# Software Design Document
for
# Land Deed Information Retrieval System
Government Engineering College, Thrissur

Ajay Mathew
Amal Krishna JK
Praveen K Nair
Vaishnav NT

December 7, 2022

# Contents

# 1 Introduction

## 1.1 Purpose

The purpose of this document is to delineate the design of our project titled - 'Land Deed Information Retrieval System' in different Design Viewpoints.

## 1.2 Scope

The Land deed information retrieval system will be a platform used to query land records and acquire necessary information with ease. The question will be in natural language so that anyone can make use of it and the answers will be precise and to the point. This will free users from the pain staking task of having to manually search and understand the entire document in order to get some details.

## 1.3 Definitions, Acronyms and Abbreviations

### 1.3.1 Definition

- **Authentication** :- Authentication is the act of confirming the truth of an attribute of a datum or entity. In Computer Networks, it is verifying that a client is what it claims to be.

- **Supplicant**:- In IEEE 802.1x Authentication protocol Scenario, Supplicant is the devices requesting access to the network resources. It is the device that supplies information to the RADIUS Server.

- **Authenticator**:- Authenticator in IEEE 802.1x Scenario is the Access Point/Router that communicates with the Authentication Server(RADIUS, etc) to authentcate Supplicants requesting access to the network.

- **Client** :- Clients in the scope of this document are devices requesting access to the network. In the Scope of this document, Clients and Supplicants are used interchangably.

## 1.4 Abbreviations

- **LDIR** :- Land Deed Information Retrieval

# 2 References

- IEEE SDD Standard -SDD-ieee-1016-2009.pdf in local folder

-

# 3 Architectural Description

Figure 1 abstracts the overall architecture of the project. Here the Natural Language Processing techniques and the application interface constitutes Software interfaces while the client device and the central servers make up the hardware components.

## 3.1   Software Interfaces

The software interface of the system consists mainly of the NLP-based models for extracting named entity extraction as well as relation extraction. Additionally, we make use of a module for creating and updating the Knowledge Graph. All these are integrated to establish a Question Answering model that effectively understands what the user is requesting and search and identify the required answer to the query from the knowledge graph.

## 3.2   Hardware Interfaces

Since we are dealing with cross-platform as well as the cross-device types, the hardware components mainly consist of many client devices with limited computational capability.

# 4   Decomposition Description

The System is divided into 4 modules based on the functionality of the system. A concise summary of the module and its functionality is provide below. The Modules are :-

- **Authentication** - This Module mainly deals with the user Authentication. It would receive request from the user to gain access to the system. This request would be process and if verified, the the user would be authorized into the system and can make use of its functionalities. The access credentials can be generated only by the super user of the system and all users must make a request to the super user in order obtain login credentials.

- **Object Character Recognition** - This Module tackles the task of converting the land deed documents in Portable Document File format to the machine encoded textual format is needed to carry out downstream tasks. The module also handles necessary pre-processing needed to improve the quality of the inputs.

- **Named Entity Extraction** - This Module will be used to obtain the key named entities that are present in the textual data extracted from the document. This serves as a preliminary step in the construction of knowledge graph which is at the center of the system. It is also made use of while processing the user questions to understand the context and intent of the query. This will help establish an idea about what the user is requesting for and is used further down to infer the answer.

- **Relation Extraction** - This module will be used to find the underlying relationships present between the entities in the processed textual data. This module provides a key ingredient used in the construction of knowledge graph. Similar to the Named Entity Extraction, It is also made use of while processing the user query's as well.

# 5   Design

## 5.1   Data Flow Diagram

The Dependencies between the modules and the flow of information is expressed in the DFD diagrams shown in figure 2 and figure 3.

## 5.2   Use Case Diagram

A use case diagram is a graphic depiction of the interaction among the elements of a system. Figure 4 illustrates the use case diagram of the system.
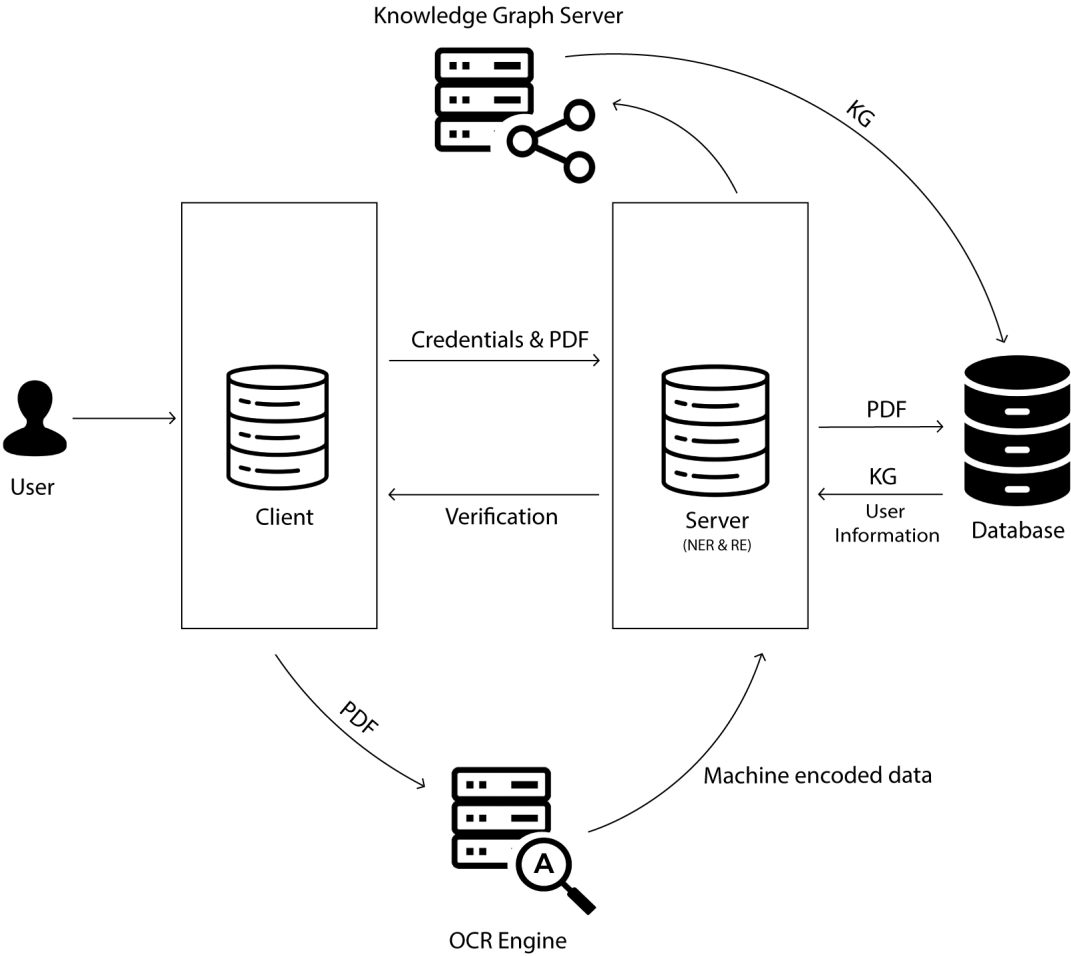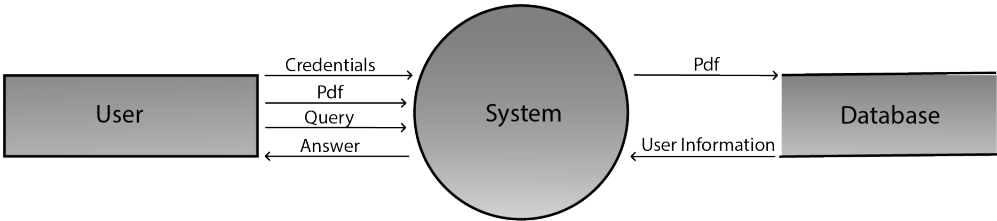
Knowledge Graph Server



Figure 1: Architecture



Figure 2: Data Flow Diagram Lvl 0
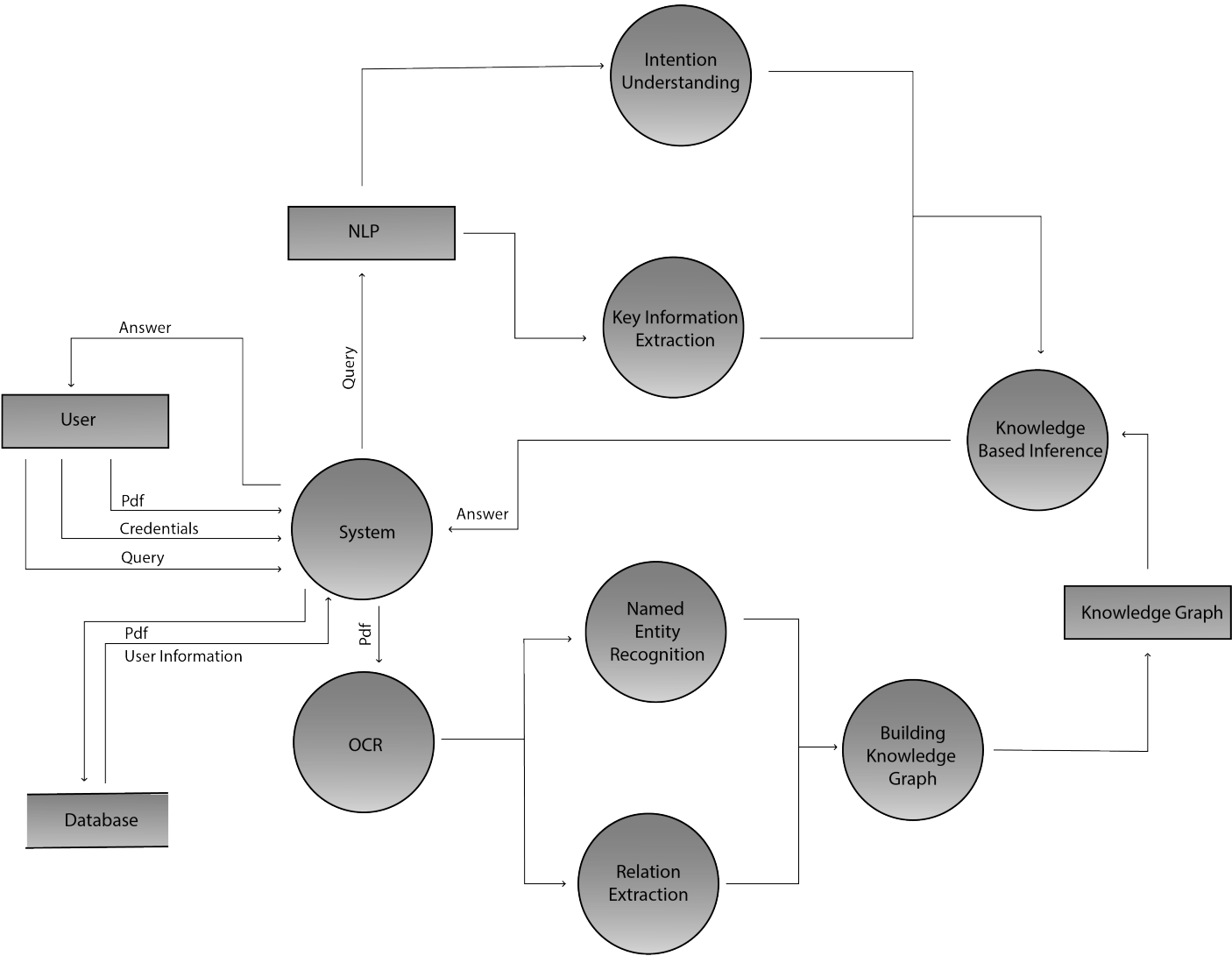
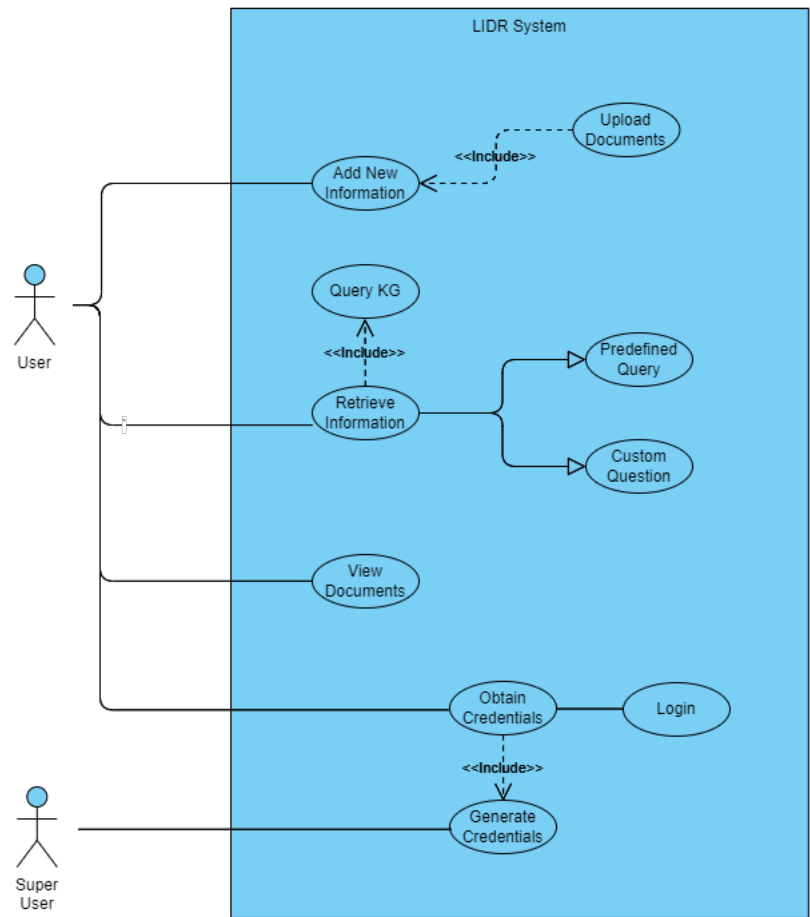Figure 3: Data Flow Diagram Lvl 1
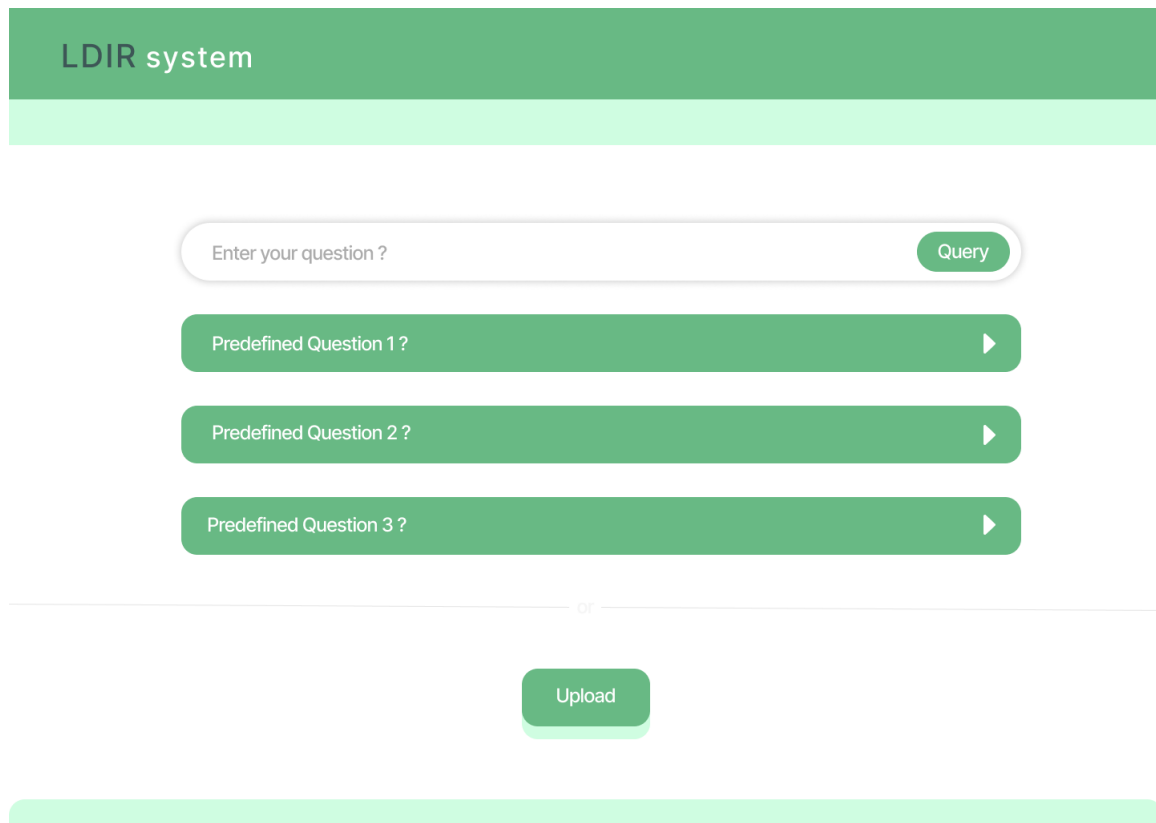
Figure 4: Use Case Diagram

# 6    User Interface Description

## 6.1    Querying User Interface

There would be a Querying User Interface which is to be used by the users to input their questions requesting necessary information or they can directly select a predefined query from the list provided below query text box. The provision to add new land deed documents to the system is also made available here under the upload option. The Querying user interface would look like :-
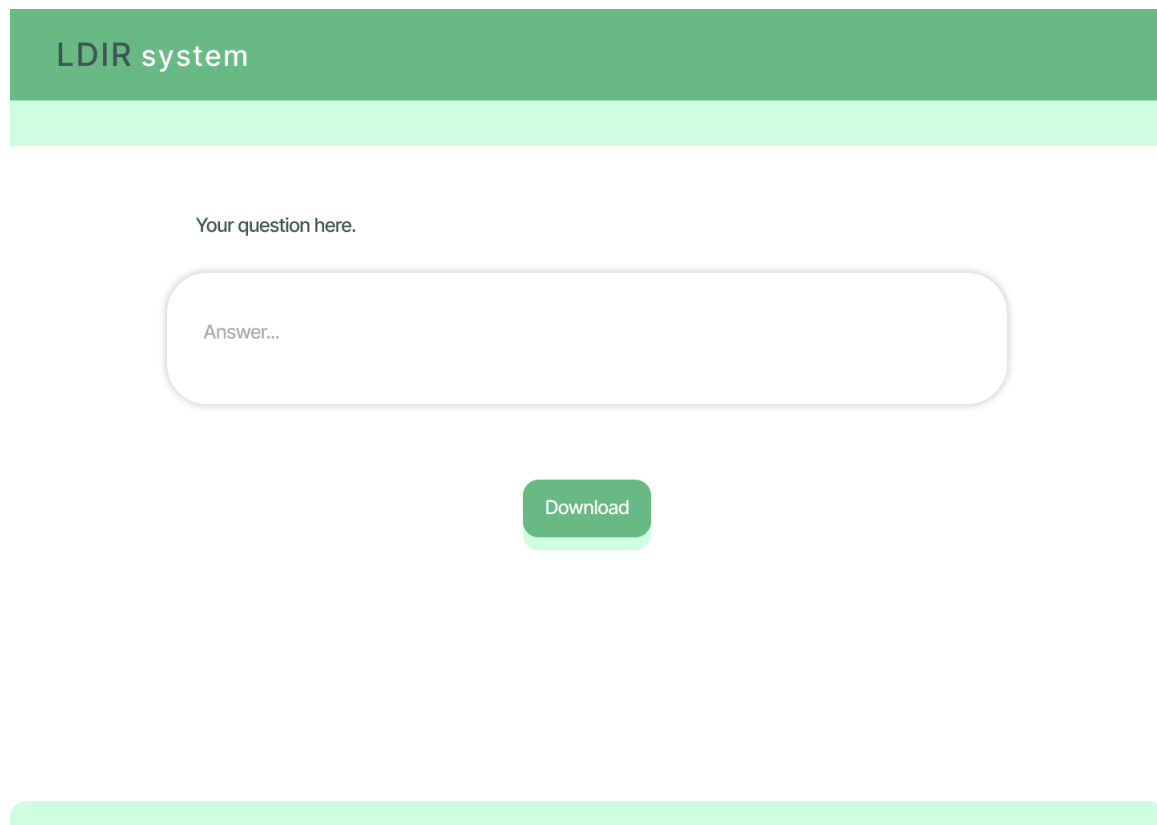


Figure 5: Querying User Interface

## 6.2    Output User Interface

The Output User Interface would consist of a field displaying the result of the user query. It also contain a button that lets you download the original document file from which the information was retrieved. The Output user interface would look like :-

Figure 6: Output User Interface

# 7    Detailed Design

## 7.1    Authentication

Authentication deals with the process of verifying the identity of client. Using a valid login credential, client can gain access to the system. Through authentication, we will crosscheck these credentials and verify that the client is registered with the system or not. Only after confirming the client's identity, will they be able to make use of the functionalities of the system. The initial credentials are are to obtained directly from the super user. These can be updated by the user in the future. Authentication helps to make the system more secure and it also helps to restrict the actions of the client with their specified limits. The credentials used for the authentication are encrypted making use of standard encryption algorithm to ensure confidentiality.

## 7.2    Object Character Recognition

OCR, Objected Character Recognition is a process of conversion of digital textual data into machine-encoded text. This process is useful for creating a textual content out of Portable Document File which is much easier to process and handle in order to obtain desired results. The computer system need data to be in a machine readable form and we can't directly make use of PDF. With the help of OCR techniques we will solve this problem.

## 7.3    Named Entity Extraction

Named entity recognition and extraction is a core part of the Information Extraction system. It facilitates the automatic detection and classification of entities in natural language text, present in the land deed document as well as the questions posed by the client into predefined categories such as the persons, organizations, locations, and so on. Named entity extraction task is crucial and along with relation extraction facilitates the construction of knowledge graph which sits at the center of the entire question answering system.

## 7.4    Relation Extraction

Relation Extraction is also an vital process that helps in the construction of Knowledge Graph. This process enables the extraction of the underlying relationships that are present between the different entities in the data. In this system, Relation Extraction is also used to understand the intent of the user query. The Relation Extraction Module will make use of the entities recognized by the Named Entity Extraction.