

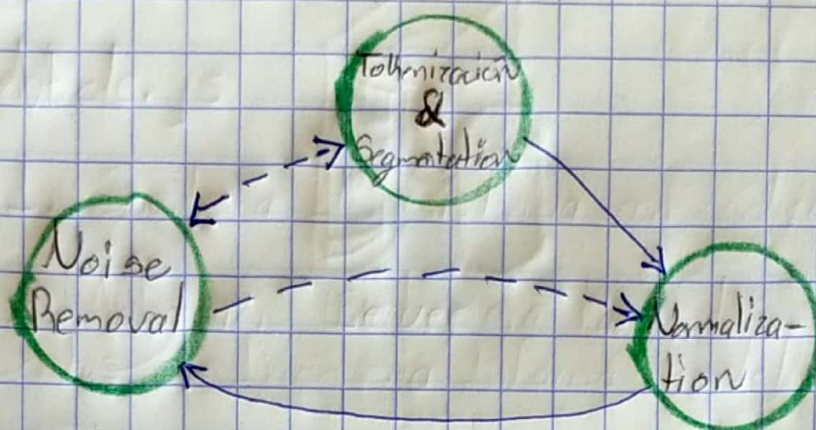
Conferencia Mag.
13:00 - 14:00

23 11 22

Ling. Juan Pablo Osuna Bonilla
Tema: "Pre-procesado, Tokenización y obtención de Embedding de textos en el procesamiento del lenguaje natural"

Institución: "Universidad Autónoma de Manizales-Colombia"

- > El objetivo es pasar de lo que describiremos como un trozo de texto (que no debe confundirse con fragmentación de texto), una cadena única larga y sin procesar.
- > Un conjunto de datos de juguetes de hecho, los pasos que se ocupan aquí para preprocesar estos datos son totalmente transferibles.



- > Dada su estrecha relación con textos específicos y su recopilación y montaje, muchas tareas de eliminación de ruido, como el análisis sintáctico de una estructura JSON.