

# GRCResponder

Brianna Steier, Liam Gass, Elijah Tavares, Cael Howard,  
June Kim, Rish Sharma, Angel Li

# Table of Contents

---

1 | Search Engine Updates

3 | LLM Updates

2 | Web Development Updates

4 | Questions

# Search Engine Updates

---

- Current Focus: Planning web scraper and defining high-level functionality
- Filtering out low information documents to decide which to scrape, process, and store
  - low information: notices, ex parte, comments, law & motion, transcripts

# Search Engine Updates

---

- Web Scraper Methods
  1. Scrape everything and save corresponding documents in directories for each case according to proceeding number
  2. Filter out unnecessary documents and only save high information files/metadata
  3. CPUC has csvs that contain links all the corresponding documents for each case
    - a. Could then filter out unnecessary documents by type and/or content while still having access to their location

# Search Engine Updates

- All filings are linked to a proceeding
- Serves as a unique identifier for each case

Proceeding Detail	
<a href="#">Subscribe</a>	<a href="#">Search Results</a> <a href="#">Main Page</a>
<b>K2210009 - Proceeding</b>	
<b>Filed By:</b>	Central Coast Community Energy
<b>Service Lists:</b>	<a href="#">Service List 1</a>
<b>Industry:</b>	Electric
<b>Filing Date:</b>	October 17, 2022
<b>Category:</b>	
<b>Current Status:</b>	CLOSED
<b>Description:</b>	Appeal of Central Coast Community Energy from Citation E-4195-0126 issued on September 16, 2022 in the amount of \$25,000 issued by the California Public Utilities Commission Consumer Protection and Enforcement Division
<b>Staff:</b>	ALJ: Debbie Chiv (Assigned Nov 4, 2022) COMMISSIONER: (Assigned Nov 4, 2022)

# Search Engine Updates

```
CREATE TABLE proceedings (  
  id SERIAL PRIMARY KEY,  
  proceeding_number VARCHAR(50) UNIQUE NOT NULL, -- "K2210009 "  
  filed_by TEXT, -- "Central Coast Community Energy"  
  service_lists TEXT, -- JSON  
  industry VARCHAR(100), -- "Electric"  
  filing_date DATE, -- October 17, 2022  
  category VARCHAR(100), -- ""  
  current_status VARCHAR(100), -- "CLOSED"  
  description TEXT, -- "Appeal of Central Coast Community Energy from Citation E-4195-0126..."  
  staff TEXT, -- e.g. "ALJ: Debbie Chiv (Assigned Nov 4, 2022) COMMISSIONER: (Assigned Nov 4, 2022)"  
  created_at TIMESTAMP DEFAULT NOW(), -- proceedings will have new documents added over time  
  updated_at TIMESTAMP DEFAULT NOW() -- this will help scheduler know when new information is found  
);
```

# Search Engine Updates

Search by proceeding id yields all related documents

<a href="#">Title</a>	<a href="#">Doc Type</a>	<a href="#">Doc Links</a>	<a href="#">Published Date</a> ↑
D2409023 DECISION DISMISSES HOTWIRE NETWORK I LLC'S APPLICATION FOR A REGISTRATION LICENSE Proceeding: A2407022	Final Decision	<a href="#">PDF</a> (84 KB) <a href="#">WORD</a> (21 KB)	09/24/2024
Resolution ALJ-176 Categorization filed by CALJI/COOKE/CPUC on 08/22/2024 Conf# 210260 Proceeding: A2407016; A2407017; A2407018; A2407019; A2407020; A2407021; A2407022; A2408002; A2408003; A2408004; A2408007	E-Filed: Resolution ALJ-176 Categorization	<a href="#">PDF</a> (177 KB)	08/23/2024
Motion filed by Hotwire Network I LLC on 08/22/2024 Conf# 210247 (Certificate Of Service) Proceeding: A2407022	E-Filed: Motion	<a href="#">PDF</a> (243 KB)	08/23/2024
Motion filed by Hotwire Network I LLC on 08/22/2024 Conf# 210247 Proceeding: A2407022	E-Filed: Motion	<a href="#">PDF</a> (230 KB)	08/23/2024
Resolution ALJ 176-3550 Proceeding: A2407016; A2407017; A2407018; A2407019; A2407020; A2407021; A2407022; A2408002; A2408003; A2408004; A2408007	Agenda Resolution	<a href="#">PDF</a> (148 KB) <a href="#">WORD</a> (27 KB)	08/21/2024
Application filed by Hotwire Network I LLC on 07/31/2024 Conf# 209481 Proceeding: A2407022	E-Filed: Application	<a href="#">PDF</a> (7356 KB)	08/01/2024

# Search Engine Updates

- Document table has a foreign key referencing corresponding proceeding
- Suppose three months later, a new document gets filed for that same proceeding - just insert a new document record

```
CREATE TABLE documents (  
  id SERIAL PRIMARY KEY,  
  proceeding_id INT NOT NULL,  
  doc_type VARCHAR(100) NOT NULL,      -- "APPLICATION", "MOTION", "DECISION", "RULING"  
  filing_date DATE,                    -- "August 22, 2024 "  
  filed_by TEXT,                       -- "Hotwire Network I LLC"  
  description TEXT,                   -- "MOTION TO WITHDRAW APPLICATION FOR REGISTRATION AS AN INTEREXCHANGE..."  
  pdf_url TEXT,                       -- generated from CPUC - "https://docs.cpuc.ca.gov/PublishedDocs/Efile/6000/M538/K691/538691839.PDF"  
  doc_text TEXT,                      -- for key word search  
  
  CONSTRAINT fk_proceeding  
    FOREIGN KEY (proceeding_id)  
    REFERENCES proceedings(id)  
    ON DELETE CASCADE  
);
```



# Design Decision - Chunks Stored in Postgres

## A: Chunking

### Pros

- Concise search results
- Improved imbedding context

### Cons

- Storage/performance overhead

## B: No Chunking

### Pros

- Easy implementation and low overhead on insertion
- Keyword search still achievable

### Cons

- Vague results
- Performance sacrifices
- Potential accuracy issues

# Search Engine Updates

Sentence transformer -

[All-mpnet-base-v2](#)

- Well vetted model
- Reported to perform well with technical legal jargon

[nlpueb/legal-bert-base-uncased](#)

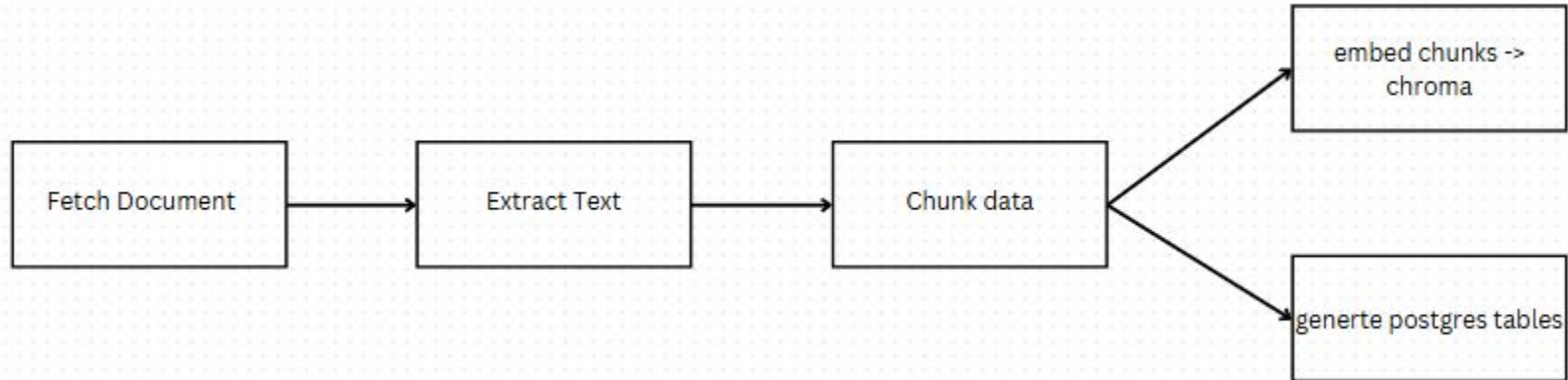
- Alternative model we could look into
- trained solely on legal documentation:

I will generate test queries and run cosine similarity to evaluate effectiveness

- Find a broad selection of example docs ~10
- Create meaningful queries for comparisons

# Search Engine Updates

## Current Ingestion pipeline



# Web Dev Updates

---

---

- CI/CD + Environment Setup ✓
- Page Outline ✓
  - Home page = GRCResponder Chat Assistant
  - (optional) document import
  - (optional) document/chat export

# LLM Updates

---

- Environment setup
  - LangChain, Chroma, Ollama
- Initial prototype of RAG system

---

# Questions?

---