The Write-Up for Homework 2

In this homework, I basically had to deal with three problems and tried to find some solutions by googling.

The first problem was to split the text based on percentages into training, development and testing groups. To be honest, I couldn't find a way though I tried many codes that I found some websites such as stackoverflow. So, I did this splitting part manually by calculating 80%, 10% and 10% of the text, respectively.

The second problem was to save the output training, development and testing files as a list. When I wrote the following code, it gave the error "write() argument must be str, not a list":

With open((args.train), "w") as output:
        Output.write((train))

Firstly, I simply turned the code into:

With open((args.train), "w") as output:
        Output.write(str(train))

However, when the output file was a string, there was another problem that I couldn't manually check the length of the output file by using the code:

wc -l train.tag dev.tag test.tag

When I omit the -l part and enter the code it was giving me the length of the files but based on strings, not lists. So, I found the following code to save it as a list:

with open(args.train, 'w') as filehandle:
        for listitem in train:
            filehandle.write('%s\n' % listitem)

After saving it as a list, I could also manually check the length of the output file by using the code above.