

Discover the power of semantic search in
government press releases with ease

SemanticPress

Presentation

Problem

Government agencies such as the Department of Justice regularly release press statements to inform the public about their activities and decisions. These press releases contain important information that can be difficult to find and understand due to their volume and complexity.


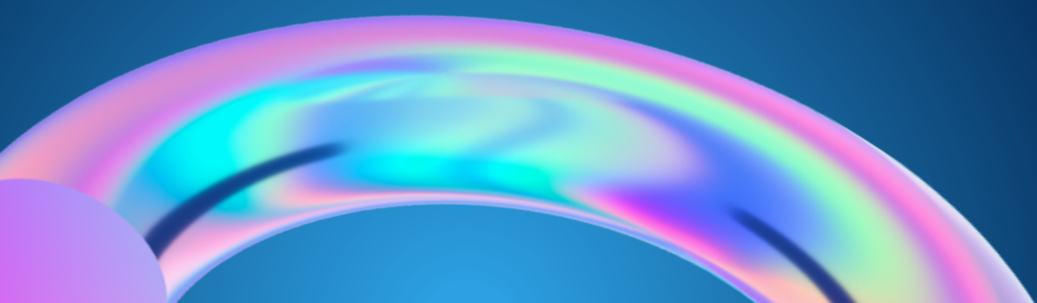
Currently, there is no easy way for individuals to search through these press releases and find relevant information. Traditional keyword-based search engines often return irrelevant or outdated results, making it difficult for users to find what they're looking for.



Solution

To address this problem, we propose the development of an app that provides semantic search results for queries on press releases from the government of justice.

This app will use advanced natural language processing techniques to understand the meaning behind user queries and provide accurate and relevant search results.



Tech stack

The logo for Cohere, featuring the text "co:here" in a white, lowercase, serif font on a black rectangular background.

Cohere
Cohere Embed model

The logo for Qdrant, featuring a red and white 3D cube icon to the left of the word "drant" in red lowercase letters. Below it, the text "Vector Search Engine" is written in white lowercase letters on a dark grey background.

drant
Vector Search Engine

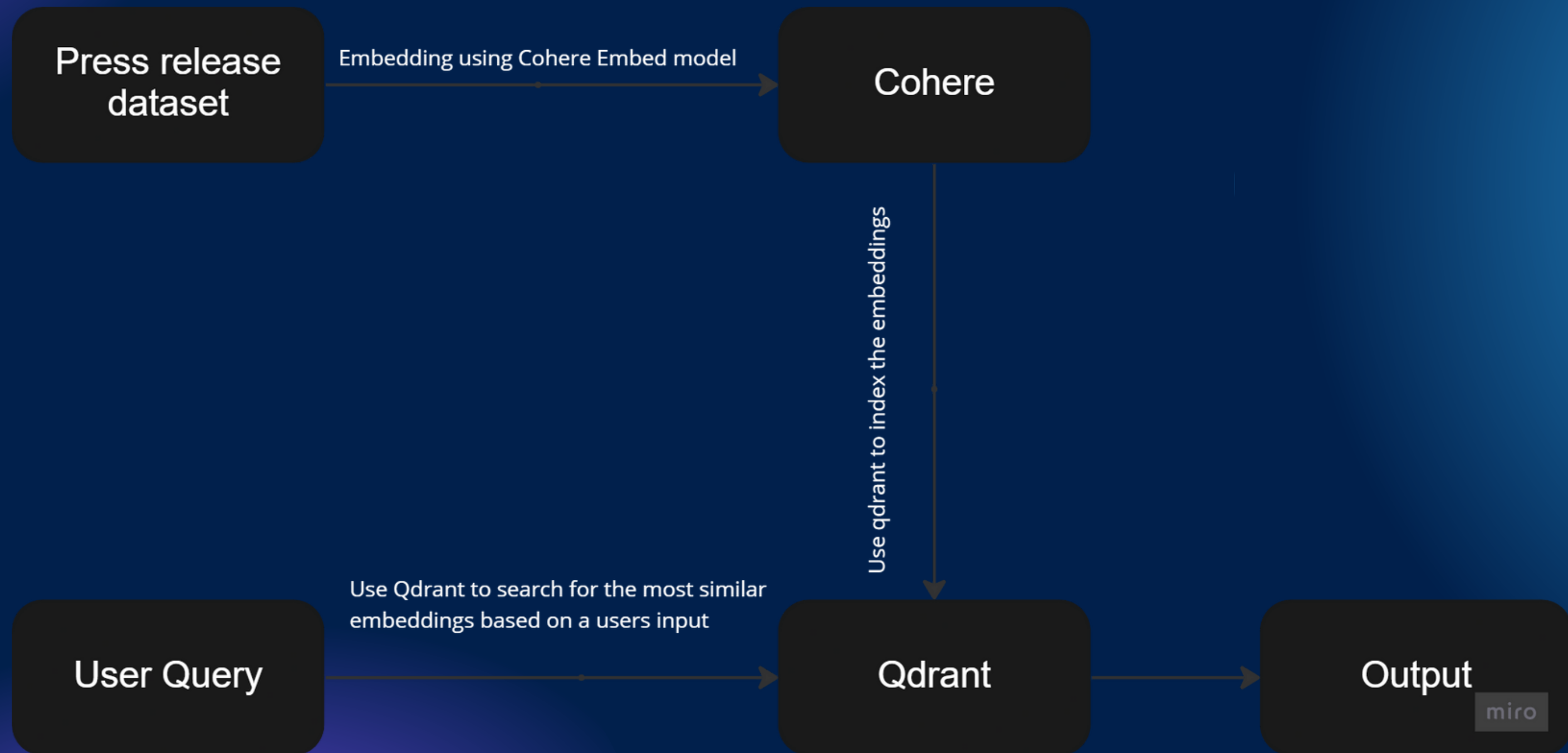
Qdrant
Qdrant Vector Database

The logo for Streamlit, featuring a white crown icon above the word "Streamlit" in a white, sans-serif font.

Streamlit

Streamlit
Web App

Workflow



Output

The data is from Department of Justice 2009-2018 Press Releases - [Kaggle](#).

Context

Content

The contents are stored as newline delimited JSON records with the following fields:

- id: Press release number (can be missing if included in contents)
- title: Title of release
- contents: Text of release
- date: Posted date
- topics: Array of topic tags (if any provided)
- components: Array of agencies & departments (if any provided)' The dataset has been modified and reduces to 100 titles.

Suggested search

SemanticPress

Discover the power of semantic search in government press releases with ease - Powered by Qdrant x Cohere

Enter your query

Racketeer Influenced and Corrupt Organizations Act (RICO)

Press release - 1

```
{id='6bc89cf4-8de5-4253-aed2-b5682b0be759' version=0 score=0.53700787 payload={'contents': 'A federal indictment was unsealed today in the District of Puerto Rico charging 50 members of La Asociación ÑETA with racketeering, drug trafficking and murder, announced U.S. Attorney Rosa Emilia Rodríguez-Vélez for the District of Puerto Rico. \xa0The defendants include the leadership of the enterprise, its drug trafficking network and members who perpetrated murders in furtherance of the enterprise. \xa0This prosecution is the product of an FBI investigation into the drug trafficking and violent activities perpetrated by members of the gang known as La Asociación ÑETA. According to the indictment, the defendants are charged with being part of La Asociación ÑETA, a prison gang that operates in the prisons of the Puerto Rico Department of Corrections and Rehabilitation (PRDCR).\xa0 La Asociación ÑETA is a criminal organization that engages in drug trafficking and murder. \xa0The main purpose of the organization is to make money. \xa0Inmates originally formed La Asociación ÑETA as a means to advocate for their rights within the PRDCR. \xa0The enterprise, however, evolved into a criminal organization that engaged in drug trafficking and murder within the prisons of the PRDCR. \xa0The enterprise makes money by introducing multi-kilogram quantities of drugs into the PRDCR prisons for profit and by engaging in murders for hire. The indictment
```

Team



Aadarsh K

Student

@dotAadarsh



Thank You

@dotAadarsh

aadarshkannan.vercel.app