MICHIGAN STATE UNIVERSITY

Project Plan Presentation

The Summarizer

The Capstone Experience

Team Auto-Owners

Ian Byram
Jiahui Hu
Jonathan Kippe
Thinh Nguyen
Brylee Pavlik
Tyler White

Department of Computer Science and Engineering
Michigan State University

Spring 2023

From Students…
…to Professionals

Project Sponsor Overview

- Fortune 500 Company
- A++ rating by AM Best Company
- Started in 1916 and now has over 48,000 agents
- Main headquarters in Lansing, Michigan

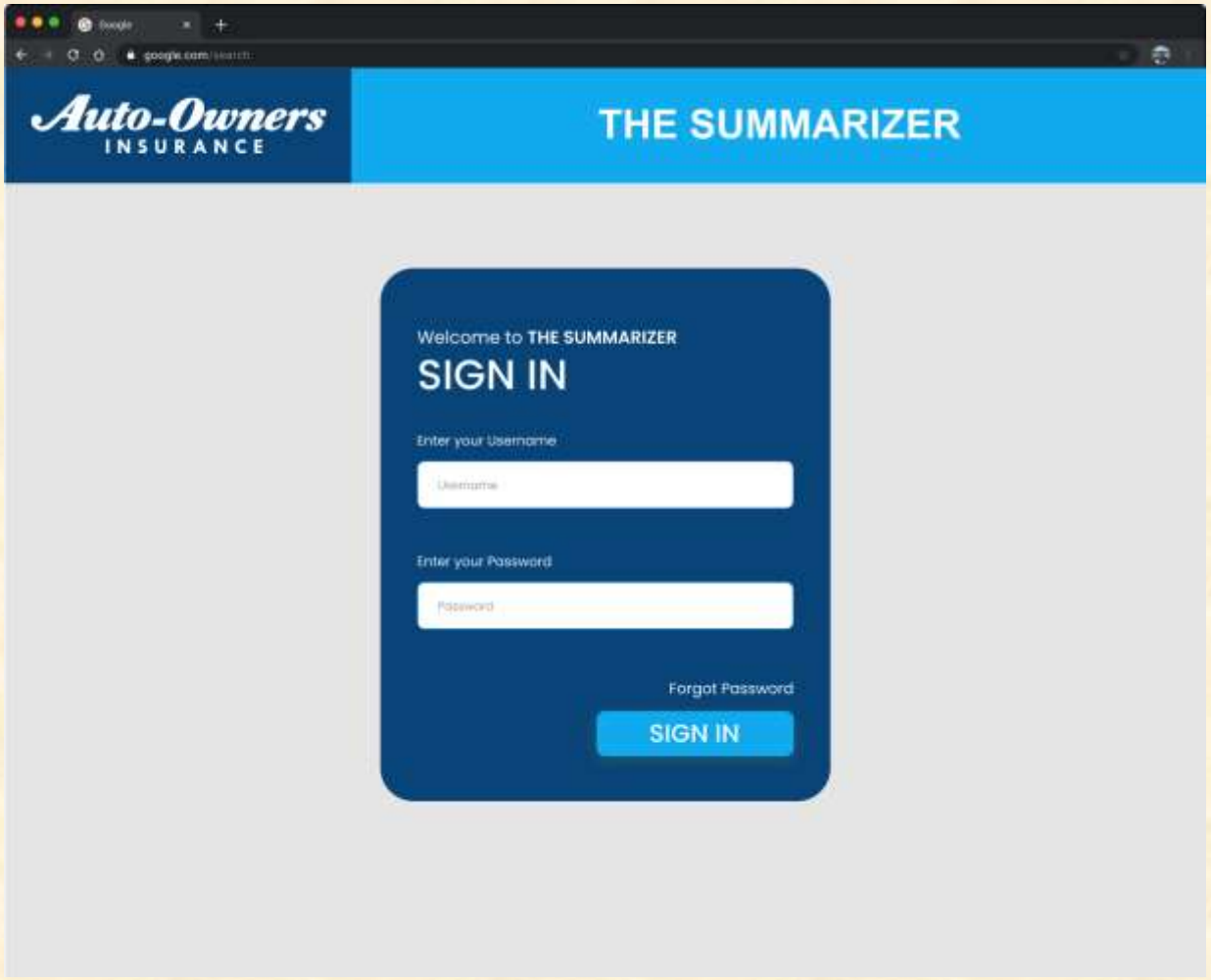
Project Functional Specifications

- Better analysis and decision making for customers' and employees' needs
  - Reviewing all the text and analyzing key metrics from a large amount of documents.
- A time-consuming and labor-intensive task
  - Manually processing; Large text; No standardization
- An automated and more efficient web application
  - Quickly summarize documents, extract relevant key metrics, and analyze the sentiment from them.
- Knowing the positive and negative part of their current service and product
- Log in -> submit file -> check report -(if admin)-> access history reports

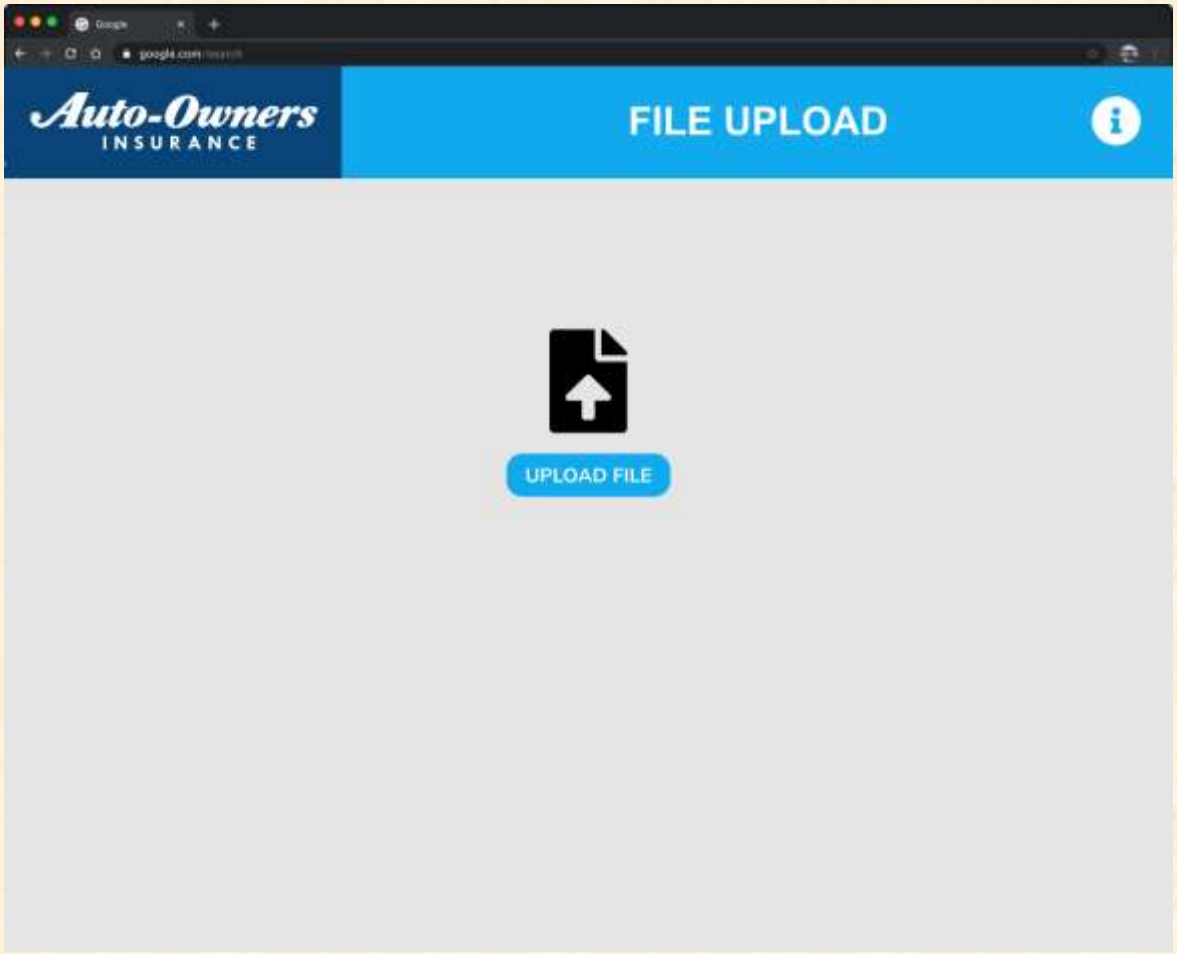
Project Design Specifications

- Major Features:
  - Login page, file upload page, file processing report, text analysis report (with sentiment analysis), summarization report
- Vertical sidebar navigation menu
- Information symbol for project explanation

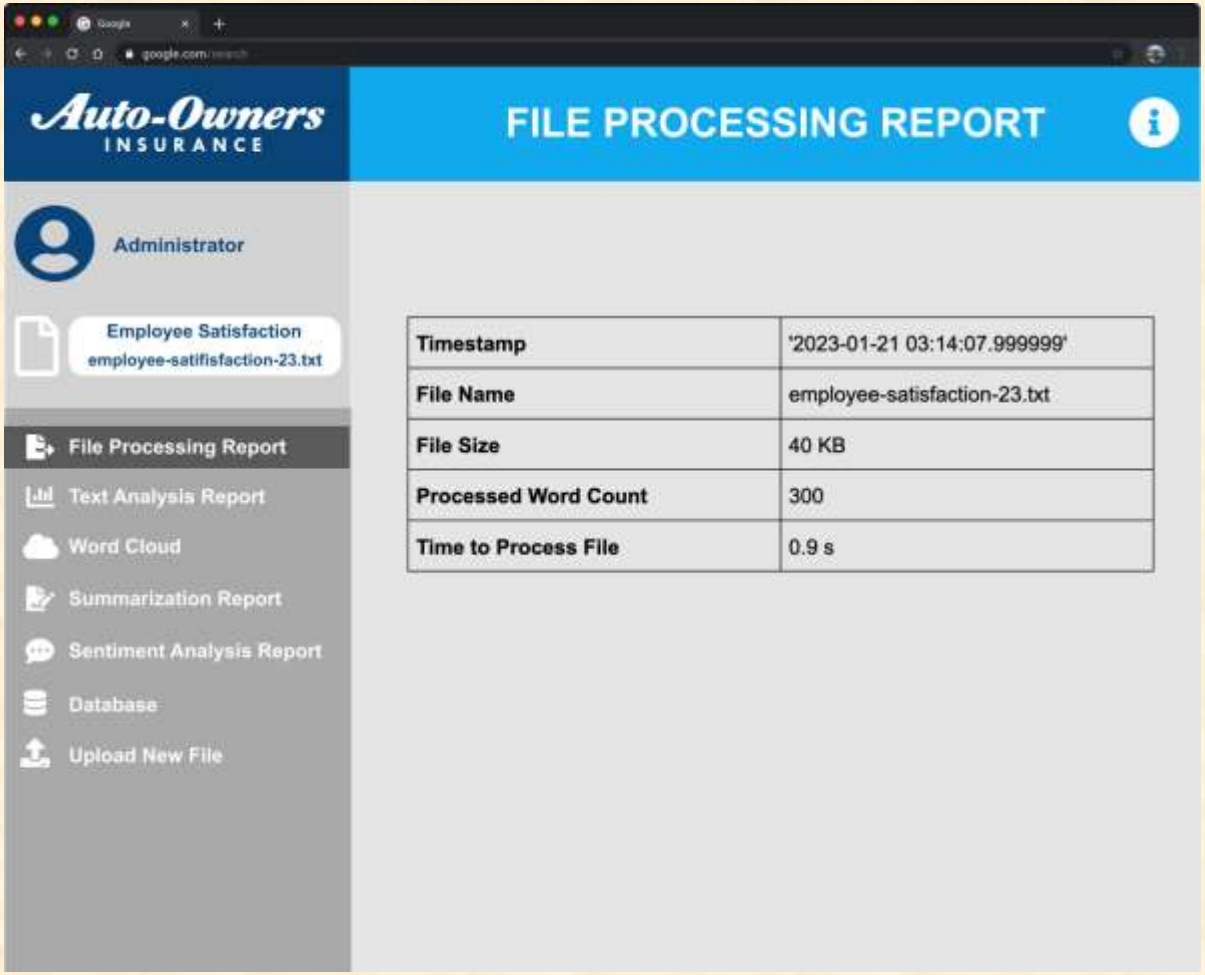
Screen Mockup: Login Page



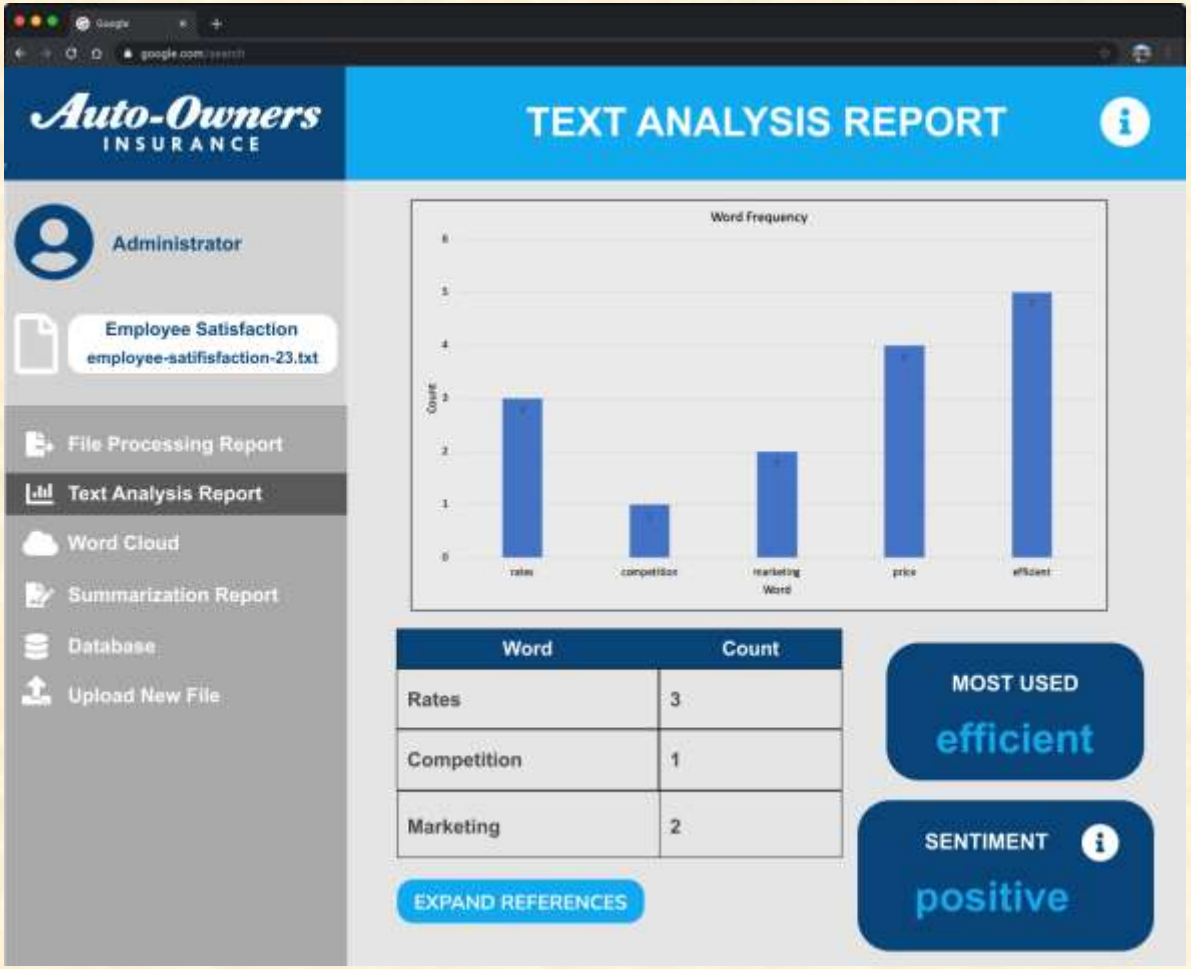
Screen Mockup: File Upload



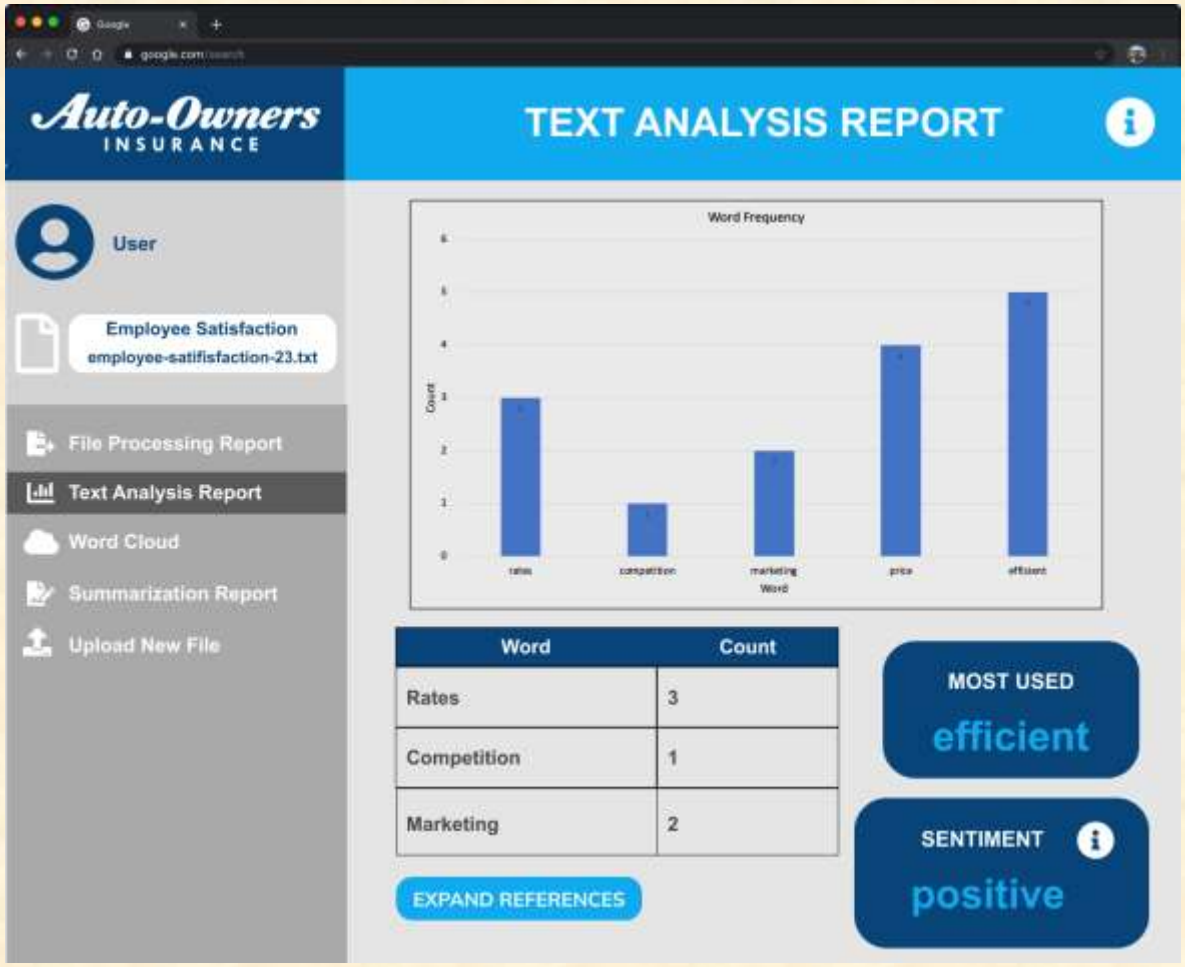
Screen Mockup: File Processing



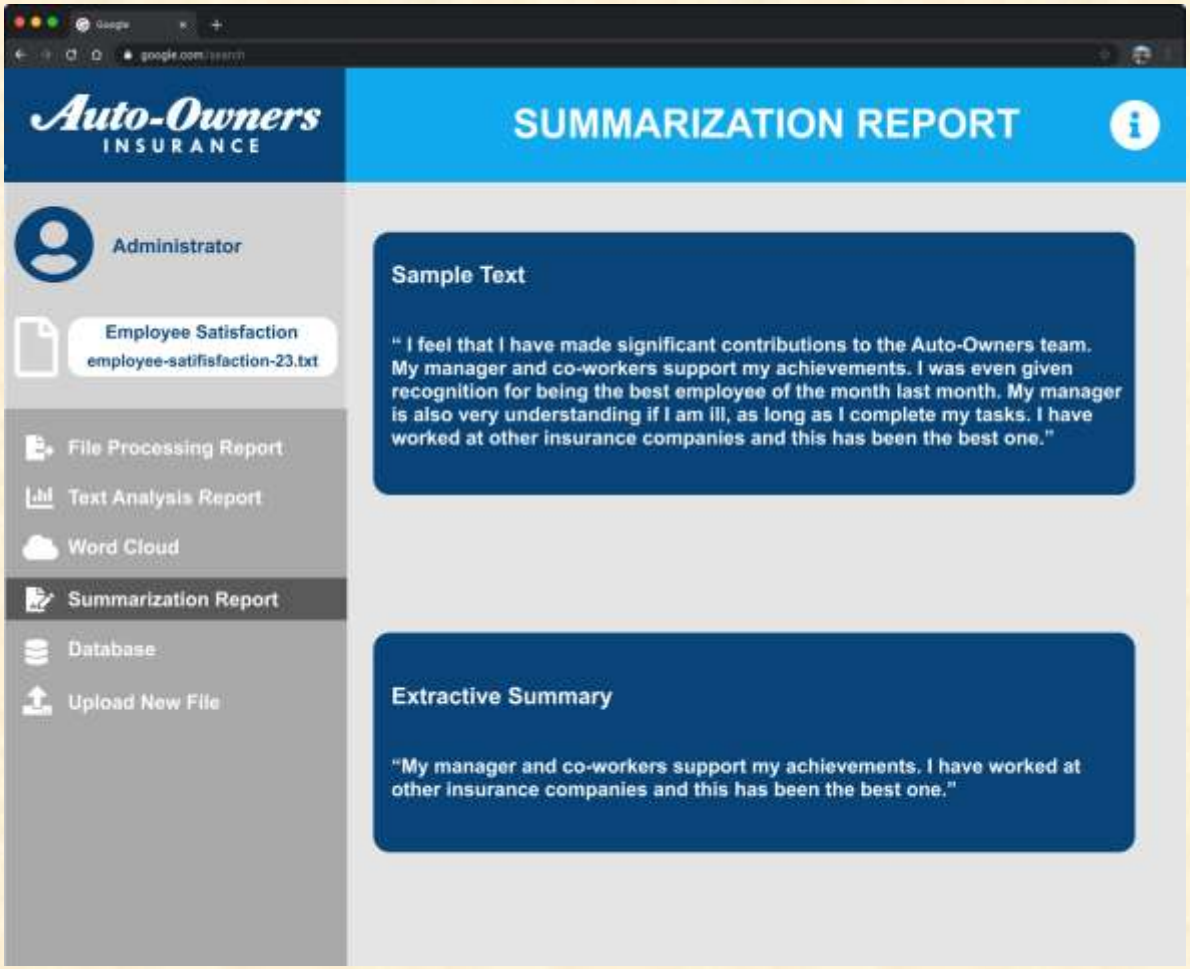
Screen Mockup: Text Analysis



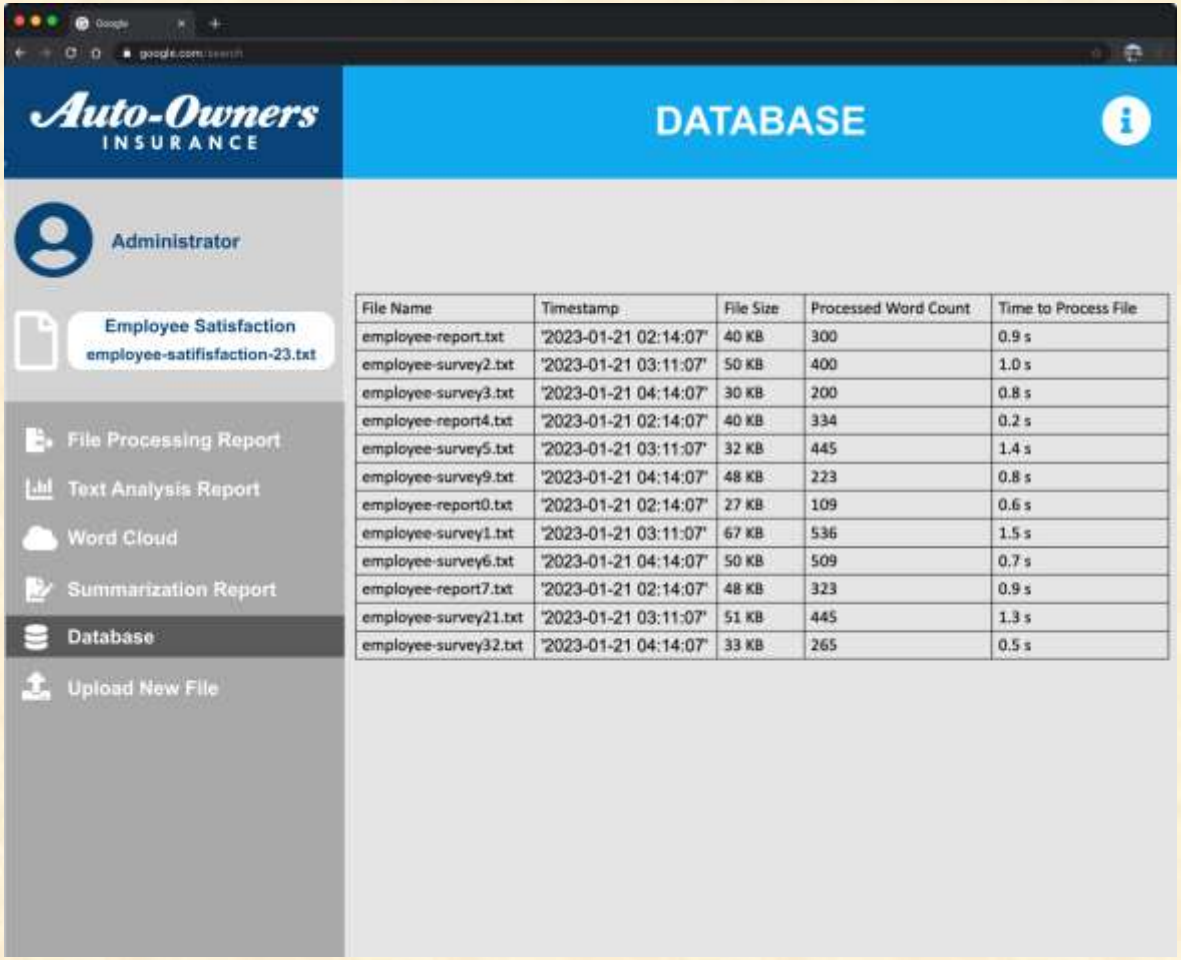
Screen Mockup: User View



Screen Mockup: Summarization Report



Screen Mockup: Database

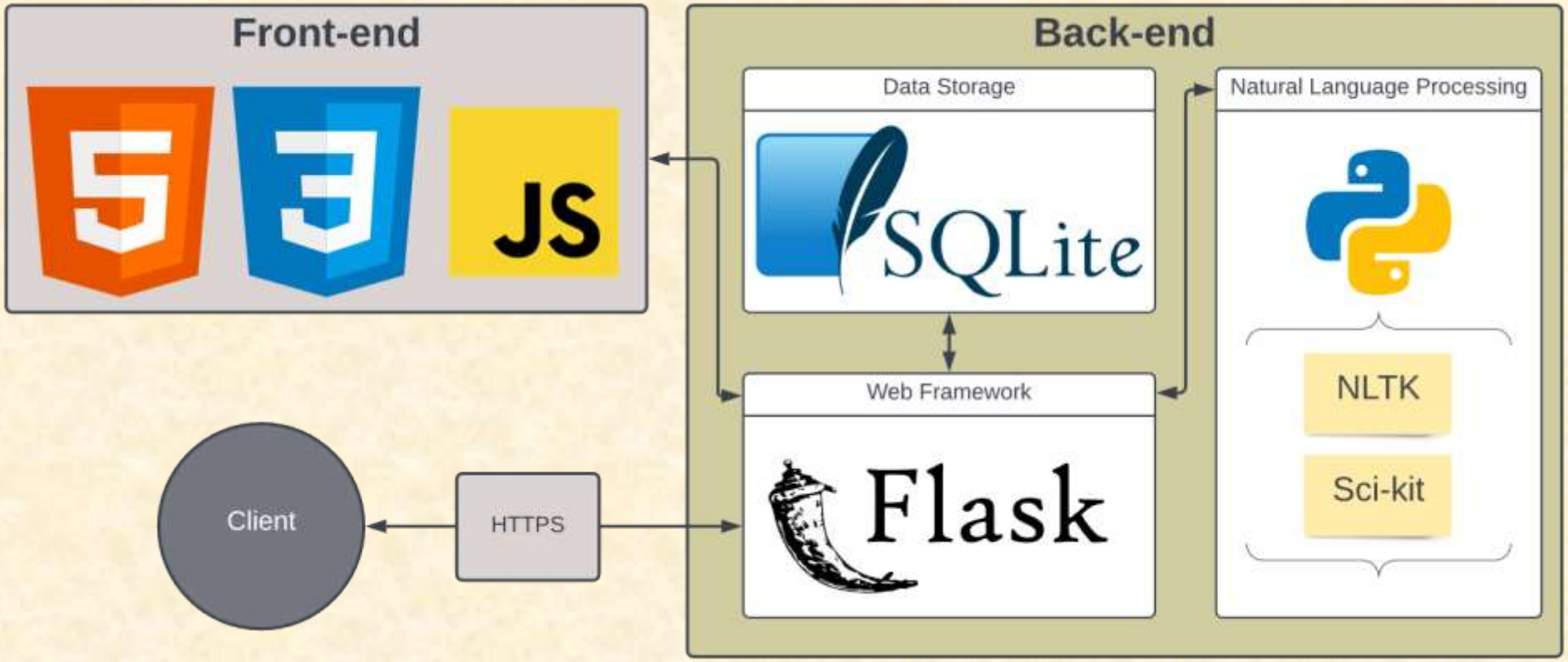


File Name	Timestamp	File Size	Processed Word Count	Time to Process File
employee-report.txt	'2023-01-21 02:14:07'	40 KB	300	0.9 s
employee-survey2.txt	'2023-01-21 03:11:07'	50 KB	400	1.0 s
employee-survey3.txt	'2023-01-21 04:14:07'	30 KB	200	0.8 s
employee-report4.txt	'2023-01-21 02:14:07'	40 KB	334	0.2 s
employee-survey5.txt	'2023-01-21 03:11:07'	32 KB	445	1.4 s
employee-survey9.txt	'2023-01-21 04:14:07'	48 KB	223	0.8 s
employee-report0.txt	'2023-01-21 02:14:07'	27 KB	109	0.6 s
employee-survey1.txt	'2023-01-21 03:11:07'	67 KB	536	1.5 s
employee-survey6.txt	'2023-01-21 04:14:07'	50 KB	509	0.7 s
employee-report7.txt	'2023-01-21 02:14:07'	48 KB	323	0.9 s
employee-survey21.txt	'2023-01-21 03:11:07'	51 KB	445	1.3 s
employee-survey32.txt	'2023-01-21 04:14:07'	33 KB	265	0.5 s

Project Technical Specifications

- Lightweight
- Front-end
  - HTML5, CSS, JavaScript
- Back-end
  - Flask – connects events from front-end to back-end
  - SQLite – serverless database
  - Python Libraries – NLTK, Sci-Kit

Project System Architecture



Project System Components

- Hardware Platforms
  - No hardware platforms
- Software Platforms / Technologies
  - HTML, CSS, JavaScript
    - Develop and style the front-end of our website
  - Flask
    - Easy to use, microframework in python
  - Python
    - Used for back-end text processing
    - Sci-Kit and NLTK
  - SQLite
    - Used as a simple way to implement a database

Project Risks

- Developing a Sentiment Analysis Model
  - The team is unfamiliar with how to create sentiment analysis model and maximize its performance
  - Begin by researching common sentiment models/techniques and implement simple versions. Then develop more complex models and compare the simple baseline results
- Performing Abstractive Text Summarization
  - The team is unsure how to create a successful abstractive text summarizer
  - We will implement quick and simple models at first and test them on small pieces of text
- Machine Learning Performance on Large Documents
  - The models need to process documents of up to 100 pages quickly and effectively
  - We will use large text files from the Canterbury corpus to frequently test the speed of the models on large documents as we develop them
- Acquiring Relevant Data for Machine Learning
  - The team doesn't have access to Auto-Owners corporate data for training the machine learning models
  - We will combine several public datasets made of different types of text to create a more generalized model

Questions?