

Software Engineering 491 - sddec19-01

Web Crawling for Data Breach Reports

Week 8 Report

4/13 - 4/19

Client: Benjamin Blakely

Faculty Advisor: Dr. Daniels

**Team Members:**

Mark Schwartz - Scraping Team

Alec Lones - Project Leader - -Machine Learning Team

Nolan Kim - Scraping Team - Git Master

Jeremiah Brusegaard - Machine Learning Team

**Weekly Summary:**

Working on fixing bugs and looking into scraper issues. Somewhat of a slow week due to midterms and other classes for our team. We are getting ready for the presentation as well.

**Past Week Accomplishments:**

Bug Fixes, Demo

**Pending Issues:**

Need database

Better way of storing lemmatized sites

Scraping seems to be currently single threaded and is bottlenecking the system. Might consider multi threading this?

Feature selection for vectorization might be broken or not done correctly

**Individual Contributions:**

Team Member	Contribution	Weekly Hours	Total Hours
Mark Schwartz	<ul style="list-style-type: none"><li>• Worked on demo</li><li>• Helped Jeremiah fix web scraper bugs</li></ul>	~6	~60
Alec Lones	<ul style="list-style-type: none"><li>• Worked on demo</li><li>• Helped Jeremiah on continuing to scrape data and machine learning model</li><li>• Investigated database ideas for project</li></ul>	~6	~60

Nolan Kim	<ul style="list-style-type: none"> <li>• Worked on demo</li> <li>• Continuing to work on multithreading the scraper</li> <li>• Worked on presentation</li> </ul>	~6	~60
Jeremiah Brusegaard	<ul style="list-style-type: none"> <li>• Worked on fixing feature selection and testing vectorization methods</li> <li>• Helped Mark with demo</li> </ul>	~6	~60

**Plans for upcoming week:**

- Mark Schwartz:
  - Finish demo
  - Keep optimizing crawler
  - Revise project plan
- Alec Lones:
  - Continue working on demo
  - Continue investigating databases
  - Finish presentation
  - Finish revising project plan
  - Finish up any other 491 assignments
  - Continue assisting Jeremiah with machine learning stuff
- Nolan Kim:
  - Continue working on demo
  - Finish presentation
  - Finish revising project plan
  - Continue working on scraper multithreading
- Jeremiah Brusegaard:
  - Work on project plan revision
  - Continue fixing vectorization and web scraper bugs
  - Work on presentation

**Summary of weekly meeting:**

Talked about what we need to do for the upcoming week and bug fixes that need to be done still. Also talked about what to use for the demo and how to present it. Also talked about meeting for fixing our project plan and how we would go about that.