

ASSIGNMENT_4
NATURAL LANGUAGE PROCESSING (CSE 5XX)
PAVAS HANDA (MT16040)

The assignment implements CKY parser which can be run by following command :

python Assignment_4.py

Basically the parser is designed in following steps:

- Context free grammar of ATIS is loaded which is not in CNF format, that the CKY requires. In the initial step, the CFG is converted to CNF by applying rule of breaking large productions into productions with rhs of size 2 with both non-terminals, ("2.txt" representing productions after removal of large productions). Further, unit productions are removed such as $A \rightarrow B$ and $B \rightarrow 'c'$ with $A \rightarrow 'c'$.
- After getting final CNF, the algorithm of CKY is applied onto the grammar storing all the productions in the matrix relevant to corresponding word in sentence diagonally, in same format all the cells in matrix are filled. In order to get all the parse trees, back_tracker keep track of all the productions with left and right child.

final_count.txt represents the sentence seperated counts and parse_trees_sentence_2, parse_trees_sentence_3 gives the parse trees of respectives.