Gregory Livingston
ITAI -2373
6-10-25

Reflection Journal – Lab 02

This lab really helped me understand how important it is to prepare text before using it in any kind of AI or machine learning project. I used to think computers could just read text the way we do, but now I know there are a lot of steps involved to make language clear for a machine. Just like how we clean up a messy room before using it, text has to be cleaned and organized so the computer can work with it properly. Learning this gave me a new appreciation for the work that happens behind the scenes in things like chatbots, translation apps, and search engines.

One big lesson I learned was that different tools handle language in their own ways. For example, when comparing NLTK and spaCy, I saw that they break words down differently. A contraction like Its might be split into two words in NLTK while spaCy treats it another way. This kind of small difference can lead to big changes in tasks like emotion analysis or chatbot conversations. It helped me see that every step in cleaning data must match the purpose of the system. Taking out a word like not can totally change the meaning of a sentence and mess up the results.

I also had a hard time at first understanding the difference between stemming and lemmatization. Both are used to make words simpler but they work differently. Stemming cuts off the ends of words without checking if the new word makes sense. For example the word studies turns into studi which is not a real word. Lemmatization does a better job by turning

studies into study which is a proper word. I learned that if your goal is accuracy and understanding meaning then lemmatization is the better choice.

Another thing I noticed was that stop word lists are not the same across different tools. NLTK had fewer stop words than spaCy and the kinds of words they removed were not always the same. I realized this can really change the outcome of a task. Words like not are important in phrases like not good because they change the message. If those words are removed the system could get the meaning wrong. This taught me that removing stop words should not be automatic. You have to think about the goal of the project.

I liked how this lab made connections to things I see every day. If I was working on customer reviews I would want to keep strong emotion words and punctuation. A sentence like This is AMAZING shows strong feeling and if you take out the capital letters or punctuation you lose part of that meaning. If I was building a search engine that needs to be quick I might use faster tools like stemming. But if I was building a chatbot that talks to people I would want to use smarter tools that keep the meaning and sound natural.

While doing the lab I started thinking about new questions. I wonder how these tools work with languages that follow different grammar rules than English. I also thought about how the type of text might change what stop words you should use. A tweet is very different from a formal document. I hope we talk more about these ideas in future lessons.

The part I found most useful was building and testing my own steps to clean text. I can now see how these tools can help in real jobs like sorting reviews or helping a chatbot talk better. I used to think cleaning text was just a starting task but now I know it is a key part of making smart language systems work the right way.