Daily Log

Monday, September 16

I debugged my script to ensure that the script could send a request and get a 200-code response. I created code to parse the problems being returned by the request.

Tuesday, September 17

The response from the site was a site that displayed an error message, so I played around with different cookies and options to try and make the script return the correct data. I talked to the sysadmins and fixed some minor issues in my code that I was not aware of previously, but still couldn't get the response to work.

Thursday, September 19

I tried fixing the webscraper for a little bit more time, but I gave up and moved on to try to clean up the problems I already had. I removed all problems that didn't appear to be standard (i.e. didn't have any point names or named objects) and cleaned up some of the remaining LaTeX formatting.

Timeline

Date	Goal	Met
9/2	Scrape at least 400 problems from the AoPS website, through brute force and operating on the Contest Collections.	No, as earlier years have a Short-list with only 3-4 relevant geometry problems apiece. I got in the neighborhood of 200-270 problems through manual scraping.
9/9	Scrape at least 1000 problems from the AoPS website – in particular, the High School Olympiads (HSO) forum.	No, but the scraper that I have is nearly functional. Up to scaling, my script can scrape an arbitrary amount of the problems on the page. I did not get many new problems this week.
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.	N/A
9/30	Write code creating a graph structure corresponding to the problem statement, with objects as nodes and relations as edges.	N/A

Reflection

This week was a bit of a failure when it came to accumulating more problems, despite using most of the means at my disposal to try and scrape more problems off of the AoPS website. I spent a lunch in the syslab on Tuesday getting their input while trying to fix the scraper, but I couldn't get it to function as I wanted to in the end. I have a suspicion that this idea was a bit of a lost cause to begin with, considering that they probably have good web devs to prevent the kinds of things I wanted to do to the site.

I moved on to take on the actual problem by establishing first the types of objects I'll be working with. I want to clean up and standardize the set of geometry problems I do have, which should all have named objects (points, lines, planes, shapes) and relations, which can be thought of "verbs" in our category between two items. I will need to create two "dictionaries" of these two kinds of terms in the coming week or so, and begin to identify these objects in relations in problems, which is the main problem that I have to tackle.