Software Engineering 491 - sddec19-01
Web Crawling for Data Breach Reports
Week 5 Report
3/8 - 3/15
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:
Team decided to get a head start on the Design Document. We finished a good portion of it prior to Spring Break. Team members also worked on individual scrapers and language processors.

Past Week Accomplishments:
- Got a prototype web crawler to crawl seed urls and urls from those. It has a blocked domain feature and now lemmatizes the text.
- Started a chunk of the design document.

Pending Issues:
None

Individual Contributions:

Team Member	Contribution	Weekly Hours	Total Hours
Mark Schwartz	● Started the design doc ● Played around with beautiful soup. ● Attached beautiful soup to my crawler	~6	~36
Alec Lones	● Got a head start on the design doc ● Spent a little time working on my food scraper	~6	~36

	● Most of my time was spent working on the design doc		
Nolan Kim	● Worked on the design document ● Continued to experiment with the Scrapy library	~6	~36
Jeremiah Brusegaard	● Worked on fixing the beautiful soup bug ● Also it the crawler lemmatizes the text from the site ● Worked on design document ● Looked into vectorization as well	~6	~36

Plans for upcoming week:

- Mark Schwartz:
  - Finish design doc
  - Continue to finish and debug crawling and lemmatizer.
- Alec Lones:
  - Enjoy Spring Break
  - Finish Design Doc
  - Continue working on food scraper, lemmatizer, stemmer, and vectorizer
- Nolan Kim:
  - Finish Design Doc
  - Continue learning Scrapy over spring break
- Jeremiah Brusegaard:
  - Finish Design doc
  - Possibly add vectorization, need to consult with Ben how he recommends going about doing that

Summary of weekly meeting:

We had scheduling conflicts and didn't have an in person meeting so we just set goals for ourselves this week. Also we are actively working on finishing the design doc.