

sddec19-01

Jeremiah Brusegaard

Alec Lones

Nolan Kim

Mark Schwartz

Progress Report Week 1

Summary of Progress

Our team setup the IDE Pycharm and started learning python. Nolan starting building a small crawler to learn the scrapy library. We got more requirements from the client, which include the end users, the machine learning algorithms we're going to be using, and an outline of the data flow of the program. We also are beginning to work on the project plan and exploring libraries we see fit.

Individual Contributions

Jeremiah:

- Worked on setting up Pycharm so I can begin creating and running prototyping
- Working on learning Python and becoming familiar with it because we are planning on using it as the main language of this project
- Researching libraries for Machine learning and web crawling to have a better idea of what to use for our intended use cases

Alec:

- Completed setting up pycharm environment
 - Hello world works fine
 - Continuing to learn python programming language
- Explored Natural Language Processing links given to us by Benjamin
 - Learned about Stemming and Lemmatization
 - Will continue to explore this subject area
- Will explore scrapy and write a small webcrawler of my own. I plan to feed a NLP library outputs from my scraper so I can get experience with stemming and lemmatization.

Mark:

- Learning the basics of python,
- Set up a virtual env and Pycharm IDE
- Started to learn some of the libraries we plan to use.(Especially Scrapy)
- Plans to write a small test crawler in Python

Nolan:

- Learned to use Python and the PyCharm IDE.
- Started to familiarize myself with Scrapy.
- Wrote a small crawler that scrapes iastate.edu for pdf files and downloads them.
- Browsed some of the links sent to us by Ben Blakely.

What We're Doing Next

Along with continuing to familiarize ourselves with Python and the Scrapy library, we will be researching machine learning and natural language processing concepts such as stemming, lemmatization, and vectorization. We will also work on developing the models required for the project plan.