Team:

Ashish Jain 1PI11IS020

Akshay Shah 1PI11IS009

Anusha B Rao 1PI12IS403

Lokesh Jain 1PI11IS054

About the Program:

Given:

We were given the training file where we replaced the words with count of less than 5 with \_RARE\_. This file was then written to the training tree and the new counts of the words were generated in the file parse\_train\_count.out.

Three dictionaries were created for non-terminals, unary and binary rules were counts of each production were listed.

The functions “q\_unary” and “q\_binary” calculate the probabilities of each production.

The CKY function starts off with the base case where probablity values might be the terminal itself or zero(0). The recursive part of the algorithm computes only the pi value of each production and calculates the maximum. The base pointer(bp) for each maximum production value is also calculated.

Parts left to be done:

Filling the data json tree format.