Below questions are from the end-of-chapter 01 questions of the book Natural Language Processing With Python, by Steven Bird et. el. Please review chapter 1, and answer the questions on a Jupyter Notebook, and submit your work on the Magnimind platform.

<https://www.nltk.org/book/ch01.html>

10. ○ Define a variable my\_sent to be a list of words, using the syntax my\_sent = ["My", "sent"] (but with

your own words, or a favorite saying).

a. Use ' '.join(my\_sent) to convert this into a string.

b. Use split() to split the string back into the list form you had to start with.

19. ◑ What is the difference between the following two lines? Which one will give a larger value? Will this be the case for other texts?

>>> sorted(set([w.lower() for w in text1])) >>> sorted([w.lower() for w in set(text1)])

20. ◑ What is the difference between the following two tests: w.isupper() and not w.islower()?

21. ◑ Write the slice expression that extracts the last two words of text2.

22. ◑ Find all the four-letter words in the Chat Corpus (text5). With the help of a frequency distribution (FreqDist), show these words in decreasing order of frequency.

25. ◑ Define sent to be the list of words ['she', 'sells', 'sea', 'shells', 'by', 'the', 'sea', 'shore']. Now write code to perform the following tasks:

a. Print all words beginning with *sh*.

b. Print all words longer than four characters

26. ◑ What does the following Python code do? sum([len(w) for w in text1]) Can you use it to work out the average word length of a text?

27. ◑ Define a function called vocab\_size(text) that has a single parameter for the text, and which returns the vocabulary size of the text.

28. ◑ Define a function percent(word, text) that calculates how often a given word occurs in a text and expresses the result as a percentage.