

Olivia Malcolmson
Nyah Harrison
Carl Ekholm
Amanda Anowi

Progress Report

Current Progress:

Since milestone 1, we have focused on collecting and gathering the remaining data we need for analysis, as well as processing/cleaning the data as we intend to display it graphically.

In terms of additional data collection and querying, we took our original list of domains and utilized the network activity tracking devtool feature to obtain a list of additional sites that were used to retrieve and load additional web content. However, parsing through every additional website tracked when visiting our 21 original domains would not be feasible, so we chose the first 5 domains that appeared and manually compiled this data into the topdomainsDev.txt file within the RawData folder. For each of these supporting domains, we ran traceroute and wrote the output to a text file for each of the original domains (i.e., the traceroute output for each supporting domain tracked by devtool feature is compiled into one file whose filename is the original domain queried). All of this data is stored in a separate folder titled TracerouteOutputDevTool. We also ran our get_location.py script on each of these traceroute outputs to compile geolocation information for each domain and stored the data in devtool_geolocations.txt. We parsed through the geolocations.txt data set to get a count of all of the unique locations, and how many domains were associated with each, as well as a count of the top 10 locations, and how many domains were associated with each. To organize the data and search for trends, the location information has also been compiled into two CSV files – one that tracks all intermediate locations when running traceroute for the original domains, and another for the supporting domains found using the devtools network activity tracking feature.

We also have begun to compile our information into graphs and maps that we aim to display on a poster. Some of the graphs have been created and we have set up a template for our poster, but the bulk of our analysis for the poster has yet to be completed.

Future Tasks:

In the last week of our project we hope to complete our analysis of the information that we have successfully collected over the past couple weeks as well as the results that we have concluded from the collected data. In order to do this we aim to complete a poster that contains different graphs as well as text that explains these graphs in order for people to easily read and understand with no prior knowledge of how networks work and what happens in the background when you log onto instagram.

How do our future tasks align with our original project plan:

Since our project aimed at figuring out what goes on behind the scenes when a person queries for their most commonly used websites, this meant that we wanted our analysis to be easy to understand for someone who does not have any networking experience. The easiest way to understand something for someone with no networking experience would be a simple easy format that you can read through quickly without feeling too heavy. Because of this we chose the poster idea with graphs that easily visually represent the results that we found from doing this project.