Exploring the Natural Language Toolkit

Eckel, TJHSST AI2, Spring 2023

Background and Explanation

Using AI algorithms to process large amounts of language is a fascinating and productive area of recent research. In this assignment, you'll use the Natural Language Toolkit to play with some of this kind of work.

This is a completely independent assignment; the goal here is to introduce you to a tool so you know enough about it that it's available for future passion projects or senior research topics. Mastery is not the plan, unless you're passionate.

Installing NLTK

There is an instructions page linked below; in my experience, take it with a grain of salt. For instance, it does not mention that it works with Python 3.9 or 3.10 but I ran a couple of tests and it seemed fine? But maybe I'll run into an incompatibility later – I don't know. The moral is: like any situation where you're installing a package for the first time, some experimentation and resilience in the face of annoying compatibility problems may be necessary. Feel free to experiment. The worst case scenario is that you have to reinstall Anaconda/Python, which isn't the end of the world.

If you run into technical difficulties, please reach out – in particular, if you have something specific to an unusual setup you may wish to ask in Town Square instead of asking me directly. Either way: plan to leave some extra time to solve any bumps in the road.

Extremely Important Tip!

It turns out that since the nltk package is in a file called "nltk", if you name YOUR file "nltk.py" you can't import nltk in it.

So: PLEASE NAME YOUR FILE SOMETHING THAT ISN'T "nltk.py". If you make this mistake and then message me about it, I reserve the right to mildly mock you.

Specification for 1 RED Credit

Complete all of this work to get full credit.

1) Go here: https://www.nltk.org/book/ and install NLTK by typing `pip install nltk` at your command prompt of choice. Read the introduction to the book and Chapter 1. You'll need to **follow the instructions that begin Chapter 1 to get started**; skim the rest / do whatever seems interesting.
2) In Chapter 2, complete exercises 4, 5, 7, 9, 12, 17, 18, 23, and 27.
3) In Chapter 3, complete exercises 20 and 22.
4) In Chapter 6, complete exercise 4.
5) Judiciously truncate any long output (ie, just copy/paste the first and last few outputs with a "…" in between) and put all of your answers into a document.

Submit the document described to the link on the course website (doc, docx, or pdf format please).

This assignment is **complete** if:

- You follow the instructions on the submission form to format your submission properly.
- You submit a document containing the above required exercises.

Specification for 2nd RED Credit and BLACK Credit

You can get a second RED credit and continue further for BLACK credit as well for doing some of the more difficult questions.

Each chapter, at the end, has several questions marked with a filled in black 5-pointed star. Call these **advanced questions**. Your goal here is to find advanced questions **that you are interested in** and complete them. (Important caveat: this assignment can absolutely be gamed; if you search for the easiest to code advanced questions it is possible to blow this off and get nothing out of it. I hope that if you're in a situation where you feel like your only option is to do make-work like that you instead come to me and tell me that's how you feel and we figure out something more interesting for you to do instead!)

To get RED credit, complete **five additional advanced questions**, including **one from a chapter not covered in the required assignment. Chapter 1 doesn't count.** For each question, answer it in your document **and then add 1-2 sentences about why that question interested you and what you learned.**

To get BLACK credits, complete **ten additional advanced questions**, including **questions from two chapters not covered in the required assignment. Chapter 1 doesn't count**. For each question, answer it in your document **and then add 1-2 sentences about why that question interested you and what you learned.**

If you have any confusion about this, let me know!

When you submit your document, **add a summary at the top of the first page** that summarizes whether you're going for RED or BLACK credit, all the questions you did and what chapters they came from, and anything else you'd like me to know as I grade your work. For example (and **these are not suggestions, I just randomly picked some numbers**):

> For RED and BLACK credits I did:
>
> - Chapter 2, questions 26, 27, and 28
> - Chapter 4, questions 27, 30, and 33 (one new chapter)
> - Chapter 8, questions 29, 30, 33, and 35 (a second new chapter)
>
> (Insert any special notes you'd like me to be aware of while grading here)
>
> …then put all your answers

Submit a document containing a clear summary like the one above, and then your answers for the Red/Black questions only (**not** including the required questions) to the link on the course website.

This assignment is **complete** if:

- You follow the instructions on the submission form to format your submission properly.
- You submit a document containing a sufficient number of black-star advanced questions as specified above.
- Each question is accompanied by a genuine explanation of why you found that question interesting and what you got out of it.