Resume QnA ChatBot

Overview

This script extracts text from a PDF resume, creates a Pinecone index, embeds text using Cohere, and generates answers to queries based on the context retrieved from the index.

Requirements

- Python 3.x
- pinecone-client, pinecone-datasets, pinecone-notebooks, PyPDF2, pdfplumber, and cohere libraries

Functions

extract_text_from_pdf(file_path)

Extracts text from a PDF file using pdfplumber.

- file_path: Path to the PDF file
- Returns: Extracted text as a string

retrieve_documents(query, top_k=5)

Retrieves documents from the Pinecone index based on a query.

- query: Query string
- top_k: Number of documents to retrieve (default: 5)
- Returns: List of retrieved documents

generate_answer(query)

Generates an answer based on the context retrieved from the index.

- query: Query string
- Returns: Answer string

Usage

1. Install required libraries using pip install commands.
2. Set environment variables for Pinecone API key and cloud/region.
3. Run the script, providing the path to the resume PDF file.
4. Execute queries using the queries list.

Code Structure

1. Import libraries and set environment variables.
2. Define functions for text extraction, document retrieval, and answer generation.
3. Authenticate with Pinecone and create an index.
4. Embed text using Cohere and upsert into the Pinecone index.

5. Define queries and execute them using the generate_answer function.

Notes

- Ensure the Pinecone API key and cloud/region environment variables are set.
- Update the queries list to execute different queries.
- Modify the top_k parameter in retrieve_documents to retrieve more or fewer documents.