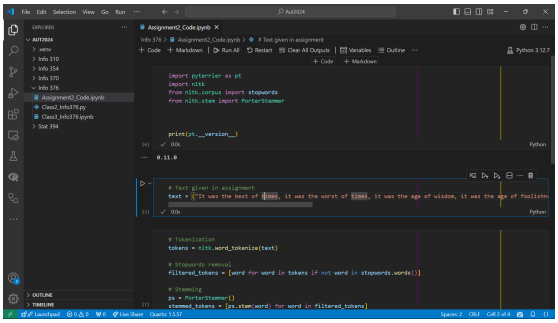
1. What is information retrieval? What is information filtering? How do you think these two relate or differ? Write in your own words.
    a. Information Retrieval (IR):
        i. This is finding “relevant” information based on a user searching/querying for the info. In this, the user has the agency and is looking for something they want.
    b. Information Filtering (IF):
        i. This is the process of narrowing down the user’s preferences and displaying results similar to/within that range. In this case, the user doesn’t have agency; instead, this is geared toward giving the user what they want without them explicitly searching for it.
    c. How do they relate/differ?
        i. IR is much more focused on one-time queries, as in finding information for the user at the time that they need it. Here, the user is active in accessing information.
        ii. IF is more focused on developing a profile for the user and keeping the user engaged in what they would like to be viewing. Here, the user takes a more passive role, letting the filtering algorithm deliver recommendations to the user based on what it knows about them.
        iii. Both are sides of the same coin, in the sense that both IR and IF are ways that the user can match data to user needs/wants.
2. **Report how you are ready to do coding assignments in this class**. You can show a screenshot of your Python+PyTerrier setup on your computer or show it working through Google Colab or some other online Jupyter notebook environment.
    a. 
    b. I will be using VSCode on my laptop.
3. What is indexing? Why is it needed for an IR system?

a. Indexing is essentially the process of creating a structured way to represent the information. This is done with a focus on keywords, terms, or metadata (from the Google article).
b. Indexing is vital because it is to locate and reference information promptly since it doesn't need to search the entire collection of information for what the user wants. Indexing also allows platforms to handle large amounts of information.

4. Consider the following text as a document. **Represent/index it as a variable/vector/frame** in Python after doing proper **tokenization**, **stopword removal**, and **stemming**:
   a. "It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair"
   b. ['it', 'time', ',', 'worst', 'time', ',', 'age', 'wisdom', ',', 'age', 'foolish', ',', 'epoch', 'belief', ',', 'epoch', 'incredul', ',', 'season', 'light', ',', 'season', 'dark', ',', 'spring', 'hope', ',', 'winter', 'despair']
   c. 

5. Now take the query "incredulity" and show how you can search through that representation (index) using **the exact match** that we saw last time.
   a.