

Project Assignment: PDF Transaction Extraction & Web UI

Deadline: 10 days from project kick-off. Please submit your GitHub repo link by end of day.

Objective

Build a full-stack solution to ingest, parse, translate and store 30 years of Tamil real-estate transactions from a PDF into a PostgreSQL database, and present them in a searchable web UI.

Scope of Work

1. PDF Ingestion & Parsing

- Accept a Tamil-language PDF containing real-estate transactions for a single survey number/district.
- Extract each transaction's fields (e.g. buyer, seller, house no., survey no., document no., date, value).

2. Translation

- Translate Tamil fields to English, preserving accuracy of names and document references.

3. Filtering & Search

- Support query parameters: buyer name, seller name, house number, survey number, document number.
- Return only matching transactions in JSON.

4. API Development

- Build a **POST** endpoint using **Node.js**, **Nest.js** and **Drizzle ORM**:
 - Accepts PDF upload
 - Parses, translates, filters, and writes transactions into **PostgreSQL**
 - Returns inserted records as JSON.

5. Web UI

- Use **Next.js** + **Tailwind CSS** to build a simple interface:
 - **Login** screen (stub authentication is acceptable)
 - **PDF upload** form
 - **Results table** showing inserted transactions
 - **PDF preview** panel alongside the table for side-by-side comparison.

6. Libraries & Tools

- You may use any **free, open-source** libraries
 - Store secrets/config in environment variables.
-

Deliverables

- **GitHub repository** containing all source code.
 - **README.md** detailing:
 - Setup & run instructions
 - Assumptions made (e.g., PDF layout variations)
 - Architectural overview & chosen libraries
 - **Database schema** (Drizzle migrations or SQL).
 - **Demo credentials** (for UI login stub).
-

Evaluation Criteria

- **Correctness:** Accurate extraction and translation of all transactions.
 - **Code Quality:** Clean architecture, modular code, meaningful variable names, comments.
 - **User Experience:** Intuitive UI for upload, preview, and results.
 - **Performance:** Reasonable processing time for a 30-year PDF.
 - **Documentation:** Clear README, schema diagrams, and setup steps.
-