Eric Zhang
Title of paper: Detecting influenza epidemics using search engine query data

The technology discussed in this paper involves using search engine queries to detect influenza pandemics. The problem it addresses is the seasonal influenza that poses a threat to public health which has the potential to cause respiratory illnesses that could lead up to 500,000 deaths around the world. Being able to detect these influenza epidemics at an earlier time will enable a rapid response that can save more people from this epidemic. This technology addresses this through observing health seeking behavior through web search queries online. These web search queries come from the Google search engine. This assumption that the frequency of certain types of queries relates to the level of influenza in a region of the United States is based on the fact that there is a high correlation between certain kinds of queries and patients visiting physicians with symptoms similar to that of influenza. This way of detecting influenza is much quicker compared to traditional methods that the CDC and EISS use which involves using clinical and virologic data that has a one to two week reporting lag. Using search queries on the other hand, has a reporting lag of about one day. The researchers in this paper seek to build a system that uses much more search data compared to earlier attempts through the use of five years of search queries from Google.

One weakness of this paper is the lack of visual models. This makes it more difficult to understand some concepts such as how the data flows through the system to be processed. Including graphics would make understanding some of the more difficult concepts such as the use of linear models more easier. Another weakness is the lack of headers. The paper mostly uses general headers for the first couple pages such as “Results”, “Discussion”, and “Conclusion”. This makes it harder to quickly scan through multiple sections or search for specific sections relating to a topic. Other papers often have more specific headers which makes navigating the paper much easier. In terms of strength, the paper does a good job of explaining the overall goal and solution approach clearly. It also explains well how the probability of search queries relating to influenza-like illness is calculated.

Potential improvements could include using visual models to clearly show how data is processed through the linear model. This would allow for better understanding of how linear regression is used to predict the likelihood of physician visits based on search queries. Another potential improvement is having more specific headers that better explain the sections that they represent. This allows one to more easily find specific parts of the paper relating to a certain topic.