Daily Log

Monday, September 30

I refined the `objects.txt` and `relations.txt` files by hand based on the frequency data, removing redundant words. These files list out all the abstract objects and relationships between any two objects in the problem dataset exactly once.

Tuesday, October 1

I researched the GEOS program referenced in the original paper and what it might do to help with the project with respect to its lexicon, methods, etc.

Thursday, October 3

I wrote a simple script to identify whether a word (possibly with a stem) was a word that corresponded to an object or relation in `objects.txt` or `relations.txt`, and identified them as such.

Timeline

Date	Goal	Met
9/16	Finish writing the webscraper to scrape arbitrarily many problems off of the forums, properly formatted. Begin the process of filtering posts from the dataset.	I was not able to get my webscraper to work, but I've successfully started to format the approximately 300 problems I actually have.
9/23	Filter posts that are not standard olympiad geometry problems, and construct a standard lexicon of keywords to look for in a problem both as objects and as relations.	Yes, I created two separate files for various geometrical objects and relations that I found that appeared in my data set.
9/30	Write code that creates a graph structure corresponding to the problem statement, with objects as nodes and relations as edges.	My code definitely is able to identify all the nodes/concepts and what kinds of relations are present in the problem, but not much more.
10/7	Create a logical language to give more structure and properties to the objects and relations detected in the problem.	N/A
10/14	Research how to write code to create a log-linear classifier and figure how to pass the proper structures detected in the problem statement to it.	N/A

Reflection

This week, I found myself diving into the heart of the problem for the first time. I re-read the original paper whose method I'm replicating in this project, just to be sure of the structure I wanted for the graph that the rest of the project builds upon. It turns out that their method (GEOS) relies on their own logical language of relations and objects that they used in order to label the concepts and endow with additional structure other than identifying it as a relation. This is something I didn't realize I had to create myself, so I tried to see if they published the code for the method online and hopefully look at how they structured their code. Unfortunately, I didn't find any piece of code related to their paper, and I think it'd be helpful for me to see it as a template for what my objects and relations should look like.

According to my timeline, I put down the creation of lexicon and the graph as two separate items to do, spanning 4 weeks of work. I think these two things are actually pretty related, and I'm tackling them together at this point. Technically, I already have code that finds the possible edges and vertices in my graph based on what words are in the problem, but I'm not sure at all that what I have is enough to move on to the next step. My timeline still allows me 2 weeks to figure out what additional structure I need to implement, and I want to use this time to build a more sound foundation for the model I need to train to detect geometric relations in a problem.