Enhancing Cyber Threat Analysis

Edalaine Cadiena

CyberSecurity Service Provider Internship Program, ICS 496, Fall 2023

Advisors: Dr. Curtis Arnold, Chief Scientist – Cyber, & Dr. Robert Adamove, Sr. Associate, (Core4ce)

Problem Statement:

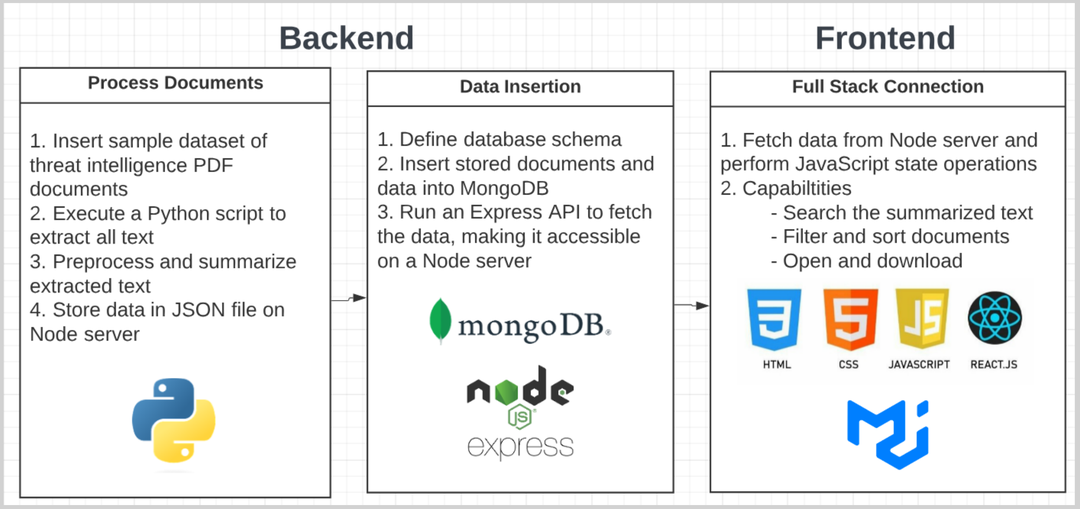
There is a proliferation of various types of documents on the internet that cybersecurity analysts must consume and analyze, which is a very time and energy consuming task. The lack of streamlined information consolidation and efficient analysis of these is what leads cybersecurity analysts to dedicate a lot of hours just parsing through these documents and online content.

Purpose:

The purpose of this study is to identify meaningful threat information from a PDF document, a repetitive and time-intensive task for cybersecurity analysts. There are a lot of PDF parsing and document analysis tools that are widely available for usage, but since NIWC is a Cybersecurity service provider they are limited with the open-source tools they are allowed to safely use with sensitive information (Ritchey, 2024). Timeliness is essential with threat intelligence, so the necessity for an analyst to efficiently analyze data and determine actionability is crucial.

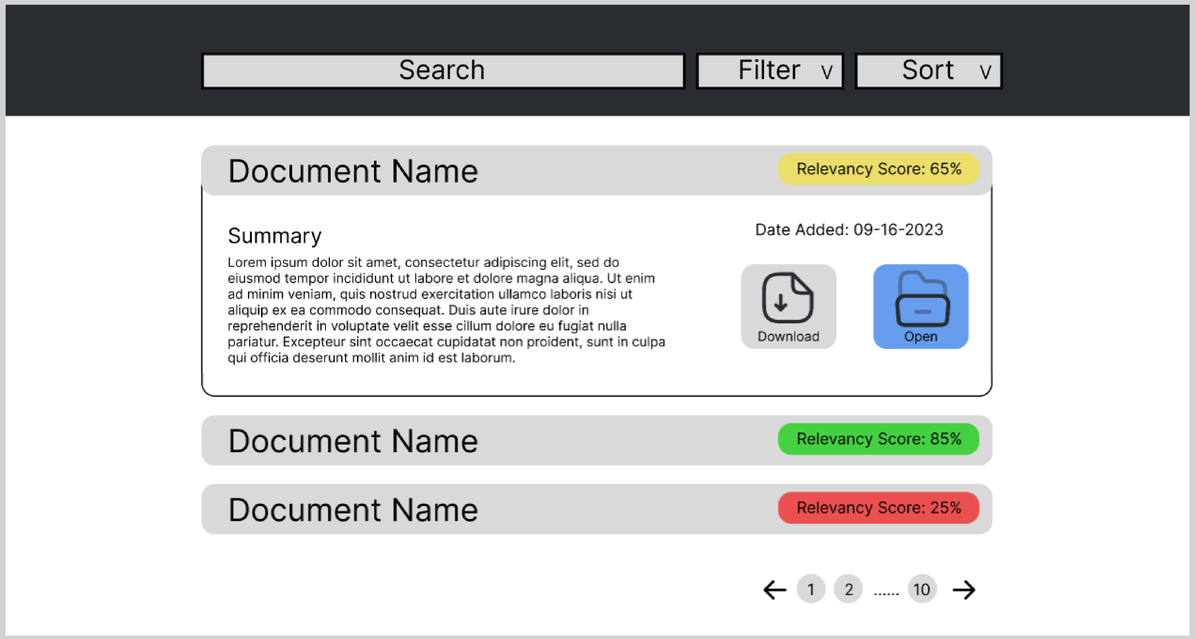
Approach:

This project utilizes the MERN (MongoDB, Express.js, React, Node.js) technology stack due to its versatility and robustness in crafting dynamic web applications. A database schema was designed that seamlessly aligned with the provided data, prioritizing efficient data storage and retrieval while developing a Python script to extract the relevant text information. A server infrastructure was established an efficient transmission of data throughout the application, primarily utilized when adding documents and searching with user input..



Discussion & Conclusion:

While reviewing the documents, challenges in parsing them for meaningful information were encountered. This difficulty arose from the inherent complexity of PDFs, which often contain a multitude of elements, including images, code snippets, and graphs, that cannot be readily converted into text format.



The application features a display of all threat documents along with a relevancy score based off user search input, and summaries generated from Python scripts.

Path Forward:

The current version of the application serves as a minimum viable product exclusively for handling PDF documents. Recognizing the diversity of document types in the real world, the next steps involve expanding and enhancing the software to accommodate various document formats. This evolution will empower the application to effectively process and manage a wide range of document types, delivering a more comprehensive and versatile solution to users.

Acknowledgement:

Paul Ritchey – Cybersecurity Data Engineer at Core4ce