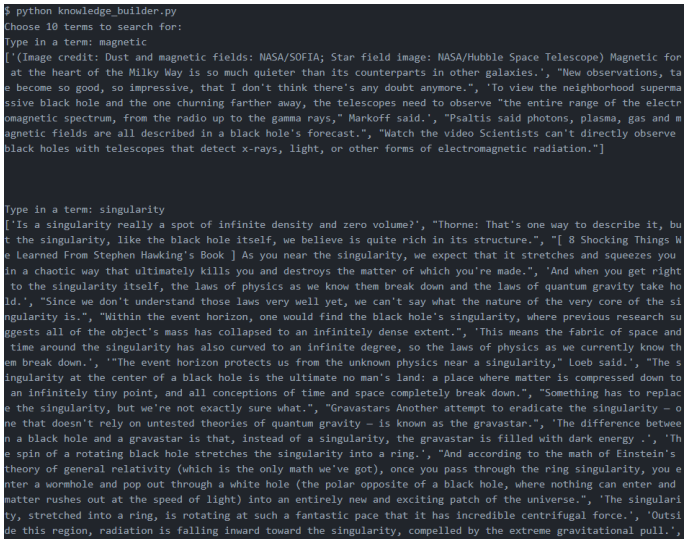
Kyle Chan kxc180021, Ryan Banafshay rcb170002

Web Crawler Report

Our knowledge base was created by using a web crawler to extract content from different web pages relevant to blackholes. That data was stored in separate text files. We analyzed the text using TF-IDF to extract the top 50 relevant words in the files. TF value gives a frequency score of items in a document and IDF gives information relative to rarity in a collection of documents. We multiply these terms together to extract what we consider to be the most important terms from each of the documents. Once we got the top 50 or so terms, we manually picked 10 of them to build our knowledge base around. The 10 terms we picked were magnetic, singularity, horizon, galaxies, interstellar, wormhole, gravastars, telescope, suns and shadow. Our knowledge base is represented through a python dictionary with values containing sentences relating to the term which we then pickled. Below are screenshots of our knowledge base.



If we develop a chatbot around this knowledge base it should be capable of giving sophisticated dialogue back to the user. An example of this could look like this:

User: What is another type of blackhole?
Chatbot: A wormhole is another well known phenomena similar to blackholes.
User: What is a wormhole?
Chatbot: A wormhole is different from a blackhole as a blackhole has an end with no escape known as a singularity. According to the math of Einstein's theory of general relativity (which is the only math we've got), once you pass through the ring singularity, you enter a wormhole and pop out through a white hole (the polar opposite of a black hole, where nothing can enter and matter rushes out at the speed of light) into an entirely new and exciting patch of the universe.