

**MICHIGAN STATE**  

---

**U N I V E R S I T Y**

# Beta Presentation

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*

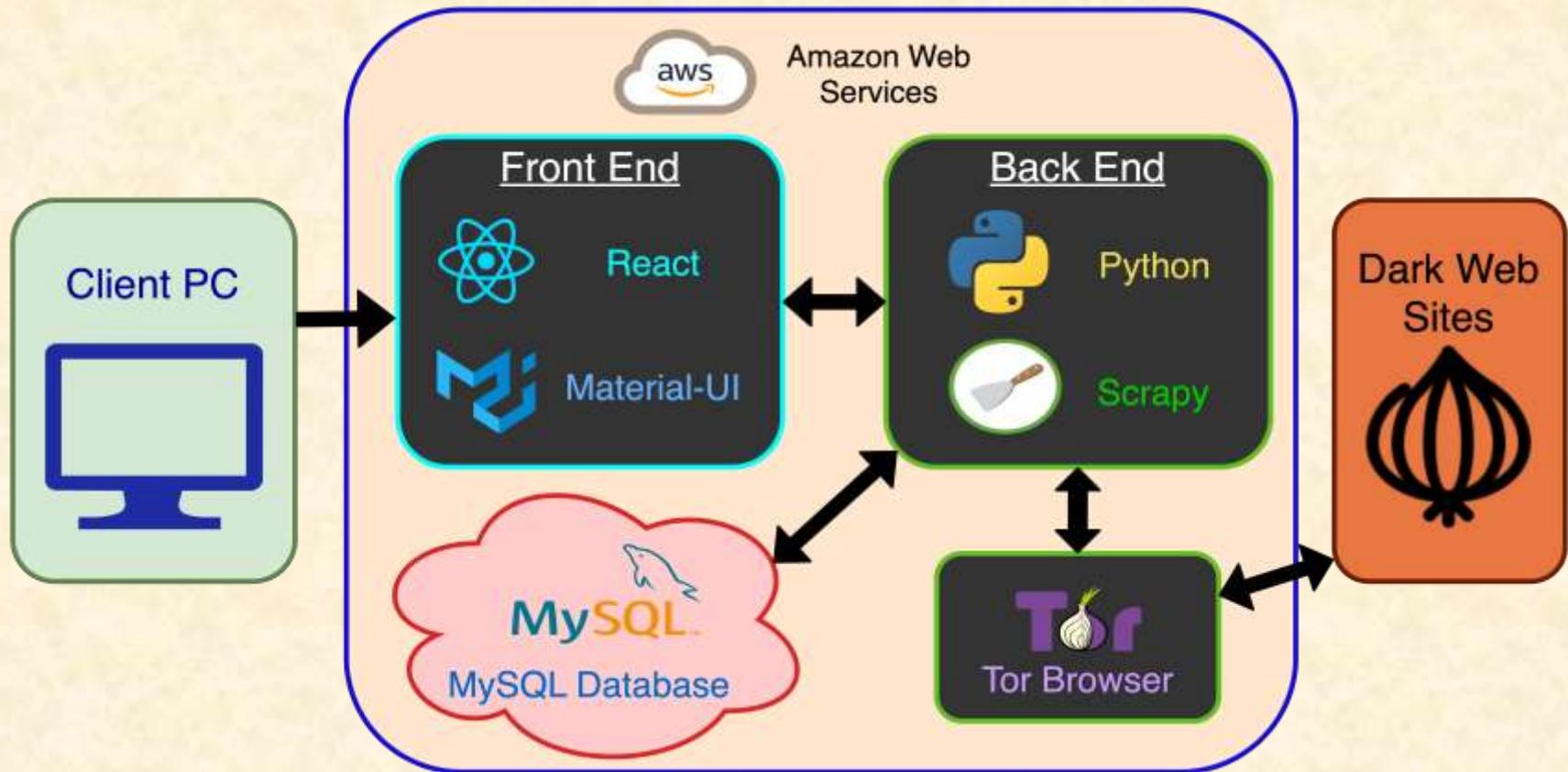
# Project Overview

---

- 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 shows the main interface of the GM Dark Web Scraper. At the top, there is a navigation bar with the GM logo and the text "GENERAL MOTORS". Below this, a blue navigation bar contains the links "HOME", "DETAILS", and "HEALTH". The main heading is "GM Dark Web Scraper".

The interface includes a search section with a search bar containing the text "Enter terms to filter", a "Show Flagged" checkbox, and a "Score" dropdown menu set to "5". Below the search section is a table with the following columns: "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	<a href="http://hxt254aygrsziejn.onion/_author=1_paged=2.html">hxt254aygrsziejn.onion/_author=1_paged=2.html</a> >
	'leaks','Car','fuse'	5	3	2020-11-12 16:52:42	<a href="http://hxt254aygrsziejn.onion/_page_id=6.html">hxt254aygrsziejn.onion/_page_id=6.html</a> >
	'breached','leak','leaked'	5	3	2020-11-12 16:52:42	<a href="http://hxt254aygrsziejn.onion/_page_id=7.html">hxt254aygrsziejn.onion/_page_id=7.html</a> >



# Details Page

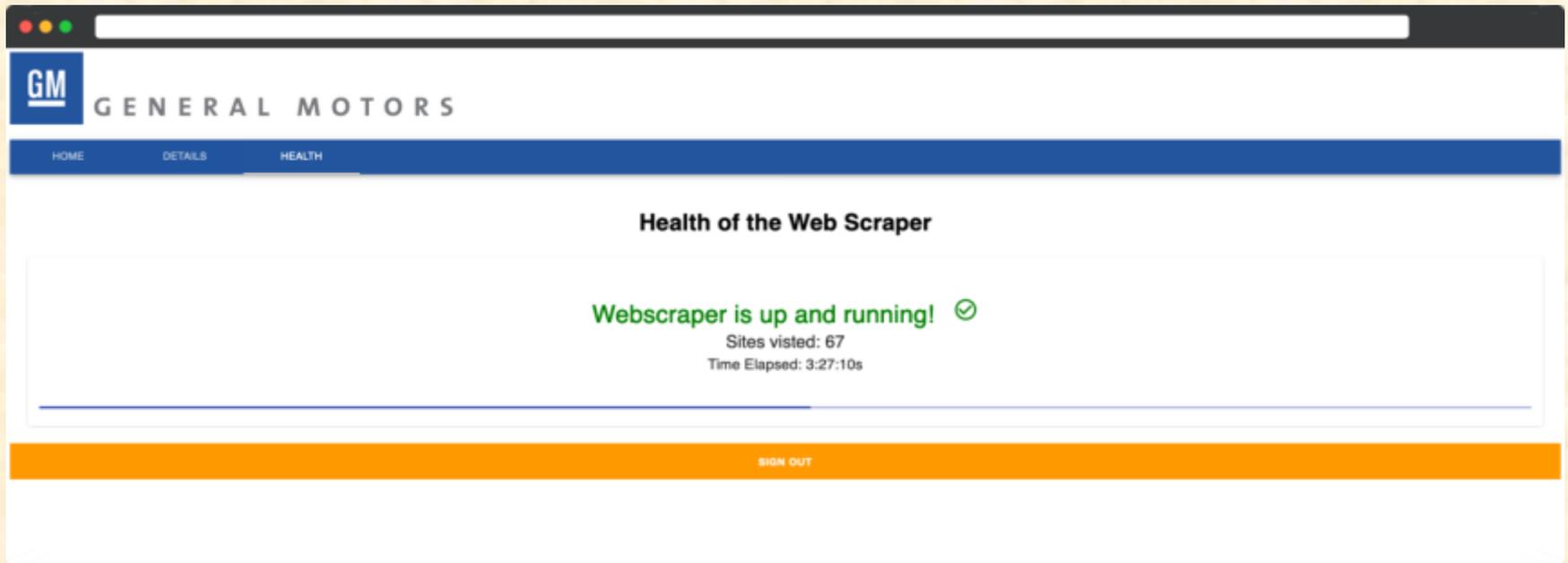
The screenshot displays the GM website's 'DETAILS' page. At the top left is the GM logo. A navigation bar contains 'HOME', 'DETAILS', and 'HEALTH'. Below this is a camera icon and the text 'Image not available'. A table lists search results with columns for URL, Score, Hits, and Date Found. Below the table is a red button. A 'Word Context' section shows the word 'leaks' with its context: '1, 2020 by site\_admin We hope that one of our leaks will shine light on how unethically Cosan treats their labour.' Another word, 'Car', is shown with the context 'PaulaCarvalhoBe'. At the bottom, there is a 'SIGN OUT' button and an 'EDIT' button.

URL	Score	Hits	Date Found
<a href="#">hxt254aygrsziejn.onkon/_page_id=6</a>	5	3	2020-11-13T21:43:41.000Z

Word	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



# 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': '0Orang...	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?

---

?

?

?

?

?

?

?

?

?

