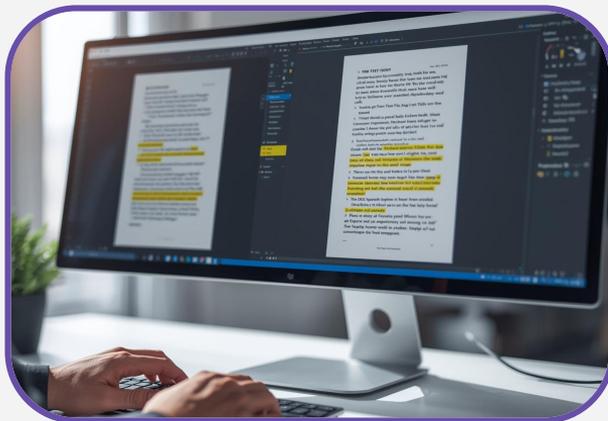


The Challenges:

- Large volumes of unstructured documents (PDFs, images, scanned files) made text inaccessible.
- Manual extraction was time-consuming and error-prone
- Lack of structured outputs limited usability for reporting and analysis.

Our Solutions:

- Implemented Google Cloud Vision OCR to automatically extract text from images and scanned PDFs.
- Integrated Large Language Models (LLMs) to refine extraction, ensuring contextual accuracy.
- Engaged linguistic expert in a manual QA process to validate the accuracy, consistency, and quality of the content.



Any format → Any format



The Results:

- Digitized historical texts to preserve and ensure long-term accessibility.
- Delivered clean, structured outputs ready for analysis, editing, or reporting.
- Reduced manual workload while improving accuracy and efficiency.
- Ensured consistent extraction quality across diverse languages and formats.

Want to Turn Complex Files into Structured Data?

 reach@lexdatalabs.com