



Beta Presentation

Automotive Specific Dark Web Threat Intelligence

The Capstone Experience

Team GM

Anthony Ajlouny

Mitch Neelis

Kevin Kimmel

Ben Kelenski

Yuxin Li



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

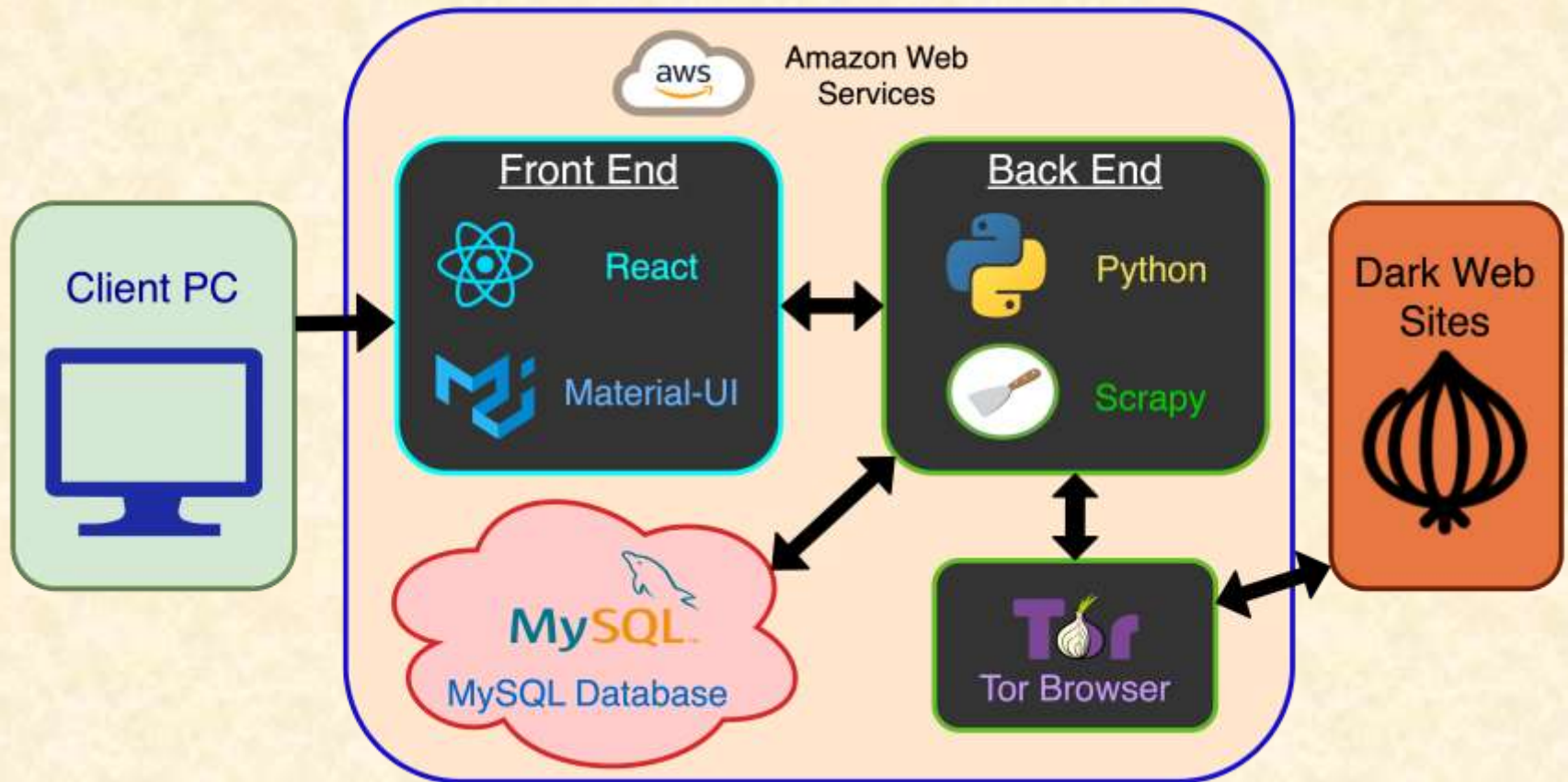
Department of Computer Science and Engineering
Michigan State University
Fall 2020

Project Overview

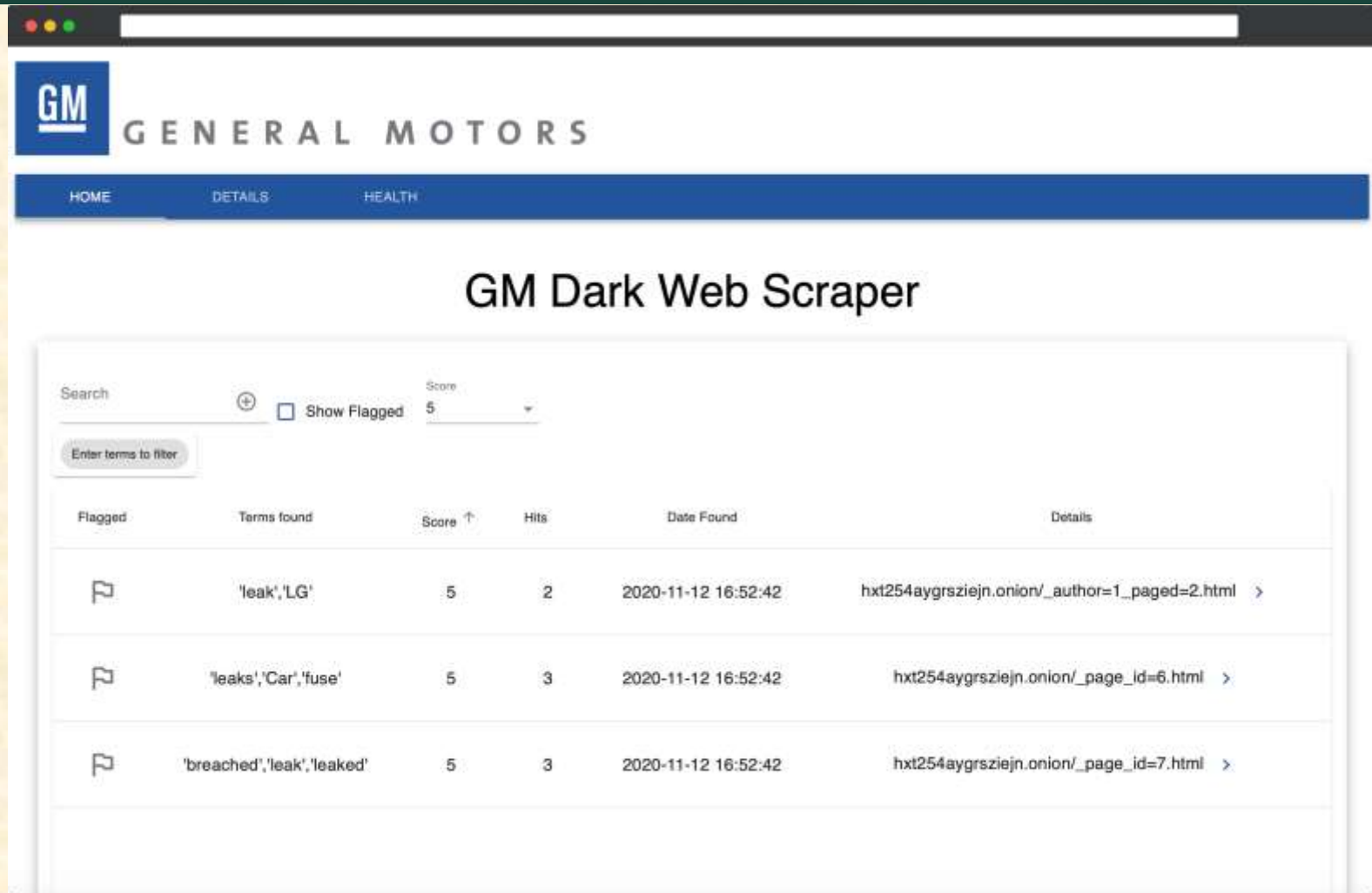
- The program scrapes through the dark web looking for defined key words
- Text found on the websites are scraped and saved in the database, including possible images that may exist
- The contents run through our scoring system to check for severity of information
- This info is pushed to the database, where can be viewed and sorted in the web app



System Architecture






Scraper Main Page



The screenshot displays the GM Dark Web Scraper application. At the top, there is a browser window header with a GM logo and the text "GENERAL MOTORS". Below this is a navigation bar with links for "HOME", "DETAILS", and "HEALTH". The main heading is "GM Dark Web Scraper".

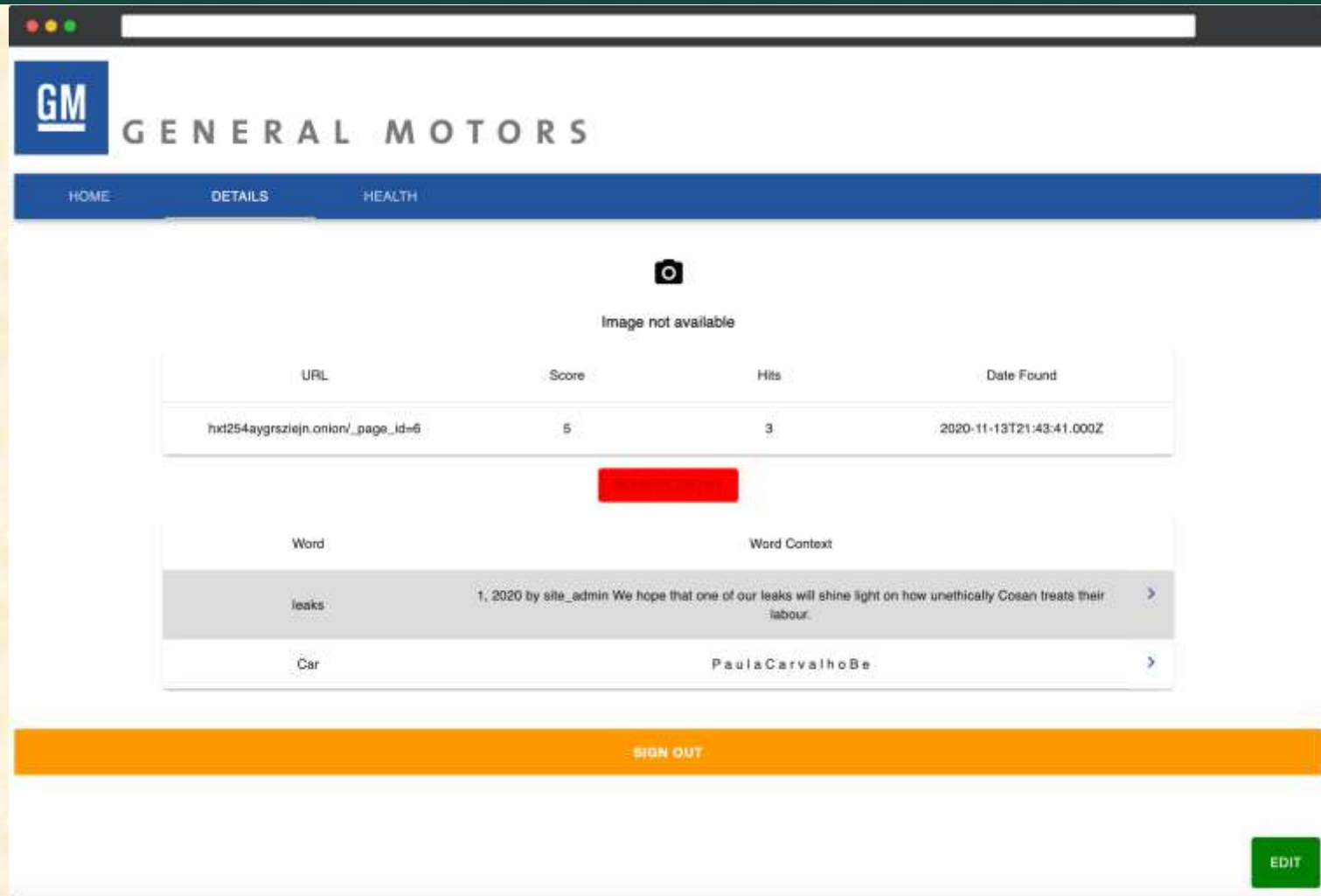
The interface includes a search section with a "Search" input field, a "Show Flagged" checkbox, and a "Score" dropdown menu set to "5". A button labeled "Enter terms to filter" is also present.

Below the search section is a table displaying search results. The table has columns for "Flagged", "Terms found", "Score", "Hits", "Date Found", and "Details".

Flagged	Terms found	Score	Hits	Date Found	Details
	'leak','LG'	5	2	2020-11-12 16:52:42	hxt254aygrsziejn.onion/_author=1_paged=2.html >
	'leaks','Car','fuse'	5	3	2020-11-12 16:52:42	hxt254aygrsziejn.onion/_page_id=6.html >
	'breached','leak','leaked'	5	3	2020-11-12 16:52:42	hxt254aygrsziejn.onion/_page_id=7.html >



Details Page



The screenshot shows a web browser window displaying the GM (General Motors) website. The page is titled "DETAILS" in the navigation bar. The main content area features a camera icon and the text "Image not available". Below this is a table with four columns: URL, Score, Hits, and Date Found. The table contains one row of data. Below the table is a red button. Further down is another table with two columns: Word and Word Context. This table contains two rows of data. At the bottom of the page is an orange "SIGN OUT" button and a green "EDIT" button.

GM GENERAL MOTORS

HOME DETAILS HEALTH

Image not available

URL	Score	Hits	Date Found
hxt254aygrszlejn.onkon/_page_id=6	5	3	2020-11-13T21:43:41.000Z

Word Context

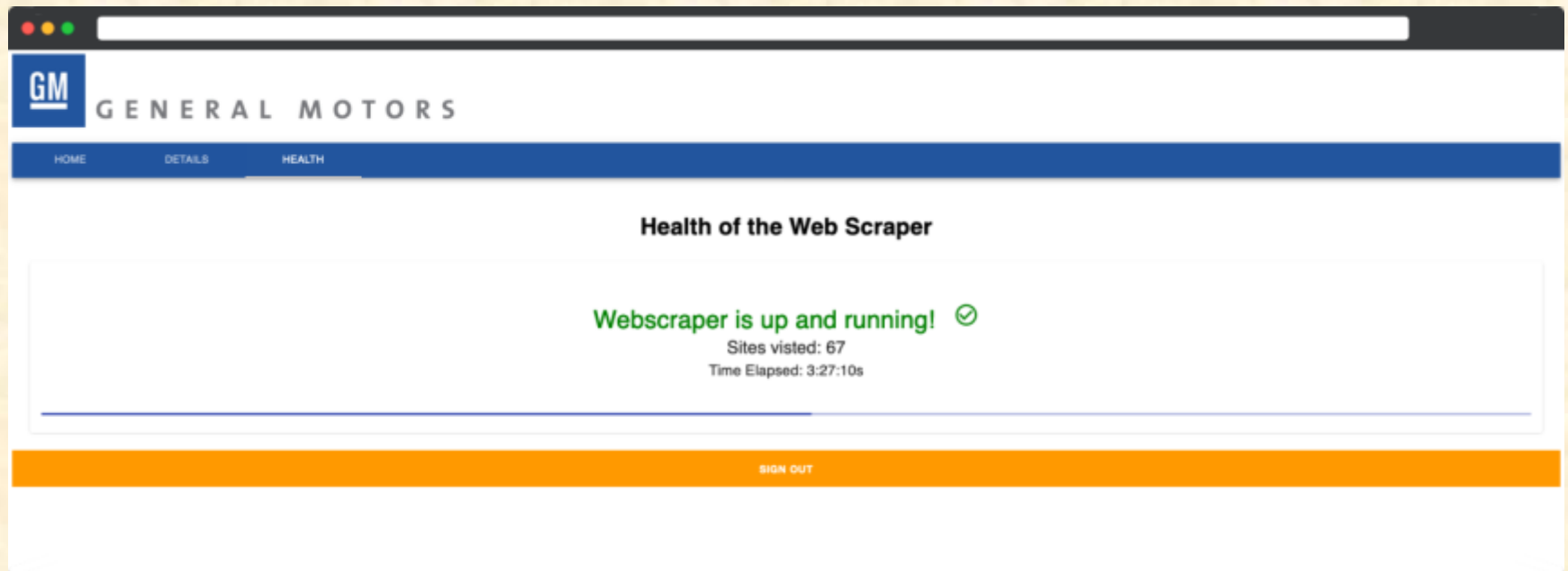
leaks	1, 2020 by site_admin We hope that one of our leaks will shine light on how unethically Cosan treats their labour.	>
Car	PaulaCarvalhoBe	>

SIGN OUT

EDIT



Scraper Health Page



Scrape Data Table

	ID	Score	URL	Terms found	Date Found	Favorite
0	73	3	hxt254aygrsziejn.onion/_page_id=6	{'term_1': {'term': 'breached', 'context': 'co...	2020-11-15 15:59:45.047723	False
1	12	2	hxt254aygrsziejn.onion/_p=382	{'term_1': {'term': 'breached', 'context': 'in...	2020-11-15 15:59:45.047723	False
2	1	2	hxt254aygrsziejn.onion/_p=303	{'term_1': {'term': 'breached', 'context': 'in...	2020-11-15 15:59:45.047723	False
3	45	2	hxt254aygrsziejn.onion/_p=205	{'term_1': {'term': 'breached', 'context': 'in...	2020-11-15 15:59:45.047723	False
4	4	2	hxt254aygrsziejn.onion/_cat=9	{'term_1': {'term': 'breached', 'context': '&r...	2020-11-15 15:59:45.047723	False
5	53	2	hxt254aygrsziejn.onion/_p=166	{'term_1': {'term': 'breached', 'context': 'Of...	2020-11-15 15:59:45.047723	False
6	109	2	hxt254aygrsziejn.onion/_cat=7	{'term_1': {'term': 'leak', 'context': 'July 5...	2020-11-15 15:59:45.047723	False
7	55	2	hxt254aygrsziejn.onion/_p=427	{'term_1': {'term': 'breach', 'context': 'to c...	2020-11-15 15:59:45.047723	False
8	65	2	hxt254aygrsziejn.onion/_p=136	{'term_1': {'term': 'leak', 'context': 'April ...	2020-11-15 15:59:45.047723	False
9	75	1	hxt254aygrsziejn.onion/_author=1	{'term_1': {'term': 'leak', 'context': '5.Post...	2020-11-15 15:59:45.047723	False
10	70	1	hxt254aygrsziejn.onion/_p=346	{'term_1': {'term': 'leak', 'context': '2, 202...	2020-11-15 15:59:45.047723	False
11	69	1	hxt254aygrsziejn.onion/_cat=26	{'term_1': {'term': 'leak', 'context': '00Orang...	2020-11-15 15:59:45.047723	False
12	67	1	hxt254aygrsziejn.onion/_p=142	{'term_1': {'term': 'breached', 'context': 'no...	2020-11-15 15:59:45.047723	False
13	63	1	hxt254aygrsziejn.onion/_p=392	{'term_1': {'term': 'leak', 'context': 'July 1...	2020-11-15 15:59:45.047723	False
14	0	1	hxt254aygrsziejn.onion/_p=362	{'term_1': {'term': 'leak', 'context': 'the th...	2020-11-15 15:59:45.047723	False
15	59	1	hxt254aygrsziejn.onion/_cat=5	{'term_1': {'term': 'leak', 'context': '5.Post...	2020-11-15 15:59:45.047723	False
16	78	1	hxt254aygrsziejn.onion/_cat=10	{'term_1': {'term': 'leak', 'context': 'August...	2020-11-15 15:59:45.047723	False
17	58	1	hxt254aygrsziejn.onion/_cat=11	{'term_1': {'term': 'leak', 'context': 'June 2...	2020-11-15 15:59:45.047723	False
18	57	1	hxt254aygrsziejn.onion/_p=202	{'term_1': {'term': 'leak', 'context': 'Symbol...	2020-11-15 15:59:45.047723	False
19	50	1	hxt254aygrsziejn.onion/_author=1_paged=2	{'term_1': {'term': 'leak', 'context': '9, 202...	2020-11-15 15:59:45.047723	False



What's left to do?

- Debug Scraper Health Page
- Allow user to manually run the scraper
- Display possible images found for wider security coverage
- Neatly display full text file when prompted
- Debug parsing algorithm for URL's and terms
- Improve security on the web app

Questions?

?

?

?

?

?

?

?

?

?

