NLP Homework 4
Total Points - 100

Due Date: Nov 04, 2022

1. [**Points 30**] Given CFG grammar below.

S→ NP VP	1.0	S = Sentence, VP = Verb Phrase NP = Noun Phrase PP = prepositional phrase DT = determiner Vi = intransitive verb Vt = transitive verb NN = noun IN = preposition
VP → Vi	0.4	
VP→ Vt NP	0.4	
VP → VP PP	0.2	
NP→ DT NN	0.3	
NP→ NP PP	0.7	
PP → IN NP	1.0	
Vi → sleeps	1.0	
Vt → saw	1.0	
NN→ man	0.7	
NN → woman	0.2	
NN→ telescope	0.1	
DT → the	1.0	
IN → with	0.5	
IN → in	0.5	

i) Parse the sentence using the above grammar and show parse tree. [**Points 10**]

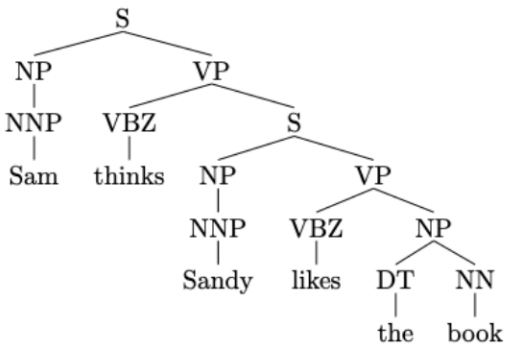
Sentence: ***The man saw the woman with the telescope***

(ii) Show probabilistic parse tree with probability at each node in the tree for the following sentences [**Points 20**]

Sentence 1: ***The man saw the woman with the telescope***

Sentence 2: ***The man sleeps***

2. [**Points 10**] Show bracketed notation for the following tree.



3. [**Points 15**] Consider the grammar G given by:

   S→ AB | XB
   T → AB | XB
   X → AT
   A → a
   B→ b

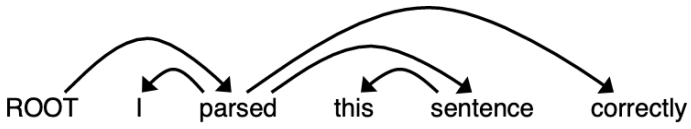
   Validate your answer using CKY algorithm.

   a) Is w = ***aaabb*** in L(G)? [**Points 7.5**]
   b) Is w = ***aaabbb*** in L(G)? [**Points 7.5**]

4. [**Points 20**] How are strings stored in spacy? Describe the document similarity checking mechanism in spacy? How would you compare similarity between two documents using spacy? Please write down an example code for document similarity checking. What are the spacy pipeline component initiated while using default nlp(). How would you add a custom component in the spacy nlp pipeline? Write a sample code which utilizes only two component tokenizer and add custom component in npl pipeline.

5. [**Points 25**] Go through the sequence of transitions needed for parsing the sentence "***I parsed this sentence correctly***".
   The dependency tree for the sentence is shown below.

   a) At each step, give the configuration of the stack and buffer, as well as what transition was applied this step and what new dependency was added (if any). [**Points 15**]



   b) A sentence containing n words will be parsed in how many steps (in terms of n)? Briefly explain in 1-2 sentences why. [**Points 10**]

Submission Instructions:
Please note that rephrasing or rename variables, functions names does not mean you are not copying/cheating. For this similar situation, you will be graded zero.

You should not zip your submission.

Late submission or Extension: Late HomeWorks/assignment will not be accepted unless an extension is approved by me in advance. Requests for extensions must be made at least three days before the due date with valid reason. **3 points** will be deducted for each day after the submission deadline from your grade even if you are approved for extension. For details, please see the **Homework and Exam Policies** section of your syllabus for more details.

Grading Policy/Rule: Copying/cheating/plagiarism is strictly prohibited as mentioned in our introductory lectures and syllabus. This policy holds for each assignment/homework/exam. In case of copying/cheating/plagiarism etc.

you will be graded zero for the assignment as well as 'F' for the subject. Note that the first incident of cheating will result in the student getting a final grade of 'F' for the course. The second incident, by CCSE rules, will result in a semester suspension from the College.