

5.2 Proceso Creación Chatbot con Python - Caso Práctico

Chatbot Servicio Crucero



CORPUS = Base de Conocimiento (FAQs, normas, descripción de productos/servicios, soluciones técnicas,...)

1. Definición del CORPUS

2. Preprocesamiento del Texto con NTLK **CORPUS**

2. Preprocesamiento del Texto con NTLK **MENSAJE DEL USUARIO**

Tokenizar
Eliminar Palabras de Parada
Derivación
Lematización

3. Evaluar Similitud **CORPUS** - **MENSAJE DE USUARIO**

4. Definición de Coincidencias Manuales

Saludo,
agradecimiento,
despedida

Palabras muy frecuentes (artículos, preposiciones,...) comienzan a dominar en el documento (mayor puntuación), pero es posible que no contengan tanto **contenido informativo**.

- **TF-IDF** (Frecuencia de documentos de frecuencia inversa a término): Escalar la frecuencia de las palabras según la frecuencia con la que aparecen en todos los documentos/textos. El resultado tras aplicarlo a un texto es un vector del valor real en el espacio vectorial.
- **Similitud de coseno** es una medida de similitud entre dos vectores distintos de cero. Se usará para encontrar la similitud entre las palabras ingresadas por el usuario y las palabras en el corpus.