FEASIBILITY STUDY
LEGAL INFORMATION RETRIEVAL SYSTEM FOR SRI LANKA

06/08/2025

Date	Version	Description	Author
02/08/2025	1.0	Initial Feasibility Report	Niduni Kasige

Team Members:
220311C - Kasige N. D.
220315R - Kavinda L. G. N.
220323N - Kodithuwakku K.S.

Group ID: G-15
Project ID: 11
Mentor: Dr. Nisansa de Silva

TABLE OF CONTENTS

1. Introduction
    1 Overview of the Project
    1.2 Objectives of the Project
    1.3 The Need for the Project
    1.4 Overview of Existing Systems and Technologies
    1.5 Scope of the Project
    1.6 Deliverables.
2. Feasibility Study
    2.1 Financial Feasibility
    2.2 Technical Feasibility
    2.3 Resource and Time Feasibility
    2.4 Risk Feasibility
    2.5 Social/Legal Feasibility
3. Considerations
4. References

1. Introduction

1.1 Overview of the Project

This document outlines the key stages and the development plan of the proposed system of Legal Information Retrieval System, which facilitates providing legal information and advisory support. This system needs to be a reliable and user-friendly with personalized legal documents recommendations. To carry out this purpose, system is built with commonly used technologies and industry best practices.

1.2 Objectives of the Project

The objective of this project is to simplify the information retrieval process, providing legal information in more reliable and understandable manner. In addition to that this project aims to give personalized recommendations for legal documents enabling users to find the documents they need the most.

1.3 The Need for the Project

Retrieving legal information in a human understandable way is a challenging task for a system to deal with. The legal constraints on collecting data and the high accuracy of the results to be produced, make it more challenging. Furthermore, finding the right legal documents is also hard. This project was proposed to support the above-described problems.

1.4 Overview of Existing Systems and Technologies

To address the legal information access problem, there are some similar projects that provide legal information via an AI agent or legal document library. For example,

- **Lanka Law**
  This system consists of AI assistant to provide legal information and advisory services, research services with generated reports, access to legal documents via e-law library feature with the support for all 3 languages.

- **AIPazz**
  This system provides facility to search across legal documents, get AI powered assistance, receive updates on newly issued legal documents.

1.5 Scope of the Project

This project mainly focusses on providing users the ability to get legal information with our virtual assistant, personalized legal documents recommendation (to support users to find the most needed documents using previous search history), get access to legal information in their preferred language. (Current implementation plan is to support Sinhala and English languages)

1.6 Deliverables

This project delivers following key deliverables:

- A fully functional **web-based application** deployed on a production server.
- A responsive graphical user interface (GUI) for all users.
- Complete technical documentation including architecture diagrams, API specifications, and evaluation reports.
- A user manual for non-technical stakeholders.
- A demonstration video of the system.

2. Feasibility Study

2.1 Financial Feasibility

The system uses open-source technologies such as Hugging Face Transformers, FAISS and GitHub Actions, significantly reducing licensing and operational costs. Infrastructure expenses are manageable with cloud-based deployment (AWS/GCP), with the possibility of scaling as needed (with the facility to pay as you go). Therefore, the financial investment is minimal compared to the value provided by simplifying the legal information retrieval and update process.

2.2 Technical Feasibility

The proposed solution is technically feasible as it uses current industry tools and frameworks. The tools and frameworks planned to use are listed below.

- **Chatbot & RAG**: Retrieval-Augmented Generation is planned to implement using vector stores (FAISS) and transformer-based LLMs.
- **Multilingual Support**: Language-agnostic embeddings such as LaBSE or multilingual SBERT support Sinhala and English effectively.
- **Recommendation System**: Vector similarity with user history storage is supported using Python and libraries like scikit-learn or sentence-transformers.
- **Automated Updates**: Scheduled scripts via GitHub Actions support efficient fetching and embed new legal documents.

The system is developed using modern AI and NLP tools and technologies which secure the system's long-term stability, and adaptability to evolving technologies.

2.3 Resource and Time Feasibility

The system can be developed with a proficiency in Python, NLP, and web technologies. Hardware requirements include machines with GPU access during the model training and embedding stages. But the inference can be optimized for CPU.
In terms of project timeframe, the development is estimated to complete within 10–12 weeks, covering data processing, model integration, system testing, deployment and documentation with all the deliverables.

2.4 Risk Feasibility

This system has several risks and issues which can be mitigated using modern technologies and best practices. Key risks and mitigation strategies include:

- **Model Bias or Inaccuracy**: Removed by regular testing, dataset validation, and user feedback loops.
- **Language Handling Challenges (Sinhala)**: Addressed through evaluation with validation datasets and LLMs.
- **Document Update Failures**: Handled by logging and alerts during automated pipeline failures (via GitHub Actions).
- **Data Privacy Concerns**: Addressed through minimal data storage and compliance with local legal data regulations.
- **Legal Concerns in Retrieving Information**: Handled by using previously retrieved documents dataset (which includes data on each document, such as document type, language, URL to pdf document etc.)

2.5 Social/Legal Feasibility

The system enhances public access to legal knowledge, promoting transparency and legal literacy. It supports bilingual access (Sinhala and English), improving inclusivity across Sinhala and English speakers.

From a legal perspective, all documents used are government-published and publicly accessible, and the system will use applicable data protection laws and ethical AI use guidelines.

To ensure the system is usable for any user, it is leveraged with simple and user-friendly interfaces. For a better understanding a demonstration video will also be also provided.

3. Considerations

The system design prioritizes the following aspects:

- **Performance**: High performance is ensured using caching strategies, and lightweight front-end design for quick response times
- **Security**: Security concerns of the system are reduced with role-based access control, HTTPS, and secure authentication mechanisms
- **Usability**: Usability of the system is supported using intuitive interfaces with clear navigation, tooltips, and responsive design
- **Maintainability**: Modular codebase with clear documentation to facilitate updates
- **Scalability**: The system development is done using an incremental approach. This model facilitates introducing newer features and functionalities to the system. Also, architecture supports future growth in users or features.
- **Reliability**: The reliability of the response is also a concern. It is mitigated with robust content retrieval approach (using vector similarity and keyword search).

4. References

Dias, P. S. P., Weerasinghe, T. W. N. A., & Perera, A. R. T. K. (2024). *An integrated approach to enhance legal information retrieval of Sri Lankan Supreme Court verdicts*. ResearchGate

Wijeweera, S. N., & Samarasinghe, W. M. D. N. (2017). *ICT-based legal information retrieval system for Sri Lanka*. APJCECT

LankaLaw Platform: https://lankalaw.net/

AIPazz Legal AI Tool: https://www.aipazz.com

LaBSE: Language-Agnostic BERT Sentence Embedding: https://huggingface.co/sentence-transformers/LaBSE

GitHub Repository of legal document data: https://github.com/nuuuwan/lk_legal_docs

FAISS Documentation: https://python.langchain.com/docs/integrations/vectorstores/faiss/

BM25 for lexical search: https://huggingface.co/blog/xhluca/bm25s

Embedding model for Legal Documents: https://huggingface.co/nlpaueb/legal-bert-base-uncased

Sentence Transformers: https://huggingface.co/sentence-transformers