



Project Plan

Automotive Specific Dark Web Threat Intelligence

The Capstone Experience

Team GM

Anthony Ajlouny

Mitch Neelis

Kevin Kimmel

Ben Kelenski

Yuxin Li

Department of Computer Science and Engineering

Michigan State University

Fall 2020



*From Students...
...to Professionals*

Functional Specifications

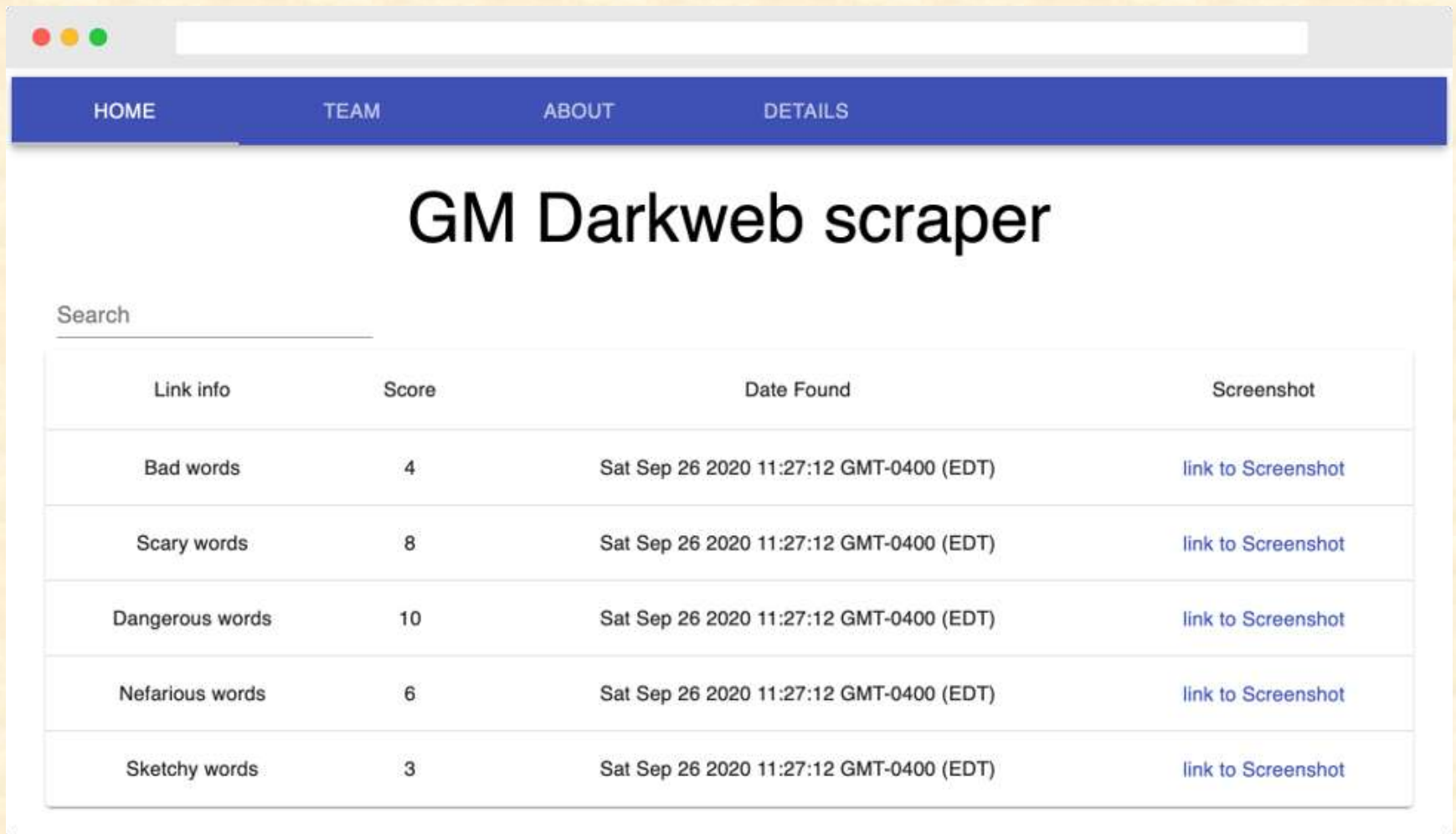
- Big companies like GM need to keep watch for threats to company safety and security, especially on the "dark web".
- A web app that makes searching and identifying threats easy.
- Crawl through blogs to find any suspicious activity, then save screenshots in the web app database.
- Ranking system to help identify major threats from the minor ones.
- Organization will help people find the biggest problem areas and act upon their findings.
- This information can help the security team at GM identify sensitive information on the internet, and act upon a possible security breach.



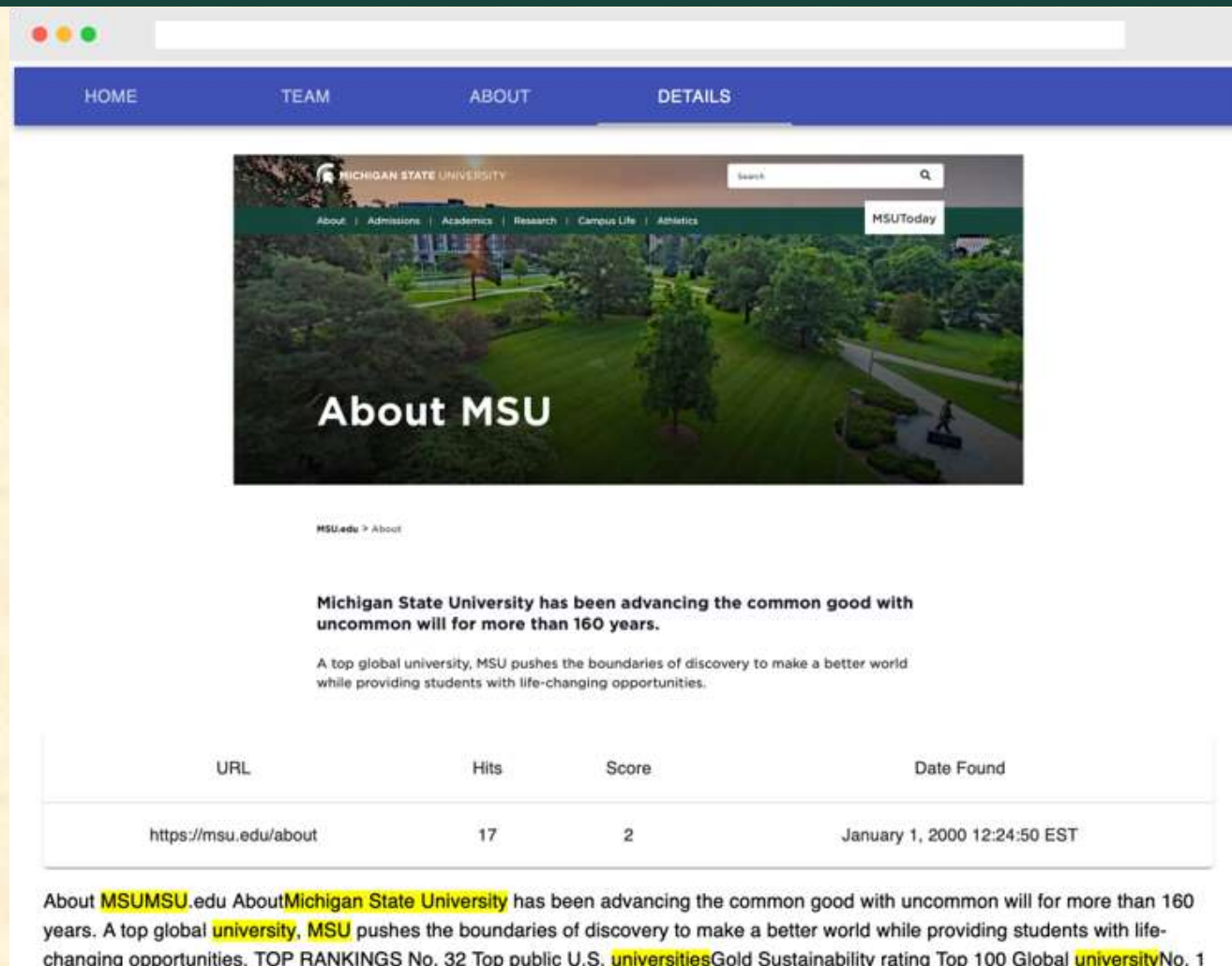
Design Specifications

- The project's application will allow authorized users to run crawling code through suspicious websites, searching for specific buzzwords and saving them for inspection.
- The application will have a ranking system that shows items that prioritizes more "hits".
- After distinguishing threats, the data is formatted for a database where it can be referenced and acted upon if needed.

Screen Mockup: Main User Interface



Screen Mockup: Details Page



Screen Mockup: Run Scraper

The mockup shows a web browser window with a dark blue navigation bar containing links for HOME, TEAM, ABOUT, DETAILS, and RUN. The RUN link is highlighted. The main content area is white and contains three input fields with labels: 'How often to run (minutes):' with a value of '45', 'Websites to search (URL):' with a value of 'msu.edu/about', and 'Words to search for:' with a value of 'University'. Each input field has a blue '+' button to its right. Below these fields is a blue button labeled 'RUN SCRAPER'.

HOME TEAM ABOUT DETAILS RUN

How often to run (minutes): 45

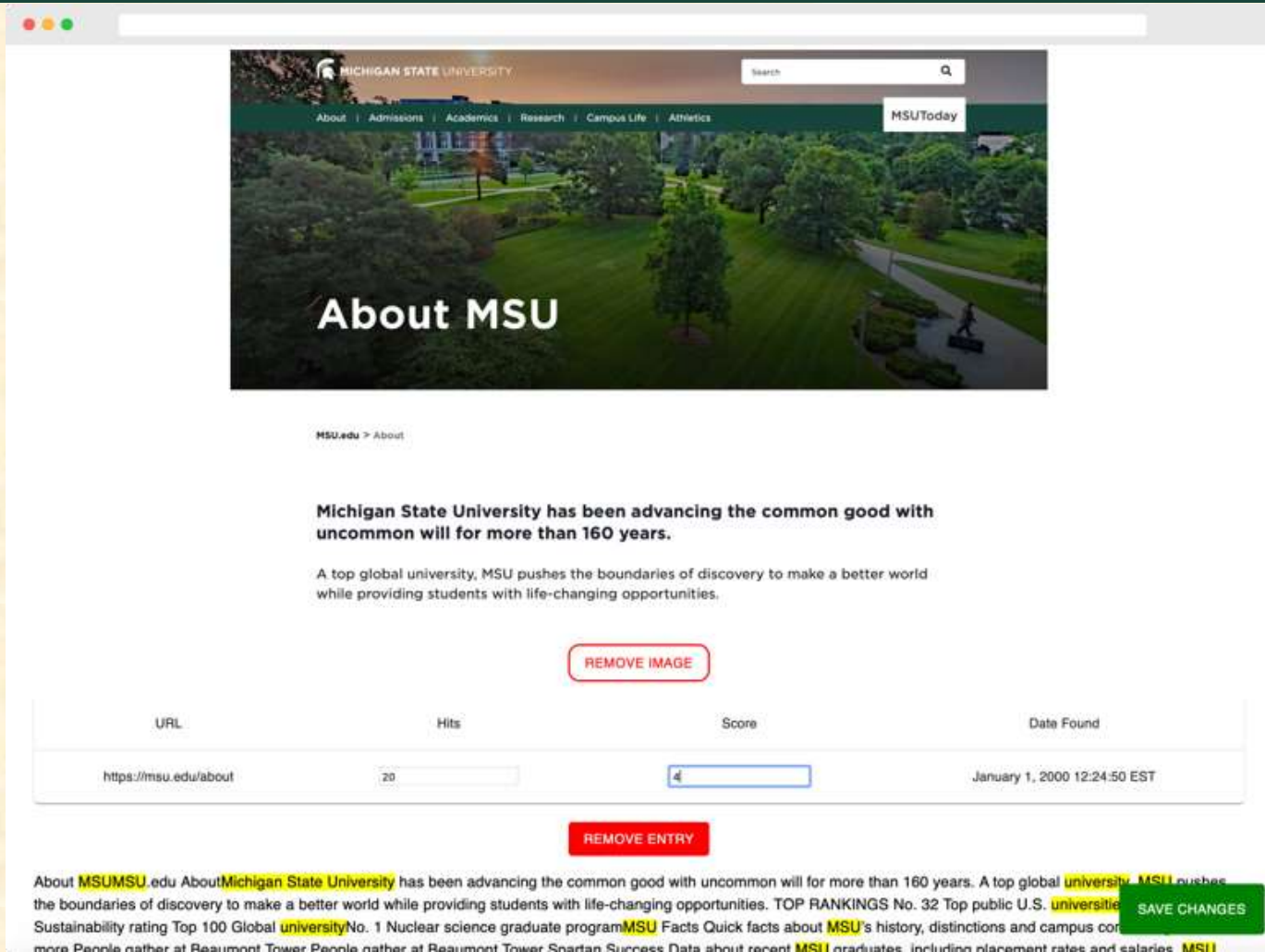
Websites to search (URL): msu.edu/about +

Words to search for: University +

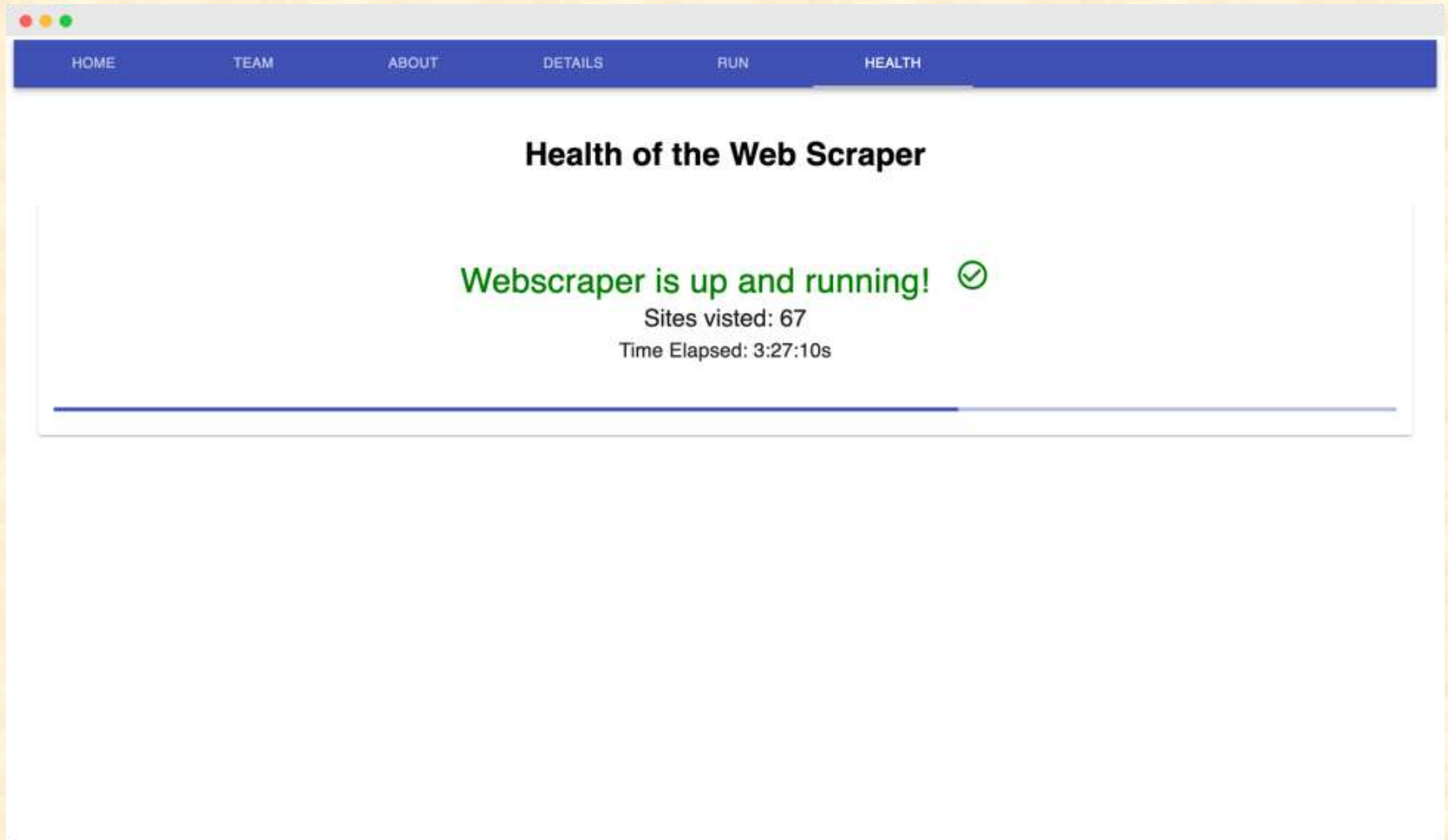
RUN SCRAPER



Screen Mockup: Edit Entry



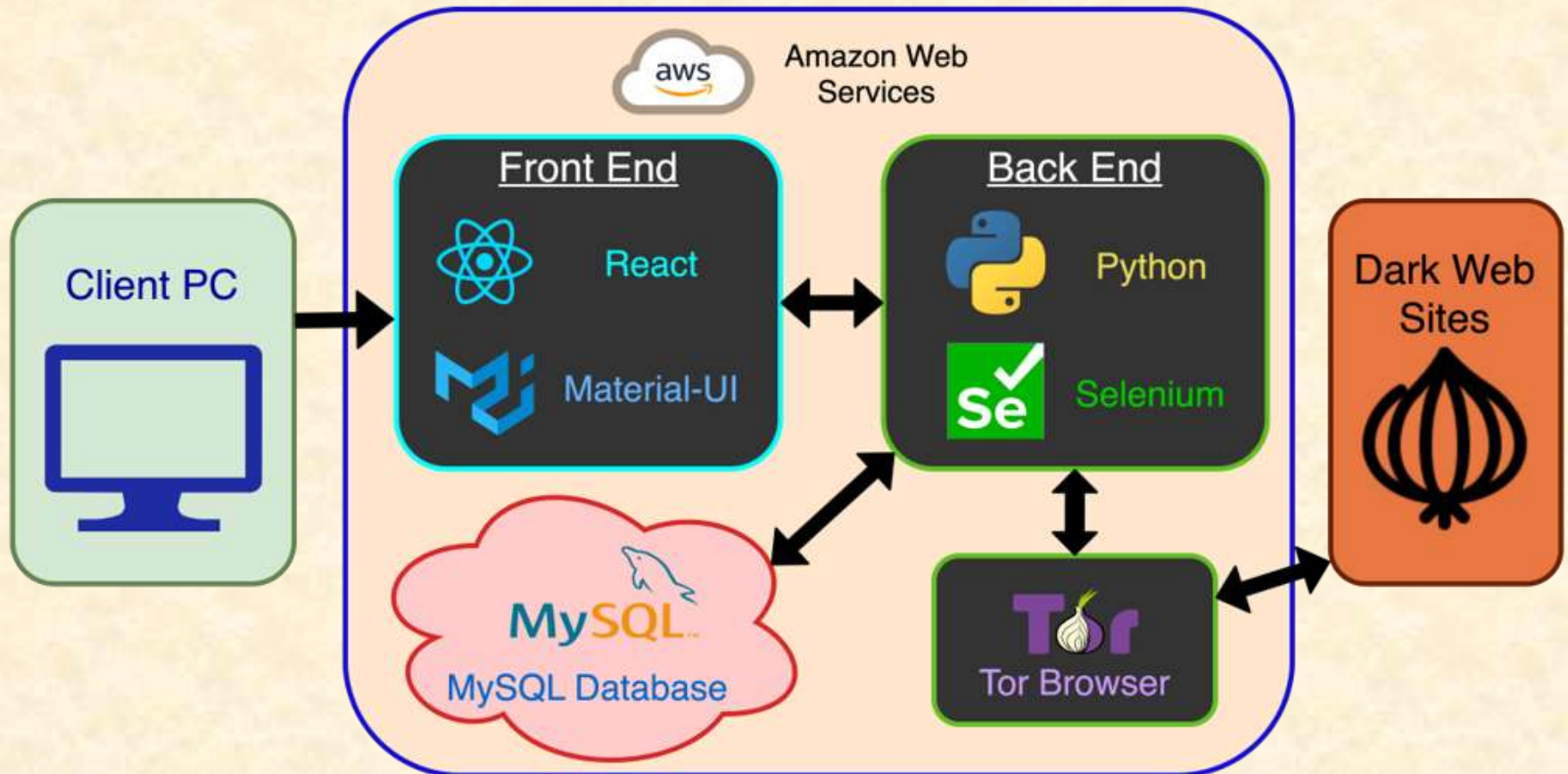
Screen Mockup: Web Scraper Health Page



Technical Specifications

- Application users will need to sign in for access to keep information secure.
- The application is semi-automatic, where it is on a timer that activates after an allotted amount of time.
- The code reads information from a text file that contains information that should be searched for, along with which websites to search through.
- Brands, employee names, IP addresses
- Reddit, onion sites
- After the program has completed, the data is inspected for threat level, and a ranking is placed.
- Amount of hits, severity of information, website it is found on
- Users will see a table of all collected data with some basic information about it. They can access the details and edit anything they feel should be changed.

System Architecture



System Components

- Hardware Platforms
 - Amazon Web Services
 - MySQL database
- Software Platforms / Technologies
 - Visual Studio Code
 - Pycharm
 - JupyterHub
 - Selenium
 - Tor

Risks

- Risk 1
 - Amazon Web Services is a new platform, so there may be some difficulties when uploading and testing our code.
 - We plan to test some skeleton programs to ensure everything is running smoothly on personal machines and AWS.
- Risk 2
 - Anything goes in the dark web. Some websites can be considered unsafe or have anti-scraping bots in place.
 - To access dark web, we are using a Tor browser to keep our location and information anonymous and storing data in AWS adds security to our data.
- Risk 3
 - AWS has a service called Lambda where you can run your code automatically based on specific events set. For example, we want to make our web scraper run automatically every hour on the hour.
 - We plan to test the scraper in Lambda as early as possible to make sure it is functional and won't encounter problems later in the project.
- Risk 4
 - General Motors wants to search dark web sites for security risks to their company and industry. GM wants to ensure that the product is secure from breaches and doesn't draw any suspicion.
 - Hosting the web app on AWS, which allows certain IP addresses to access the app. Adding a login page adds in an extra layer of security.



Questions?

?

?

?

?

?

?

?

?

?

