Homework 4A

Nolan McCafferty

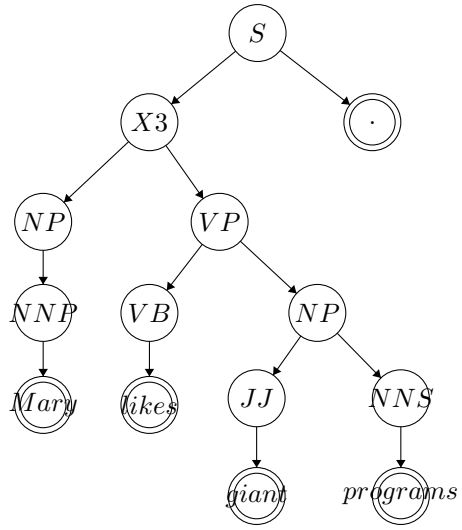
27 February 2019

1. Read!
2. `Mary likes giant programs .`

   (a) Our CYK chart from parsing the sentence.
   The bolding signifies that constituent is part of the final parse tree.

`Mary`	`likes`	`giant`	`programs`	`.`
NP: -0.15 + -0.2=-0.35 NNP: -0.2			**X3:0.0+-1.7+ -0.35=-2.05**	S: -10 **S:-0.1+-2.05+ 0.0=-2.15**
	VB: -0.6		**VP:-0.3+-0.6+ -0.8=-1.7**	S: -0.6
		JJ: 0.0	**NP:-0.8+0.0+ 0.0=-0.8**	
			NNS: 0.0	
				.: 0.0

   (b) Final Parse



3. Our structure will be somewhat of a one-way tree as each cell will know where it came from (parents), but now how or even if it is used later in the tree (children). Each entry in our CKY table will store the following:

   Each possible constituent with its associated weight. This could be stored in a hashtable with the constituent as the key and the weight as the value. The weights would be the sum of all the weights of the rules used before it. If the same constituent could be added to the same cell, then it will only keep the one with the higher associated weight.

   A pointer to the previous rule Thus, the class would have a `getPrevious` method to represent the rule that came before it in order to reconstruct the parse tree.

4. Since we will be implementing our algorithm with Hashtables, all of our lookups will be *O(1)*. Thus when we iterate over the 50,888 rules it will be much faster than iterating over all possible pairs of the two entry cells, since on average that will be $466 * 466 = 198,916$.

   We can also see that this algorithm will have less total lookups. The faster algorithm will be iterating through all of the binary rules and seeing if the corresponding parts of the RHSs match with *entry1* and *entry2*. On average there will be $446 * 2$ RHSs that we will be looking for so there are about $50,888 * 892 = 45,392,096$ total lookups.

   The alternative is much less efficient. On average there will be 446*446 = 198,916 possible combinations to check. For each one of these we will then have to go through all 50,888 binary rules to find a match. That equates to a maximum of 10,122,437,408 lookups.