

Software Engineering 491 - sddec19-01

Web Crawling for Data Breach Reports

Week 2 Report

2/09 - 2/15

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 created rough outline and started on the Project Plan. Additionally the team explored different areas of the project and read documentation on different libraries. Preliminary library testing has been started and toy examples are being written.

**Past Week Accomplishments:**

This week we made significant on the project plan document. - Everyone

- Use case diagram done
- Machine Learning Research
- Generalized pipeline diagram created

**Pending Issues:**

None

**Individual Contributions:**

Team Member	Contribution	Weekly Hours	Total Hours
Mark Schwartz	Added significantly to the first two sections of the project plan,	~6	~12

	especially the diagrams. Did more “Hello World” stuff with scraping on Python.		
Alec Lones	Worked on the Project Plan, Explored Natural Language Processing more, Explored Machine Learning, Explored Scrapey, Created General Project Pipeline Diagram	~6	~12
Nolan Kim	<ul style="list-style-type: none"> <li>• Did additional experimentation with Scrapy</li> <li>• Worked on project plan</li> <li>• Started research into machine learning concepts</li> </ul>	~6	~12
Jeremiah Brusegaard	<ul style="list-style-type: none"> <li>• Worked on project plan</li> <li>• Researching Python libraries for ML</li> <li>• Learning Python more advanced tutorial</li> <li>• Beginning research on vectorization for ML algorithm</li> </ul>	~6	~12

**Plans for upcoming week:**

- Mark Schwartz: Finish project plan document, play more/finish with scraping on Python; hopefully add some ML stuff to it.
- Alec Lones: Finish Project Plan, Work on integrating different libraries together (aka write hello world with each library then start pipelining them together)
- Nolan Kim: Finish project plan, learn more in-depth scrapy, continue machine learning research
- Jeremiah Brusegaard: Finish project plan, work on experimental code with libraries and push to git

**Summary of weekly advisor meeting:**

We checked in with everyone to make sure we were all doing ok with knowing our roles and what we should be doing for next week. Suggested that we should start doing “hello world” examples with the libraries and pipelining them to make sure they work together.