Subikhyat Pokharel

Dr. Nathan Green

IT-360-A

December 8, 2017

**Project Summary**

Net Neutrality is a principle that Internet service providers must treat all data on the Internet the same, and not discriminate or charge differently by user, content, website, platform, application or method of communication. And while the federal government had previously implemented this law, preventing ISP’s to do so the new government is attempting to get rid of this law.

My final project for data scraping on for the Topics in Information Technology class was to scrape the net neutrality subreddit and attempt to find answers to a few questions I had prepared. For scraping reddit I used PRAW, a python reddit API wrapper. To make processing easier and manageable I used the first 500 posts on the subreddit and save it to a text file for consistent data read. This list includes all the comments on each post and replies for each comment on each post. (Note: The posts I used were the newest 50 posts on the subreddit at the time of executing the script.)

One of the first questions I wanted answers for was the total number of links on the first 500 posts and after scraping the data and making a dictionary out of it, I found there was a total of 505 links on the subreddit at the date of scraping. Most of the links were repeated and to check the most common link I found that <https://www.govtrack.us/> which was repeated 50 times on the subreddit.

Another question I wanted to check was the number of times Ajit Pai’s name was repeated on the subreddit. Ajit Pai is the face of the government trying to remove the net neutrality principle, and not surprisingly his name was mentioned multiple times on the subreddit. Another part of this question was to see list words commonly used to in the same sentence, but reading the first few comments most of the comments were not safe for work and hence not appropriate for a project.

The last question I wanted to answer was the most frequently used word in the subreddit (exclusive of words such as the, a, your etc.) was FCC which was repeated a total of 208 times in the subreddit.

In conclusion, although it would be very interesting to see how this data compared to a different social media like twitter or Facebook, reddit is considered the face of the internet and hence, was a very reliable source to answer the questions I had prepared relating to net neutrality.