 100728. https://doi.org/10.1016/j.jik.2025.100728
10. S., Shahzad, K., Helo, P., & Mubarak, M. F. (2025b). Knowledge management and SMEs' digital transformation: A systematic literature review and future research agenda. Journal of Innovation & Knowledge, 10(3), 100728. https://doi.org/10.1016/j.jik.2025.100728
11. Jochim, C., Gleize, M., Bonin, F., & Ganguly, D. (2021). TDMSci: A Specialized Corpus for Scientific Literature Entity Tagging of Tasks Datasets and Metrics. In P. Merlo, J. Tiedemann, & R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (pp. 707–714). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.
12. Kitto, S., Chiang, H. L. M., Ng, O., & Cleland, J. (2025). More, better feedback please: Are learning analytics dashboards (LAD) the solution to a wicked problem? Advances in Health Sciences Education, 30(1), 69–85. https://doi.org/10.1007/s10459-024-10358-8
13. Kliimask, K., & Nikiforova, A. (2024). TAGIFY: LLM-powered Tagging Interface for Improved Data Findability on OGD portals (No. arXiv:2407.18764). arXiv. https://doi.org/10.48550/arXiv.2407.18764
14. Kulkarni, A., & Eagle, M. (n.d.). Towards Understanding the Impact of Real-Time AI-Powered Educational Dashboards (RAED) on Providing Guidance to Instructors.
15. Kurniawan, S. S., Maharani, N. Z., Sensuse, D. I., Purwaningsih, E. H., & Hidayat, D. S. (2024a). Applying User Centered Design and System Usability Scale to Design Knowledge Management System for Exam Proctors in Higher Education. Scientific Journal of Informatics, 11(4), 1043–1056. https://doi.org/10.15294/sji.v11i4.9919
16. Lhoest, Q., Villanova del Moral, A., Jernite, Y., Thakur, A., von Platen, P., Patil, S., Chaumond, J., Drame, M., Plu, J., Tunstall, L., Davison, J., Šaško, M., Chhablani, G., Malik, B., Brandeis, S., Le Scao, T., Sanh, V., Xu, C., Patry, N., … Wolf, T. (2021). Datasets: A Community Library for Natural Language Processing. In H. Adel & S. Shi (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 175–184). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-demo.21
17. Liu, M., & Xu, J. (2025). NLI4DB: A Systematic Review of Natural Language Interfaces for Databases (No. arXiv:2503.02435). arXiv. https://doi.org/10.48550/arXiv.2503.02435
18. Mehta, A. (2021). Knowledge Management in Higher Education Institutions: A Framework to Improve Collaboration. Asian Review of 129 Mechanical Engineering, 10(2), 43–47. https://doi.org/10.51983/arme2021.10.2.3179
19. Methods and Applications of Fine-Tuning Llama-2 and Llama-Based Models: A Systematic Literature Analysis. (2024). Journal of System and Management Sciences. https://doi.org/10.33168/JSMS.2024.1015
20. Mosha, N. F., & Ngulube, P. (2023). Metadata Standard for Continuous Preservation, Discovery, and Reuse of Research Data in Repositories by Higher Education Institutions: A Systematic Review.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Information, 14(8), 427. https://doi.org/10.3390/info14080427

21. Muhammad farid fadhlan, null, & Dana indra sensuse, null. (2022). Knowledge Repository Design to Improve Knowledge Management Process Capabilities: A Systematic Literature Review. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 6(2), 246–251. https://doi.org/10.29207/resti.v6i2.3929