

Práctica 3

Procesamiento de Lenguaje Natural
Facultad de Ingeniería, UNAM

Realizar un reconocimiento de Entidades Nombradas (NER) a partir de un modelo secuencial y con el corpus 'ner_dataset.csv' que se proporciona.

Se deberán seguir los siguientes pasos:

1. Obtener las sentencias a partir del csv. En este archivo, se indica cada inicio de sentencia con 'Sentence: n'. Se cuenta con 1000 sentencias que conformarán el corpus de entrenamiento y evaluación.
2. Preprocesar los datos.
3. Separar los datos en corpus de entrenamiento (70 %) y corpus de evaluación (30 %).
4. Entrenar un modelo secuencial a partir del corpus de entrenamiento. Deberán definirse los hiperparámetros.
5. Evaluar el desempeño del sistema a partir del corpus de evaluación y con la métrica de Exactitud (Accuracy).
6. Ejemplificar el reconocimiento de entidades nombradas con 5 sentencias del corpus de evaluación.

Puntos a evaluar

1. Entrega a tiempo del trabajo.
2. Código bien realizado y, principalmente, comentado adecuadamente.
3. Propocesoamiento y separación adecuada de los datos (entrenamiento 70 % y evaluación 30 %).
4. Haber desarrollado adecuadamente el modelo secuencial: buena selección de hiperparámetros, manejo adecuado de las sentencias y las palabras.
5. Evaluación adecuada (cumplimiento de los puntos 5 y 6).