



Multi-Language Document Analyzer

Upload property documents (PDF, TXT).

- PDFs are first checked for parsable text. If found, it's used directly.
- If no parsable text is found, OCR is used with Tesseract (Languages: `eng+hin+tam+tel+kan+mal+mar+pan+urd+guj+ben`).
- Text is analyzed for language. If the detected language is **not English**, it's translated to English. Otherwise, the original English text is used.
- Tables are extracted using Tabula and formatted as **plain text** (space-aligned) for the AI.
- Regex is used to extract boundary information. This information is saved to a text file and used to inform the AI.
- A semantic comparison of boundaries (North, South, East, West) between document pairs is performed using a Sentence Transformer model, with results displayed in a color-coded table.
- Documents are chunked and indexed in Milvus for semantic search (used for Q&A).
- The AI performs a comprehensive comparison of documents, generating a table with key features (Boundaries, Owner, Address, Area) and a conclusion.
- You can ask follow-up questions about the analysis.

Choose document files (PDF, TXT) - Minimum 1 for processing, 2 for comparison



Drag and drop files here
Limit 200MB per file • PDF, TXT

Browse files



Technical report (2).pdf 9.1MB



Legal report (3).pdf 461.2KB



Analysis Results & Chat



AI Document Analysis / Comparison

`http://localhost:11434/api/generat`
`e`

Vector Database (Milvus)

Milvus Host

localhost

Milvus Port

19530

Test & Connect Milvus

Milvus Connected: localhost:19530

Translation & Identification AI

Uses models from

`translation_with_ollama.py` .

Tool Setup Info

Tesseract: `/opt/homebrew/bin/tesseract` OK

Tesseract Langs (for OCR):

`eng+hin+tam+tel+kan+mal+mar+pan+urd+guj+ben`

Poppler: `/opt/homebrew/bin` OK

Okay, I'm ready to perform the analysis and generate the output table.

Output Table (Markdown):

| Feature | Legal report (3).pdf | Technical report (2).pdf |
|------------------|--|--|
| North Boundary | House belongs to Sidda Reddy, House belongs to K | HOUSE OF SIDDA REDDY HOUSE OF SIDDA REDDY |
| South Boundary | House belongs to K | HOUSE OF K.SUBRAMANYAM REDDYHOUSE OF K.SUBRAMANYAM REDDY |
| East Boundary | Land belongs to Sidda Reddy | LAND OF SIDDA REDDY LAND OF SIDDA REDDY |
| West Boundary | Cement Road | 25 FEET WIDE ROAD CEMENT ROAD |
| Owner Name(s) | K.Rupa Kumar Reddy, K.Rama Chandra Reddy | K RUPA KUMAR REDDY |
| Property Address | SY NO 291-4 D/NO 3-12 THIMMAREDDY PALLI ,PAPIREDDY PALLI NANDANOOR, NANDANURU ,Chittoor ,Andhra Pradesh 517125 | SY NO 291-4 D/NO 3-12 THIMMAREDDY PALLI ,PAPIREDDY PALLI NANDANOOR, NANDANURU ,Chittoor ,Andhra Pradesh 517125 |
| Dimensions/Area | Not Found | 3577 SQ.FT |

Final Conclusion:

Key Similarities: The documents share significant overlap in the boundary descriptions (North, South, East, West), indicating a common understanding of the property's location. The owner's name is consistent across both documents. The property address is also largely the same.

Key Discrepancies: The most notable discrepancy is the area information. The 'Technical report (2).pdf' provides a specific area of 3577 SQ.FT, while the 'Legal report (3).pdf' states "Not Found" for this field. The West Boundary description also differs slightly between the two documents.

Overall Assessment: Likely Different. While the documents share a core understanding of the property's location and ownership, the differing area measurements and slight variations in the boundary descriptions suggest that they represent slightly different interpretations or measurements of the same property. The 'Technical report (2).pdf' appears to be providing a more detailed and specific description of the property's dimensions. Further investigation would be needed to determine the precise source of the area measurement discrepancy.



Semantic Boundary Comparison (SBERT)

Comparison between **Legal report (3).pdf** and **Technical report (2).pdf**:

| Feature | Legal report (3).pdf Value | Technical report (2).pdf Value | Match % |
|----------------|------------------------------|---|---------|
| North Boundary | House belongs to Sidda Reddy | HOUSE OF SIDDA REDDY HOUSE OF SIDDA REDDY | 93% |
| South Boundary | House belongs to K | HOUSE OF K | 79% |
| East Boundary | Land belongs to Sidda Reddy | LAND OF SIDDA REDDY LAND OF SIDDA REDDY | 93% |
| West Boundary | Cement Road | 25 FEET WIDE ROAD CEMENT ROAD | 79% |



Download Processed Files

Download Regex-Extracted Boundaries (.txt)

Download final English text for individual documents:

DL Legal report (3).pdf.. (Text)

DL Technical report (2).. (Text)

 View Full Processing Details & Extracted Data (per document)



Chat About the Analysis

Ask gemma3:4b about the comparison...

