

Exploring the Natural Language Toolkit

Eckel, TJHSST AI2, Spring 2021

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 get really 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 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.

For the record, your Anaconda install should be Python 3.8 anyway. The "base" environment already comes with nltk installed; in any other new environment, just go to "Environments" and search "Uninstalled" for "nltk" and you should find it no problem.

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.

Required Tasks

Complete all of this work to get full credit.

- 1) Go here: <https://www.nltk.org/book/> and here, if you don't use Anaconda: <https://www.nltk.org/install.html> and familiarize yourself with NLTK, install it, and 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.

Extremely Important Tip!

It turns out that since the nltk package is in a file called "nltk", if you name YOUR file "nltk.py" in which you're planning to do this lab, the nltk won't work.

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

Submit the document described on the front page to the link on the course website.

This assignment is **complete** if:

- The “Name” field on the Dropbox submission form contains your **class period**, then your **last name**, then your **first name**, in that order.
- You submit a document containing the above required exercises.

For **resubmission**:

- Complete the specification correctly.

Specification for Outstanding Work (1 or 2 OW credits)

You can get Outstanding Work credit for exploring further and 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**.

To get 1 OW credit, complete **five additional advanced questions**, including **one from a chapter not covered in the required assignment**. **Chapter 1 doesn’t count**.

To get 2 OW credits, complete **ten additional advanced questions**, including **questions from two chapters not covered in the required assignment**. **Chapter 1 doesn’t count**.

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 one or two OW credits, 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 2 OW 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 OW questions only (**not** including the required questions) to the link on the course website.

This assignment is **complete** if:

- The “Name” field on the Dropbox submission form contains your **class period**, then your **last name**, then your **first name**, in that order.
- You submit a document containing a sufficient number of black-star advanced questions as specified above.

For **resubmission**:

- Complete the specification correctly.