PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE - 411043
Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043

Synopsis On
“LegalBuddy”

Project Based Learning-II
AY: 2024 – 25
Batch – E4

Group Members:

Sharv Mahajan – 21411

Satyam Mengshetti – 21412

Mufaddal Ali - 21413

PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE - 411043
Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043

Title: **LegalBuddy** – Legal Document Information Extraction and Chatbot

Introduction:

In the legal domain, the volume of documents such as contracts, agreements, legal notices, and other forms of legal paperwork is enormous. Manually extracting relevant information from these documents is time-consuming and error-prone. This project proposes a software system capable of extracting key data from legal PDF documents and integrating a chatbot interface to assist users in querying extracted data in a conversational manner. The system will allow users to easily interact with legal content, improving efficiency and accessibility.

Motivation:

Legal professionals, including lawyers, paralegals, and clients, often need to navigate large volumes of legal text. Manual search and interpretation are tedious and require expertise. By automating the extraction of critical information from legal PDFs and integrating an AI-powered chatbot, this system aims to reduce human effort, increase productivity, and make legal content more accessible to non-experts.

Objective:

1. To develop a software tool that extracts relevant legal information from PDFs (e.g., names, dates, clauses, signatures).
2. To integrate a chatbot capable of understanding natural language queries related to the extracted content.
3. To enable users to query legal documents interactively, retrieving specific information with ease.
4. To provide a user-friendly interface for legal professionals and general users alike.

PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE - 411043
Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043

S/W and H/W requirement:

Software:

- **Programming Languages:** Python (Primary), JavaScript (for web interface)
- **Libraries/Tools:**
  - PDF parsing libraries (e.g., PyPDF2, pdfminer)
  - Natural Language Processing (NLP) libraries (e.g., spaCy, NLTK, OpenAI GPT)
  - Machine Learning Frameworks (e.g., TensorFlow, PyTorch for AI models)
  - Chatbot development platforms (e.g., Rasa, Dialogflow)
  - Frontend Framework: ReactJS (for web-based interface)

Hardware:

- Personal Computer with at least 8GB RAM, 4-core processor
- Storage: Minimum 5GB available space
- Internet Access (for AI model training, API integration)

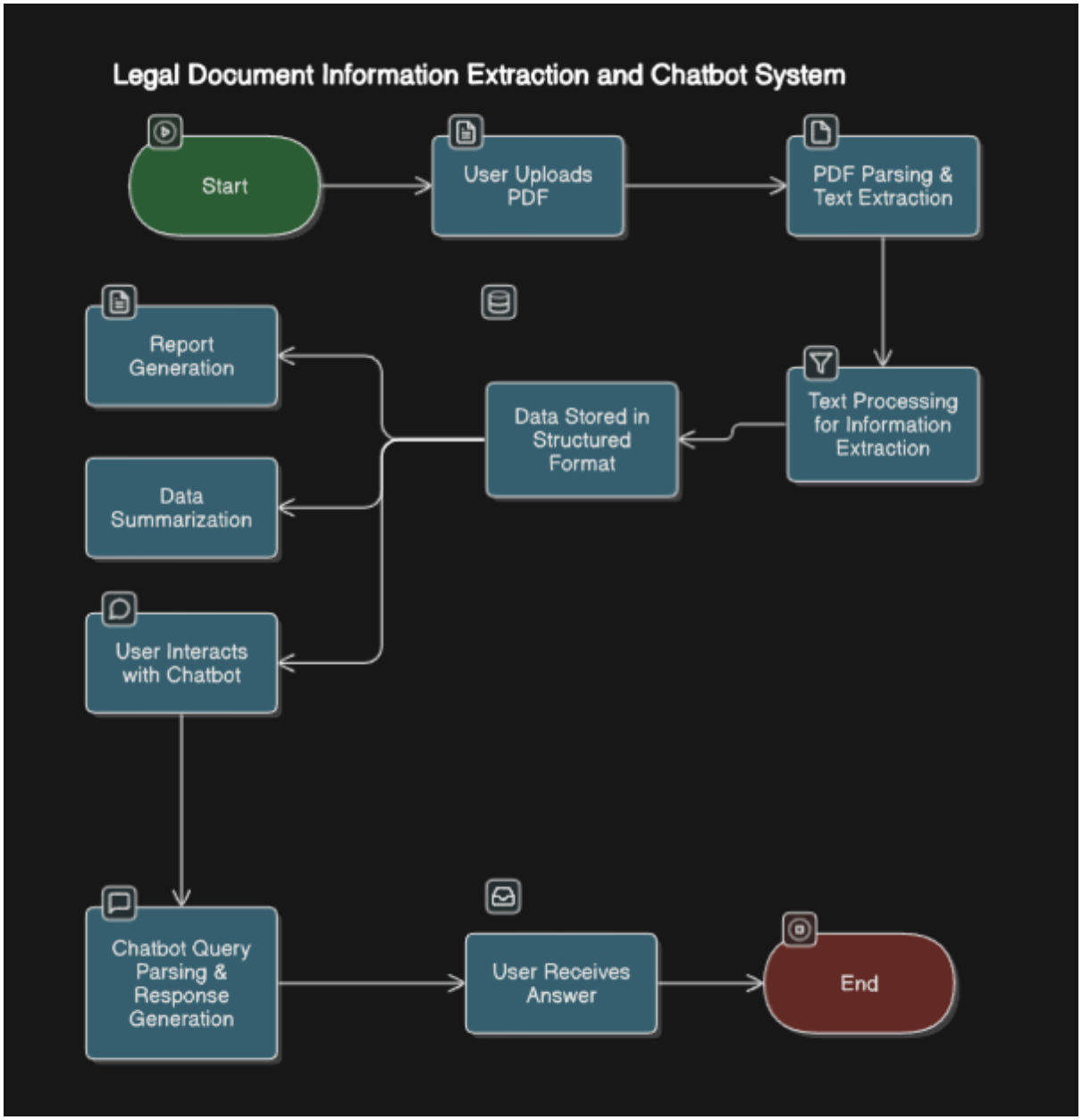
Theory/ Short Description:

- This software uses Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques to analyze legal documents in PDF format. The key components include:

- **PDF Parsing**: The system will extract text from scanned or digitally-created legal PDFs using libraries like pdfminer or PyPDF2.
- **Information Extraction**: Using NLP, the system will identify and extract relevant legal information such as parties involved, case numbers, dates, legal clauses, and signatures.
- **Chatbot Integration**: The system incorporates a chatbot powered by advanced NLP models. The user can query the chatbot about specific details like "What are the key clauses in the contract?" or "What is the termination date?"
- **User Interface**: A simple web-based UI will be developed for users to upload their legal PDFs, interact with the chatbot, and view extracted data.

PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE - 411043
Department of Computer Engineering
S.No.-27, Pune Satara Road, Dhankawadi, Pune-411043

Flowchart/Use case diagram:



References:

1. https://pythonprogramming.net/pdf-parsing-python-pdfminer-tutorial/
2. https://www.rasa.com/
3. https://spacy.io/usage
4. https://www.nltk.org/
5. https://groq.com/

Conclusion:

This software will bridge the gap between legal professionals and the vast number of legal documents they deal with daily. By extracting relevant data from PDFs and providing an interactive AI-powered chatbot interface, users will save time and effort while working with complex legal texts. This project aims to provide a modern, efficient, and user-friendly solution to legal document management.