

What have you been able to accomplish so far?

I have been able to build a simple crawler that searches google for relevant information related to a certain person and returns clean text. I have also been able to leverage Named Entity Recognition tools from spaCy and Noun Phrase extraction tools from TextBlob to count the most frequent occurrences of certain names and phrases. I have also spent many hours to try to get Relationship Extraction models OpenNRE and RESIDE to work on my machine. They do work and train, but I have encountered numerous problems that have led me to decide against using them in my final project.

What libraries are you using? Why those ones?

I am using spaCy, TextBlob, Flair, WordNet, OpenNRE, RESIDE, BeautifulSoup, Google (python package), and a few other miscellaneous packages. All of these libraries are either used for me to extract data from online, clean the data, and then use various NLP techniques to find named entities, noun phrases, and other important considerations.

What were some unexpected challenges that you've faced? How have you overcome them?

The biggest challenges that I am facing is that existing relationship extraction technology has not yet advanced enough to be reliable or usable. The models that I have tried either have failed to train because they use all of my desktop's memory, they do not have any easy or accessible way to predict novel relationships (I spend much time digging into the code and trying to write a predict method with little success), and the inputs they require are difficult to replicate for my example. Because of these roadblocks, I am considering to instead use WordNet to find the attribute name for a certain value.

Are you about halfway done with what you proposed?

I am a little less than halfway done with my project because of all the challenges I encountered with the relationship extraction tools. I also had a busy couple weeks, but I just finished many of my tests/exams.

Do you think you can still deliver on what you proposed?

Definitely! I am confident that I will have plenty of time to continue to work on my project and flush out the problems I have encountered. I have several ideas for how to proceed.

Based on how much time you have spent on the project so far, how much time would you estimate you need to spend on it until you complete it?

Approximately another 20 hours should be enough.