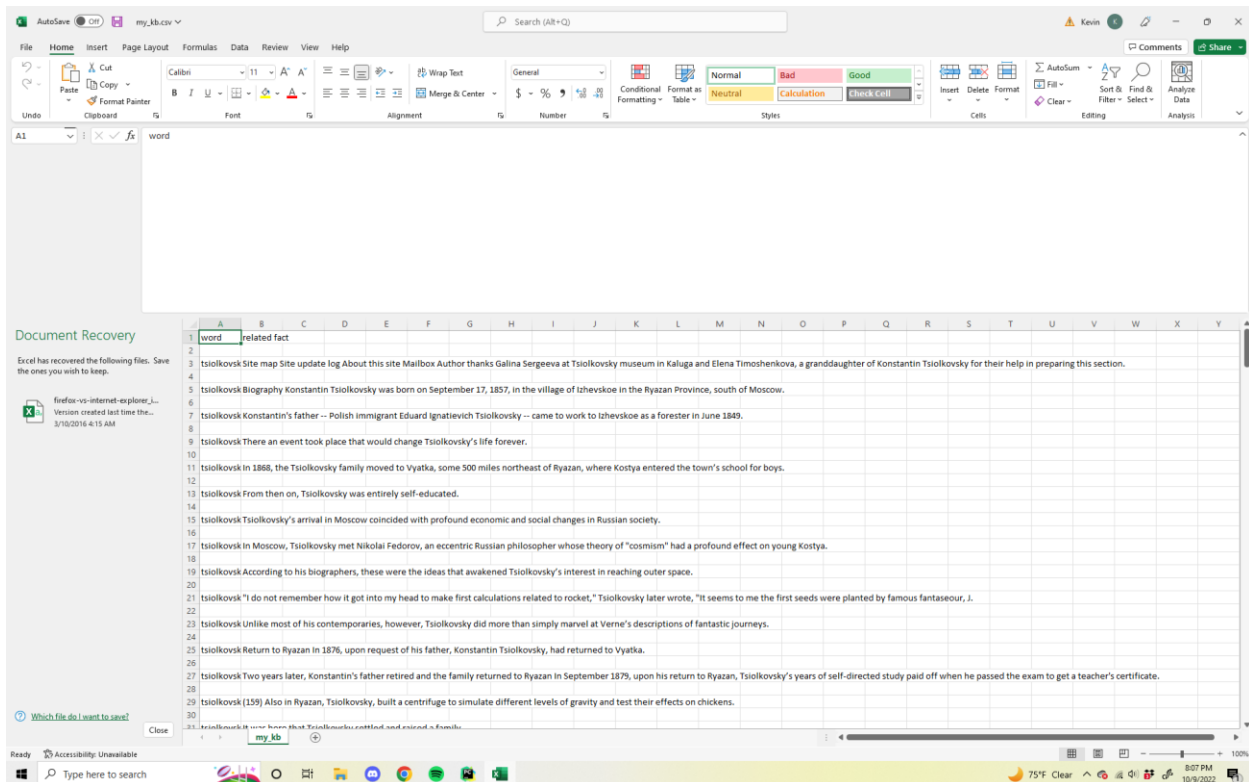


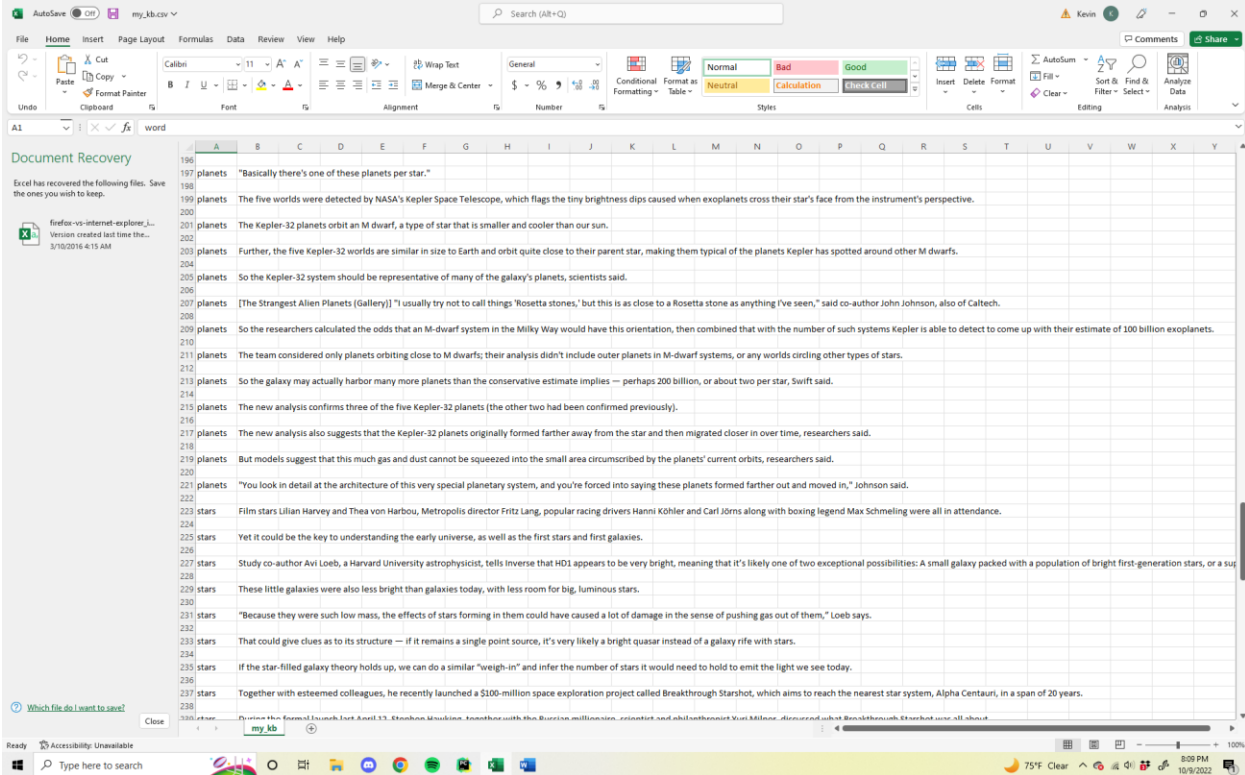
1. How I created my knowledge base

Process:

I created my knowledge base simply by using the sentences (in my corpus) that my top 10 words showed up in as facts that can be provided by my chatbot by searching things related to those 10 words. I determined my top 10 words by using calculating a tf-idf score for every term in every document and then just taking the maximum scores for each term. The problem with my tf-idf scores is that I got very high scores for words like 'enable' and 'disable,' because in one of my documents I had simply scraped 'Please enable JS and disable any ad blocker.' Because 'enable' only occurred in one document and has a very high term frequency for that particular 'error' document, its tf-idf score was very high relative to other more relevant words. As such, I had to manually prune some of the highly scored words to get more relevant ones.

Screenshots:





2. Sample dialog I would like to use

A sample dialog I would like to conduct with my chatbot might include:

Me: 'What is Mangalyaan?'

Computer: 'The spacecraft called "Mangalyaan," or "Mars-craft" in Hindi, which was launched last November, slowed down just enough to reach orbit early Wednesday, securing India a place in the elite global space club of Martian explorers.'

Me: 'When will Mangalyaan launch?'

Computer: 'Mangalyaan, which means "Mars craft" in Hindi is scheduled to lift off at 2:38 pm from Sriharikota, 80 kilometres from Chennai.'

Me: 'Which of the planets will Mangalyaan land on?'

Computer: 'The technology to reach nearby planets is possible, at least to make the first steps there.'
(*Unsure of the behavior if two words are used belonging to the knowledge base*).