

NER con CRFs

Mario Alejandro Payan Viafara
1224601

Modelos

Para los modelos se utilizó un diseño en cascada, donde cada modelo genera su correspondiente salida a partir del modelo anterior.

Modelo 1

El modelo 1 posee una nueva columna con respecto al modelo base en la cual se identifica si identifica si la palabra de la primera columna empieza en minúscula o mayúscula y colocando un 0 o 1 respectivamente en la nueva columna. Para ello, se separan todos los elementos de una fila del documento y, teniendo en cuenta que la primera columna corresponde a la palabra, se identifica si es mayúscula o minúscula, luego, en un archivo de salida se escribe la línea analizada con el dato adicional, y así por cada línea del documento.

Modelo 2

A partir de aquí, el trabajo es muy similar, ya que se realizan los mismos pasos tras cada modelo pero modificando de manera distinta la columna nueva, este modelo a diferencia del primero usa un documento con tags para poder identificar si una palabra es común, propia, o simplemente no es un sustantivo.

Modelo 3, 4 y 5

Similares tanto en funcionalidad tanto como en código, cada uno utiliza el archivo generado por el modelo anterior para agregar una nueva columna, correspondientemente al orden mencionados, estos archivos utilizan el documento de tags para generar el género de la palabra, si es singular, plural o invariable, y la persona gramatical.

Conclusiones

Se observa en las tablas de desempeño que la precisión alcanzada por cada uno de los modelos es bastante alta, seguida del recall y por último el F-score, al analizar los porcentajes entregados después de hacer el testeo se puede notar que la mayoría de los porcentajes no descienden del 75% para cada una de las etiquetas combinadas de IOB lo cual los hace resultados bastante buenos de manera general.