Rachel Shanblatt Bloch
March 8, 2020

LING 83800 - Homework 2

Initially, I had a very hard time getting started with this assignment. I was not familiar with the argparse module prior to this assignment, so familiarizing myself with the documentation, watching youtube tutorials, and completing the practicum exercises was quite time consuming. Only once I had a better understanding of the module could I begin working on my program.

Once I started coding, it came to me pretty easily. The main roadblock I hit writing the code was figuring out how to split the data properly. Initially, I thought my program split the data properly, but I had actually only set those variables to the *amount* of data (0.8, 0.1, etc) that I wanted in each set. Once I realized that and tried to fix it, my program then *did* split the data - but some of the data overlapped in the sets. After some reworking, I managed to split the data properly using indices.

My next roadblock came with trying to test my program from the command line. When I used the testing line from the assignment page, I got an error - 'No such file or directory: 'conll2000.tag'. I couldn't figure this out: I knew the file was in the right place, what was I doing wrong? I reached out to a classmate who suggested I try adding .txt to the end of the file name - and that worked. So simple yet so frustrating.

I used all of the tests recommended on the assignment page, and my program passed all tests as expected. The files were created as expected, the lengths of the files were as expected, and the SHA-256 checksums did not change after running the script over the input file again using the same seed.

All in all, this was an extremely challenging assignment for me: not due to the assignment itself, but rather because of my total unfamiliarity with this module. Learning about argparse and reviewing the documentation took a lot of time, and I found it difficult to understand. Once I understood argparse better and overcame the couple roadblocks along the way, the assignment itself was quite reasonable and straightforward to complete.