15CSE358 Natural Language Processing
Assignment-1

Total Marks: 25

1. Write programs to process the Brown Corpus and find answers to the following questions:
   a) Which nouns are more common in their plural form, rather than their singular form? (Only consider regular plurals, formed with the -s suffix.)
   b) Which word has the greatest number of distinct tags. What are they, and what do they represent?
   c) List tags in order of decreasing frequency. What do the 20 most frequent tags represent?
   d) Which tags are nouns most commonly found after? What do these tags represent?

Reference: https://www.nltk.org/book/ch05.html [5 Marks]

2. One common way of defining the subject of a sentence `S` in English is as *the noun phrase that is the child of* `S` *and the sibling of* `VP`. Write a function that takes the tree for a sentence and returns the subtree corresponding to the subject of the sentence. What should it do if the root node of the tree passed to this function is not `S`, or it lacks a subject?

Reference: https://www.nltk.org/book/ch08.html [5 marks]

3. Write a function that takes a grammar (such as the one defined in 3.1) and returns a random sentence generated by the grammar. (Use `grammar.start()` to find the start symbol of the grammar; `grammar.productions(lhs)` to get the list of productions from the grammar that have the specified left-hand side; and `production.rhs()` to get the right-hand side of a production.)

Reference: https://www.nltk.org/book/ch08.html [5 marks]

4. Implement CKY algorithm for parsing a CFG in CNF.

Reference: https://www.youtube.com/watch?v=yFLvZXGRiSE [10 marks]

Deadline: 06-10-2021, 05:00 PM