Info Challenge 2025
Team 25063 Abstract

Amtrak's Records and Information Management Program faces significant challenges in managing Freedom of Information Act (FOIA) requests due to the lack of technology enabling the categorization of FOIA requests. This creates a reliance on a single individual's knowledge for identifying appropriate contacts within the organization. To address this bottleneck, a novel solution leveraging Python libraries for computer-based processing classification has been proposed.

This innovative approach aims to create a sophisticated FOIA Knowledge Base that automatically classifies requests based on keywords. By utilizing traditional processing techniques, the system can intelligently match FOIA requests to Amtrak departments and their point of contact. This classification process eliminates the dependency on a single point of contact and streamlines routing requests, to appropriate personnel.

During our analysis, we noticed that the document containing a list of contacts and their respective departments are vague, providing little context. In some instances, different departments are stored within a single table cell, causing trouble when parsing for departments that belong to a particular request. Our solution was to return one point of contact for each department, with the exception of departments with multiple personnel responsible for particulate areas, or regions. We also performed data cleanup on FOIAs that contain the words “Duplicate”, “Not a proper FOIA request”, “unclear”, etc.

Our analysis of the data provided to us by Amtrak revealed that a majority of the data was able to be classified based on a select set of keywords, which helped us identify where each FOIA request belongs to.