

Question and Answers Assistant - Design Choices

Document Indexing

The Elasticsearch search engine was a very good option chosen for indexing the document. This decision was based on its robust full-text search capabilities and scalability, aligning with the project's requirements. The indexing process included associating each document with its metadata and generating embeddings using a state-of-the-art language model. Dense vector embeddings were selected to capture intricate semantic relationships between passages, facilitating more accurate retrieval of relevant information.

Query Processing

Query processing involved utilizing the Sentence Transformers library to generate embeddings for both user queries and indexed passages. This approach was essential in computing similarity scores efficiently. Cosine similarity was selected as the similarity metric due to its effectiveness in comparing high-dimensional vectors. This choice was particularly crucial for the embedding-based approach.

Passage Retrieval

The passage retrieval process centered around calculating similarity scores between the user's query embedding and the embeddings of all indexed passages. The passages were then ranked based on their similarity scores, and the top results were returned. By incorporating a threshold for relevance, the system ensures that only the most pertinent passages are presented to the user.

User Interface

The Streamlit framework was utilized for the user interface, aligning with the specified requirements. Streamlit offers rapid development of interactive web applications in Python, making it a suitable choice. The interface encompasses options for users to upload documents for indexing, enter their questions, and view the relevant passages alongside their relevance scores. Additionally, a button-triggered mechanism was integrated to initiate the question-answering process.

Evaluation

The evaluation process followed a structured two-step approach. First, manual ratings were assigned to determine the relevance of passages retrieved for each query. This step was crucial in establishing a ground truth dataset for accuracy assessment. Subsequently, logic was implemented to compute top-1 and top-3 accuracies based on the user ratings. These metrics provided a quantitative assessment of the system's effectiveness.

User-Friendly Interface

The design of the user interface prioritized simplicity and intuitiveness, aligning with the project's requirements. Input fields were clearly presented for users to upload documents and enter queries. The display of retrieved passages was structured for easy comprehension, featuring relevant metadata and relevance scores alongside the content. The interface also facilitated seamless interaction with the system.

These design choices were guided by the specific requirements outlined in the project's initial document. The selection of tools and approaches was directed by the predefined objectives and considerations for optimal performance within the legal domain.