

Software Engineering 492 - sddec19-01

Web Crawling for Data Breach Reports

Bi-Weekly Report 7

11/22 - 12/06

Client: Benjamin Blakely

Faculty Advisor: Benjamin Blakely

Team Members:

Mark Schwartz - Scraping Team

Alec Lones - Project Leader -Machine Learning Team

Nolan Kim - Scraping Team - Git Master

Jeremiah Brusegaard - Machine Learning Team

Bi-weekly Summary:

We looked at trying to fix the memory problem in our crawler but could not find a solution so we resorted to allowing the user to pass in a starting url from the command line so that on restart they wouldn't have to modify any code. Also we finished our poster and started working on the final report and presentation.

Past 2 Weeks Accomplishments:

- Finished up the project with the exception of the memory issue
- Poster looks nice
- Added a slight work around for having to restart crawler due to memory problem

Unresolved Issues:

- Scrapy is limited in the way that it is able to continue onto new links and due to this we could not fix the memory issue so the crawler will have to be restarted approximately every 24 hours assuming the user has 16 GB of ram

Individual Contributions:

Team Member	Contribution	Bi- weekly Hours	Total Hours
Mark Schwartz	<ul style="list-style-type: none">● Worked on poster● Helped create/polish UI● Started final presentation	~12	~84
Alec Lones	<ul style="list-style-type: none">● Worked on poster● Created frontend interface● Polished code	~12	~84

Nolan Kim	<ul style="list-style-type: none"> ● Helped work on poster creation ● Helped improve crawler's memory usage ● Worked on final presentation 	~12	~84
Jeremiah Brusegaard	<ul style="list-style-type: none"> ● Worked on poster ● Added command line argument for start url so user can rerun the program easily ● Polished code ● Fixed issue with Scraper's DB integration 	~12	~84

Plans for upcoming 2 weeks:

- Mark Schwartz:
 - Final Presentation and turn in project to client
- Alec Lones:
 - Final Presentation and turn in project to client
 - Polish frontend
- Nolan Kim:
 - Final Presentation and turn in project to client
- Jeremiah Brusegaard:
 - Final Presentation and turn in project to client

Summary of bi-weekly meeting:

We didn't meet with the client due to thanksgiving break. As a team we met to discuss final touches need for the project as well as dates to work on the poster and presentation.