

22046 – Nelson Escalante

23384 – Eliazar Canastuj

Proyecto 2 – Algoritmo CYK

Repositorio en Github:

https://github.com/can23384/Proyecto2_TC

Introducción:

El objetivo del proyecto es implementar el algoritmo CYK (Cocke-Younger-Kasami) para verificar pertenencia de oraciones al lenguaje de una CFG y construir el árbol de derivación correspondiente, además de realizar mediciones del tiempo del algoritmo.

Diseño de la aplicación:

Tecnologías utilizadas:

- **Lenguaje:** Python 3.x
- **Librerías estándar:** time, copy, re.
- **Terceros:** matplotlib.

Diseño general:

La gramática se define mediante un diccionario y sobre esta gramática se realiza una serie de pasos para simplificarla y convertirla a la Forma Normal de Chomsky (FNC). En orden: se eliminan las producciones épsilon, luego las unitarias y los símbolos inútiles. Después, se completa la conversión a la FNC.

Una vez en la FNC se aplica el algoritmo CYK para determinar si una cadena pertenece o no a esta gramática. Primero se construye una tabla donde se guardan los no terminales que pueden generar una subcadena, y una tabla para backpointers. Para cada posición i de la cadena w , se agrega el estado A a la tabla cuando existe una regla para A que produce w_i . Luego, para subcadenas de longitudes 2 o mas, se consideran las particiones apropiadas y las reglas de no terminales que producen 2 no terminales, y si ambos no terminales pertenecen a la tabla, entonces el no terminal principal también pertenece. La cadena pertenece a la gramática si S pertenece a la tabla en el índice $[0][n-1]$. Luego de este proceso, se reconstruye el árbol a partir de los datos guardados en la tabla de backpointers; el árbol se exporta en una imagen mediante la librería matplotlib.

22046 – Nelson Escalante

23384 – Eliazar Canastuj

Discusion:

Obstaculos:

Entre los obstáculos que nos encontramos, uno de los principales fue el asegurarnos que la FNC fuera correcta, ya que el algoritmo funciona sobre esta gramática y cualquier error en esta es trasladada a errores en el algoritmo CYK. Tambien, cuando hay ambigüedades pueden haber dificultades con la reconstrucción del árbol con backpointers.

Recomendaciones:

Una de nuestras recomendaciones es extender la gramática agregando mas palabras y mas categorías para así tener mayor cobertura. Tambien podríamos agregar manejo de palabras desconocidas en el algoritmo CYK.