

1. Define Objectives:

The objectives achieved by the legal document analyzer is extracting key information, identifying relevant clauses, or assessing compliance. Additionally, the analyzer provides functionalities such as store, retrieve and update the extracted data. Furthermore it also provides the capability of generating a summary of the legal document.

2. Select Document Types:

The types of legal documents used for analysis, such as contracts, agreements, or legal briefs. This will help focus the PoC on relevant use cases.

3. Data Collection:

We curated an extensive dataset from various legal sources, including law firms and public records, encompassing a diverse range of document types like contracts, agreements, ensuring comprehensive coverage for varied document structures and complexities.

4. Data Preprocessing:

The data accepted in the form of pdf or word document is processed and the extracted data is stored in the a csv file in a tabular format of key and value.

5. Feature Extraction:

The analyzer extracts queries like legal party names, key business terms such as effective start and end date, geographical region, payments due , technical specifications etc.

6. Model Selection:

'bert-large-uncased-whole-word-masking-finetuned-squad' is a pretrained natural language preprocessing model used for Question-Answering, given its fine tuning on SQuAD dataset. It can be employed to answer questions based on a given context or passage.

7. Testing:

Rigorous evaluation using separate test datasets and metrics like precision, recall, F1 score, and accuracy facilitated thorough error analysis. The iterative refinement based on testing results significantly boosted the system's accuracy and precision in handling legal document information.

8. User Interface (UI):

A user interface is designed for the users to provide easy upload of the document. After the document is processed the queries will be displayed which is further equipped with the ability to make updates. Further a button for summary is provided which displays summary of the document on the display.Ui interaction is smoother and can perform multiple tasks at a time without going to unresponsive

9. Integration with Legal Databases:

Our system seamlessly integrated with external legal databases and APIs, empowering cross-referencing and data validation. This integration enriched the extracted data with supplementary information from these sources, enhancing our system's accuracy and depth.

10. Security and Compliance:

Implementing robust encryption and access control measures ensured the protection of sensitive legal information using MongoDB Injection. These measures were meticulously designed to adhere to stringent data protection regulations and maintain compliance with legal standards.

11. Performance Optimization:

Efforts in optimizing performance, including resource allocation and the implementation of caching mechanisms, significantly enhanced processing speed and scalability. These optimizations ensured efficient handling of large volumes of legal documents, improving overall system performance.

12. Documentation:

Our comprehensive documentation, covering technical specifications, system architecture, user guides, and API documentation, served as an indispensable resource. This detailed documentation facilitated effective system utilization and seamless integration for developers interacting with our system.

13. Feedback and Iteration:

Continuous feedback gathering and incorporation ensured ongoing system enhancements. By maintaining multiple system versions, we allowed smooth iterations and rollbacks when needed, ensuring a seamless user experience during the system's evolution.

14.Algorithms:

PDF and Word Document Processing:

- Use pdfplumber for PDF extraction.

- Use docx2pdf or other libraries for Word to PDF conversion.

- Extract data row-wise from both formats.

Data Export to CSV:

- Iterate through the data and export it to CSV in the specified sequence.

Question Answering with BERT:

Utilize a pre-trained BERT model for question-answering tasks.

- Tokenize and encode the input data.

- Extract relevant embeddings and apply them to the question.

- Decode and extract the answer.

MongoDB Database Operations:

- Use pymongo for database interactions.

- Insert, update, and retrieve data from MongoDB.

- Create and update a dictionary for each document.

JSON File Creation:

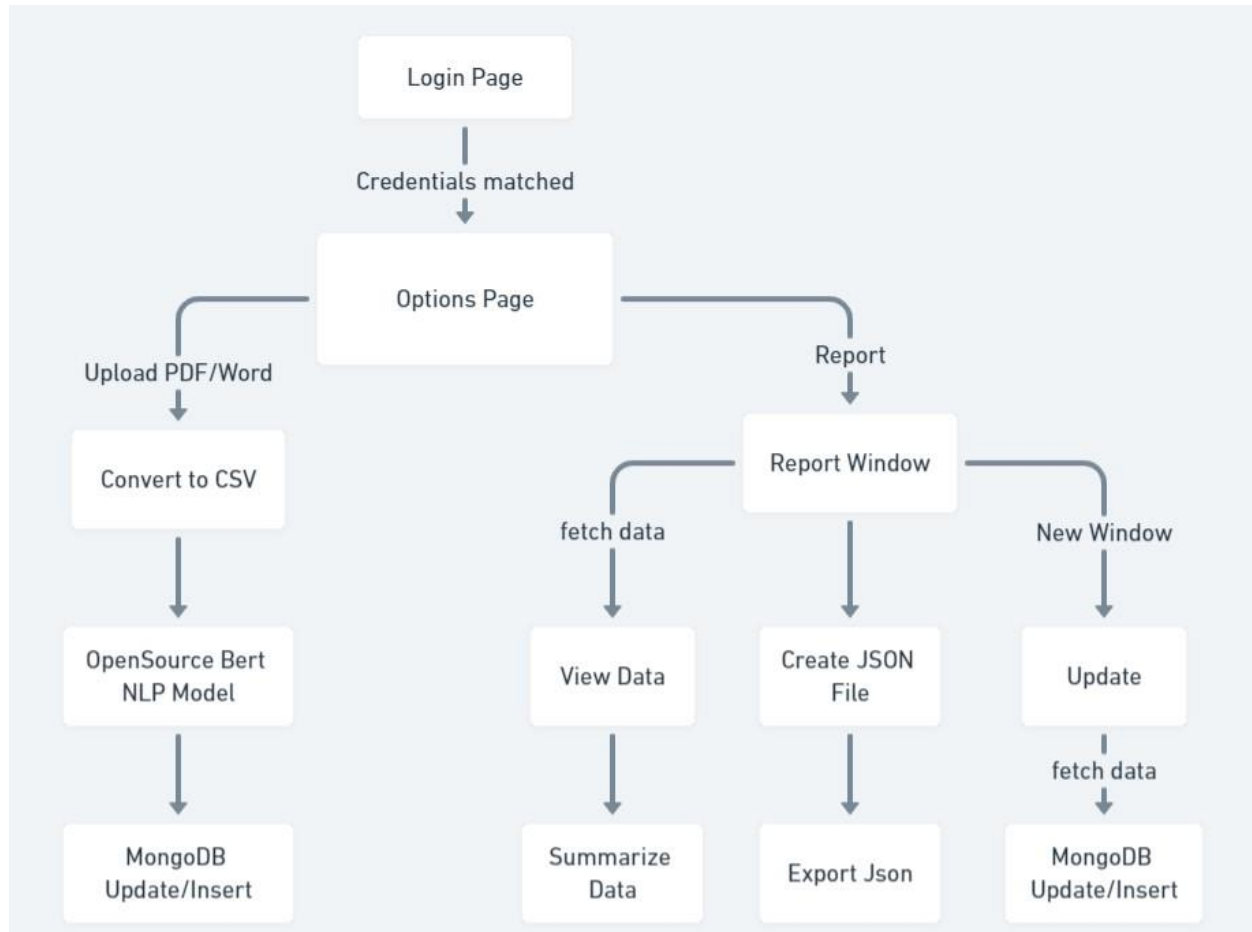
Serialize the data and save it in JSON format.

Summary Generation:

- Implement a summarization algorithm.

- Summarize the extracted data.

15.Flowchart:

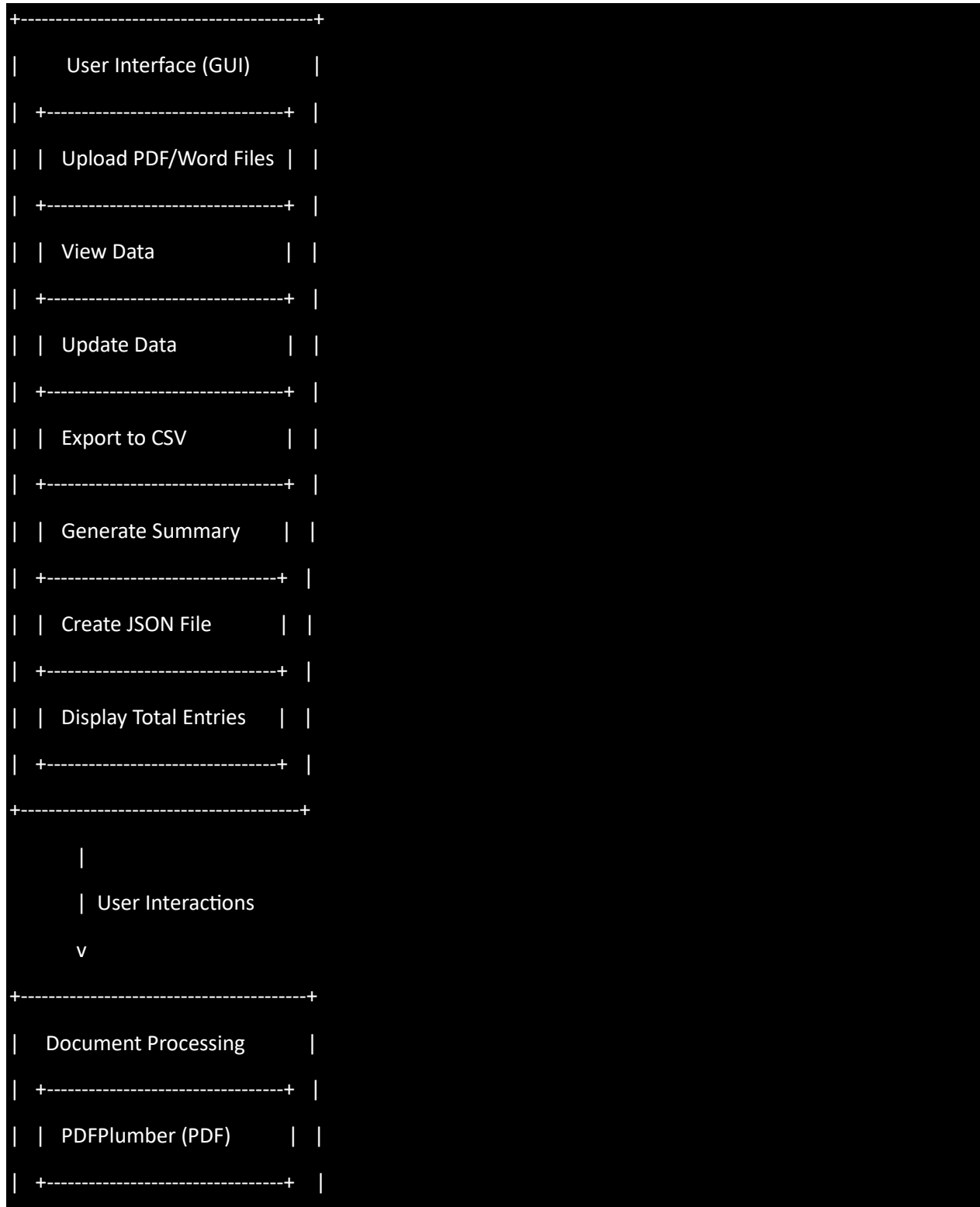


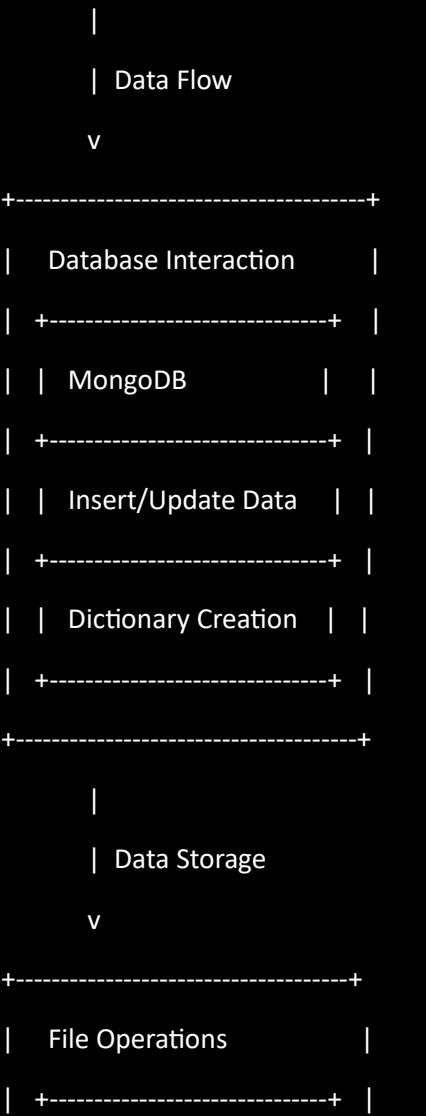
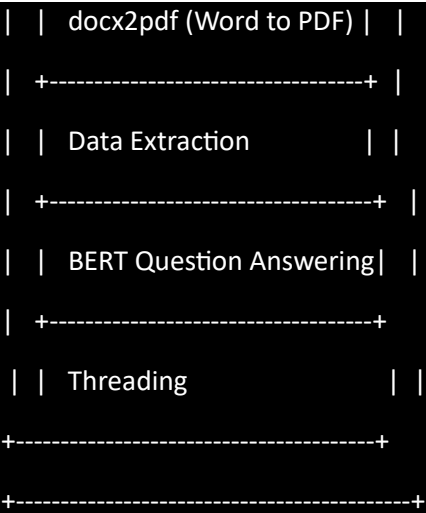
GUI Operations:

Upload, update, view options

Display total entries

15.Architecture:





```
| | Create JSON File | |
| +-----+ |
| | Summary Generation | |
+-----+
```

16.GUI:



Legal Document Processing

Developed By PICT

Upload PDF

Upload Word

Report

Update Database

Select PDF File Name:

What is Name of distributor?

What is the legal name of studio?

What is the title of the document?

From when is the term sheet effective

What are the services offered?

Which geographical areas are included?

Which other territories are included?

What is the end date of the agreement?

What is the delivery method?

Under A La Carte Rates, What is the Service Subscriber charge per month?

Under Packaged Rates, What are the Service Subscriber charges per month?

What is the video file format?

What is the audio format?

What content protection system should be used?

Where will the media be delivered?

When certain titles need to be released or revoked, how will the update take place during the term?

Fetch Data

Update Database

Update Database

Select PDF File Name:

Doc3.pdf

| | |
|---|--|
| What is Name of distributor? | ABC Media Distribution Private Limited |
| What is the legal name of studio? | DEF Broadcast Ltd |
| What is the title of the document? | BINDING TERM SHEET |
| From when is the term sheet effective | July 1, 2019 |
| What are the services offered? | MusicChannel1, in standard definition (SD) |
| Which geographical areas are included? | the United States and its commonwealths, territories |
| Which other territories are included? | Puerto Rico and the U.S. Virgin Islands |
| What is the end date of the agreement? | 30th July, 2021 |
| What is the delivery method? | Hi-Tech Distribution System |
| Under A La Carte Rates, What is the Service Subscriber charge per month? | US\$7.495 |
| Under Packaged Rates, What are the Service Subscriber charges per month? | US \$10.0 |
| What is the video file format? | digitally compressed |
| What is the audio format? | digitally compressed mode |
| What content protection system should be used? | fully encrypted |
| Where will the media be delivered? | Hi-Tech Distribution System |
| When certain titles need to be released or revoked, how will the update take place during the term? | availability of commercial advertising time for Affiliate in |

Fetch Data

Update Database

Legal Document Processing

Developed By PICT

Select PDF File Name:

Doc2.pdf

Fetch Data

Queries

What is Name of distributor?
What is the legal name of studio?
What is the title of the document?
From when is the term sheet effective
What are the services offered?
Which geographical areas are included?
Which other territories are included?
What is the end date of the agreement?
What is the delivery method?
Under A La Carte Rates, What is the Service Subscriber charge per month?
Under Packaged Rates, What are the Service Subscriber charges per month?

Data

ABC Media Distribution Private Limited
DEF Broadcast Ltd
BINDING TERM SHEET
July 1, 2019
MusicChannel1, in standard definition (SD)
the United States and its commonwealths, territories
Puerto Rico and the U.S. Virgin Islands
30th July, 2021
Hi-Tech Distribution System
US\$7.495
US \$10.0

Total Entries: 4

Name of distributor is ABC Media Distribution Private Limited.
Legal Name of Studio is DEF Broadcast Ltd.
Title of Document is BINDING TERM SHEET.
Term of sheet is effective from July 1, 2019.
Services offered are MusicChannel1, in standard definition (SD).
Geographical areas included are the United States and its commonwealths, territories.
Territories included are Puerto Rico and the U.S. Virgin Islands.

Get JSON

Summary

Update