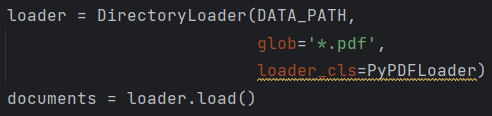
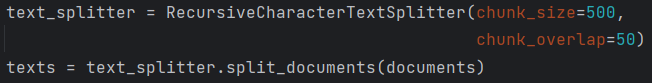
**LLAMA 2:**

**LOADING PDF TO VECTOR DATABASE:**

1. Load the pdf document.



1. Split the text into multiple chunks.



1. Used embeddings from hugging face.



1. Convert them into vector database.



1. Save the database to local directory.



1. To check the similarity of any query in vector database.



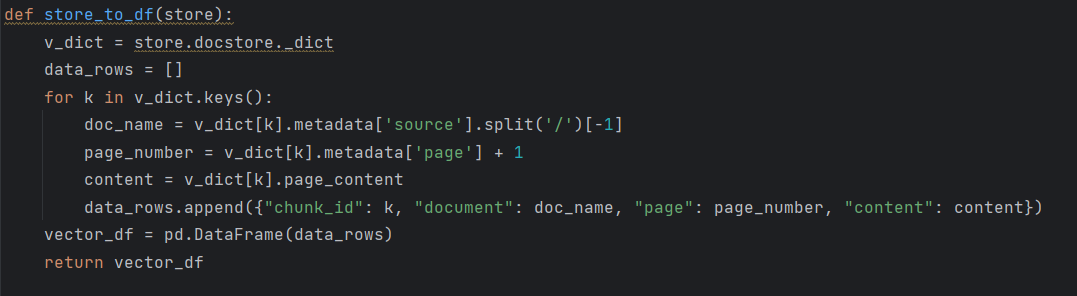
1. To delete any specific chunk id from vector database.



1. To add any new data to vector database.



1. Convert the vector database to data frame so that it’ll be easy to understand.

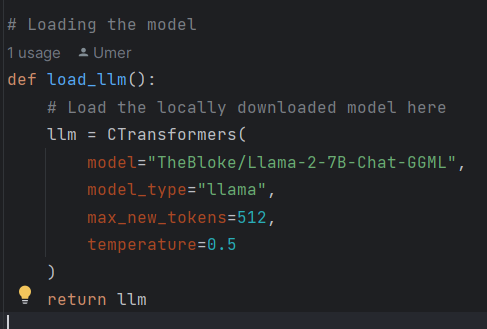


**IMPLEMENTATION OF MODEL:**

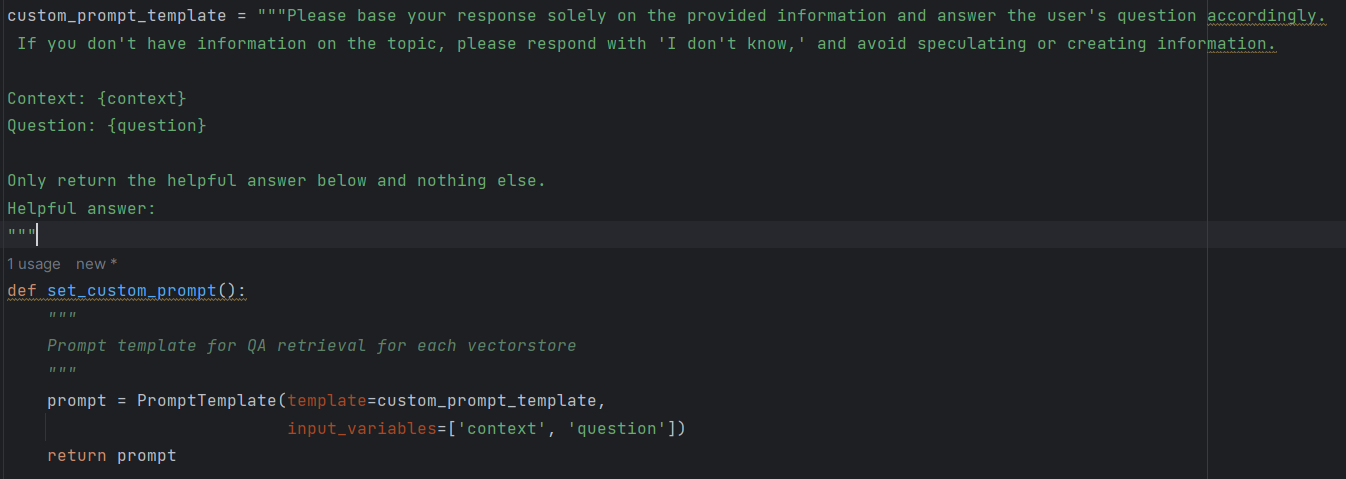
1. Load data from vector database.



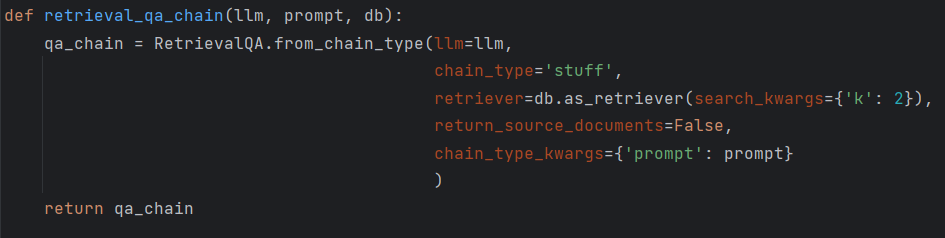
1. Load the llm model locally by using CTransformer:



1. Add the custom prompt so that model respond the query accordingly.



1. Used retrieval qa chain from langchain to make the model able to perform question and answer.



1. To get the response of any query just simply pass the query to qa\_chain.

