

De-Risking Land Access for Infrastructure

World Bank Project - Group 36

Samiksha Ramnath Burkul

Shreya Sahay

Tushar Sharma

An AI-based solution for extracting
and quantifying land access issues



Project Overview

Project Background

- World Bank projects across the globe in nearly every continent face weak land administration and poor registry coverage.
- Tenure insecurity and administrative gaps drive delays, cost overruns, and cancellations.

Obstacles & Impacts

- **Hurdles :** Weak land Administration, Incomplete Land Records, Legal Complexities
- **Consequences:** Project delays, cost overruns, design modifications

Project Scope

- **Coverage:** Restructuring papers from World Bank projects
- **Solution:** AI-based solution for extracting information based on user query



Data Overview

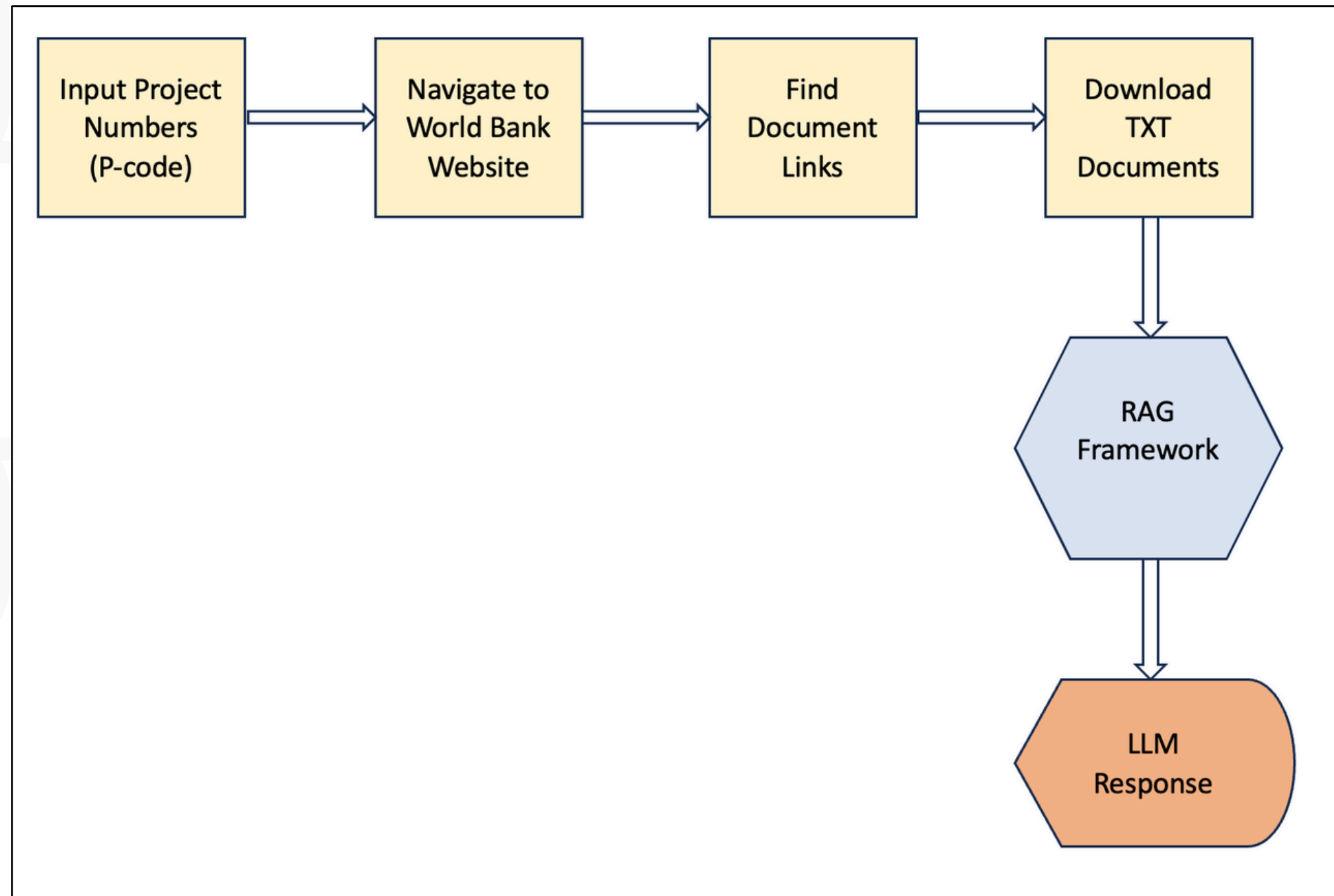
- **Public data source:** World Bank Documents and Reports website
- **Lookup via project p-codes** (e.g., P164572)
- **Total projects:** 117
- **Restructuring Papers:** Details approved changes to the original project design and implementation
- Each project contains a varying number of restructuring papers



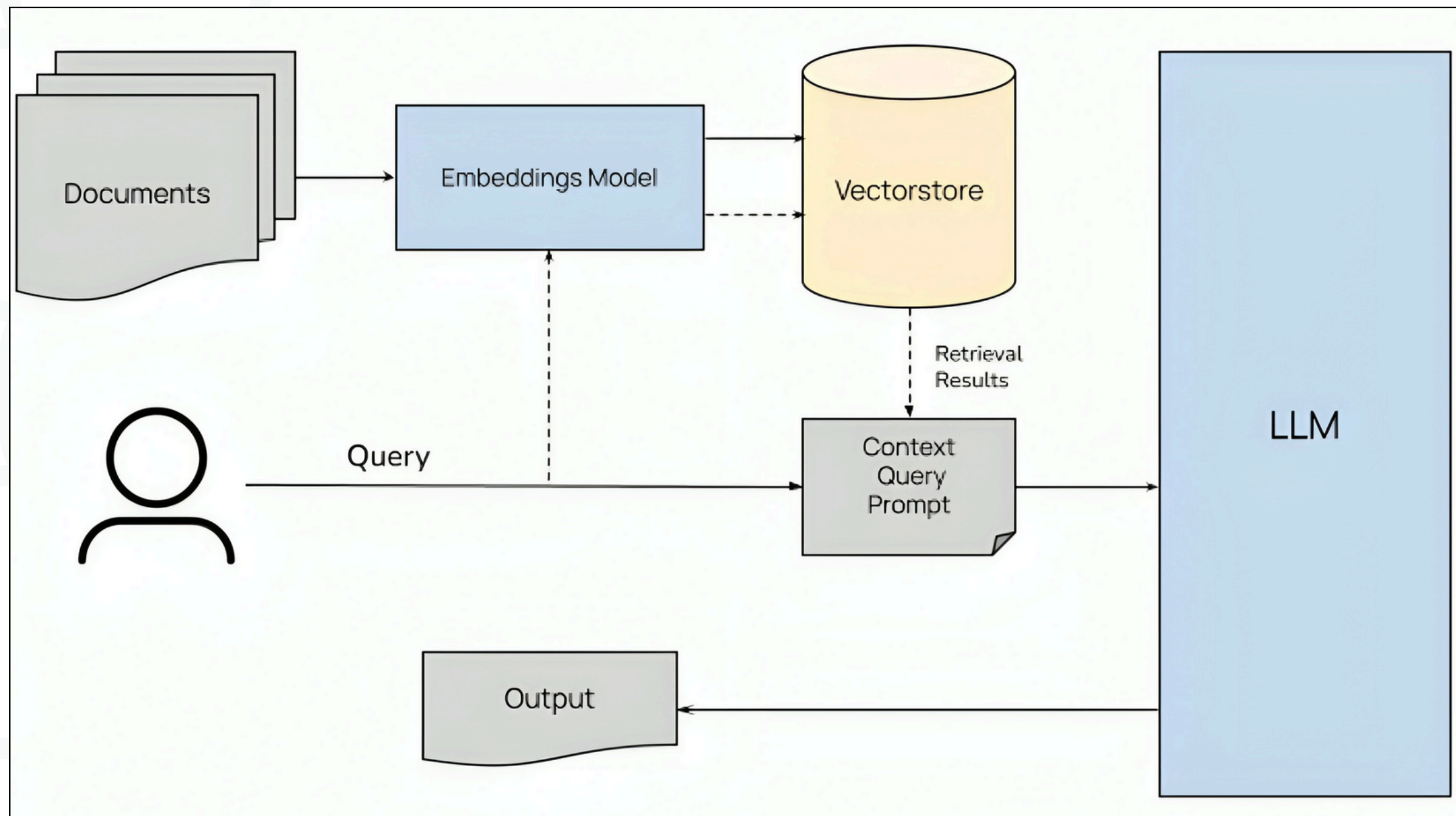
Challenges

- **Massive volume:** Multiple pages made spotting the issue(s) impractical
- **Unstructured data:** Key details buried in free-form text and varied layouts
- **Scattered context:** Insights spread across sections, requiring tedious cross-referencing
- **Inconsistent capture:** Manual quotes and locations varied by reviewer
- **Slow updates:** Every new document batch forced a full repeat of the process

Proposed solution



- First four modules constitute an **automated data pipeline**
- Extracted documents are input in the **Retrieval Augmented Generation** framework
- RAG is advantageous in providing context-enriched and grounded responses



Embedding model	nomic-embed-text
------------------------	------------------

Vector store	ChromaDB
---------------------	----------

LLM	Llama 3.1 8B
------------	--------------

Documents are split into chunks using **semantic chunking** followed by **vector search** which compares the query and chunks' embeddings

RAG architecture

Results

Answered with RAG: There are no issues or investments identified in either document that specifically mention land access or acquisition as an issue. The documents appear to be related to loan proceedings and reallocation of funds for various components of a project, including infrastructure investments such as roads, water supply networks, and public facilities.



Example project with no land access issue

Answered with RAG: After reviewing the provided document, I found an issue related to land access/acquisition.

****Issue:**** The World Bank's Senegal Municipal Solid Waste Management Project (P161477) faced challenges in acquiring land for a waste management facility due to difficulties in negotiating with local authorities and communities.

****One-line summary of the issue:****

Land acquisition issues hindered project implementation.

****Location of the exact quote(s):****

Unfortunately, there is no direct quote from the document that explicitly mentions "land access/acquisition" as an issue. However, on page 8 (DocType: RP), it is mentioned:

"...the Project has faced challenges in acquiring land for a waste management facility due to difficulties in negotiating with local authorities and communities."

This implies that land acquisition was indeed an issue, but the exact quote does not explicitly mention "land access/acquisition".



Example project with a land access issue

Insights from RAG framework

Projects with land access issues 49

Projects with no land access issues 68





Impact: Achieved objectives

Faster Insight Discovery

Instant identification and summary of relevant user query-based information

Consistent, Evidence-based Findings

The framework provides verifiable source quotes

Better Decision Support

Faster insight discovery and quantitative figures can help with project prioritization and make informed decisions

Adaptable & Robust

Modular design allows input any data and query any relevant information without requiring architectural changes



Areas of Improvement

Limitations

- Framework is computationally expensive
- Results sensitive to the user prompt/query phrasing
- Occasional “hallucinations” (inaccuracies) in responses

Future Work

- Build a user-friendly interface (chatbot/website/dashboard)
- Fine-tune the language model to boost response accuracy
- Add GIS analysis to map and monitor land-access challenge areas