Decentralized RAG: Combining Retrieval Augmented Generation with Blockchains

Reed White (12367576)

Github: <https://github.com/mist861/5576-Blockchain>

Panopto Video: <https://umsystem.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=603bf6e5-e1de-4f6d-9577-b2170161bed2>

# Weeks 5-6 Report

The following progress has been made during this sprint (weeks 5-6):

* Proof-of-Concept Smart Contract (DocumentStore.sol)
  + Some defects were cleaned up
* Proof-of-Concept Hash Smart Contract (DocumentHashStore.sol)
  + A new HashCorpus smart contract was created to be an alternative version of the above. Instead of storing entire documents in the blockchain, the blockchain stores only hashes of each document along with a location where it is stored.
* Proof-of-Concept Web3.js Hash Interaction Script (corpus\_hash\_interact.js)
  + This was created to support the above alternative hashing mechanism.
* HTML Corpus Interaction Administrative Frontend (corpus\_server.js/corpus\_index.html)
  + Some defects were cleaned up
* HTML HashCorpus Interaction Administrative Frontend (corpus\_hash\_server.js/corpus\_hash\_index.html)
  + A new frontend was created to support the new HashCorpus smart contract
* RAG Notebook (5576-0001\_project\_rag.ipynb)
  + The notebook was updated with the hashing mechanisms.
* RAG (rag\_blockchain\_streamlit.py)
  + A Streamlit Python script was developed to run the original RAG that pulls documents directly from the Corpus smart contract.
* Hash RAG (rag\_blockchain\_hash\_streamlit.py)
  + An alternative Streamlit Python script was developed to run the alternative RAG that pulls hashes and locations from the HashCorpus smart contract, documents from the locations, calculates a new hash for each pulled document, compares that hash with the original hash from HashCorpus, and loads them into the vector store only if they match.

This project was deployed via Hardhat and Node.js in an Ubuntu system with NPM installed, using the commands located in the README.md file in the project src directory.

Minimal changes have been made to existing code, the majority of work done was on the new, alternate version of the system that uses hashing and off-chain storage. Source code has been sorted into specific hash and non-hash directories, with the exception of the Hardhat smart contracts.

Additionally, work on project documentation has been continued:

* Architecture
  + This describes the connectivity between system modules and various user flows.
* User Manual
  + This describes how users (corpus administrators and end users) interact with the system.
* System Manual
  + This describes how system administrators install and administrate the system.

And, of course, the above video was created.