

## 5.- Lenguaje Natural

viernes, 1 de marzo de 2024 12:13 a. m.

**NLTK (Natural Language Toolkit):**

Propósito: NLTK es una biblioteca ampliamente utilizada para el procesamiento de lenguaje natural y la enseñanza de la lingüística computacional.

Funciones: Ofrece una variedad de herramientas para el análisis de texto, incluyendo tokenización, lematización, etiquetado gramatical, análisis sintáctico, y más.

**spaCy:**

Propósito: spaCy es una biblioteca NLP moderna y eficiente que se utiliza para realizar tareas de procesamiento de lenguaje natural.

Funciones: Ofrece capacidades de tokenización, lematización, etiquetado gramatical, reconocimiento de entidades nombradas, análisis de dependencias y más. Además, es conocida por su velocidad y eficiencia en el procesamiento de grandes volúmenes de texto.

**TextBlob:**

Propósito: TextBlob es una biblioteca de procesamiento de lenguaje natural basada en NLTK y Pattern, diseñada para la facilidad de uso y la implementación rápida de aplicaciones de NLP.

Funciones: Proporciona una interfaz sencilla para realizar tareas de NLP como tokenización, análisis de sentimientos, traducción de texto, corrección ortográfica, y más.

**Gensim:**

Propósito: Gensim se utiliza principalmente para el modelado de tópicos y el procesamiento de textos de gran escala.

Funciones: Ofrece implementaciones eficientes de algoritmos para el modelado de tópicos, como LDA (Latent Dirichlet Allocation), así como funcionalidades para la construcción de modelos de vectores de palabras (word embeddings) utilizando técnicas como Word2Vec.

**Transformers (anteriormente conocido como Hugging Face Transformers):**

Propósito: Esta biblioteca es especialmente conocida por su implementación de modelos de aprendizaje profundo pre-entrenados para tareas de NLP.

Funciones: Proporciona acceso a una amplia gama de modelos de lenguaje pre-entrenados, como BERT, GPT, RoBERTa, entre otros, para tareas como clasificación de texto, generación de texto, traducción automática, y más.

**StanfordNLP:**

Propósito: StanfordNLP es una biblioteca que ofrece modelos pre-entrenados para el procesamiento de lenguaje natural desarrollados por el Grupo de Procesamiento de Lenguaje Natural de la Universidad de Stanford.

Funciones: Proporciona herramientas para tokenización, lematización, etiquetado gramatical, análisis sintáctico, reconocimiento de entidades nombradas y más, con modelos de alta calidad.

**Pattern:**

Propósito: Pattern es una biblioteca de Python que ofrece herramientas para el procesamiento de texto y el análisis de datos web.

Funciones: Incluye funcionalidades para tokenización, lematización, análisis de sentimientos, clasificación de documentos, extracción de información, y más. Además, proporciona herramientas para acceder y analizar datos de la web

Investigar:

POS Tagging <--- Se usa con otro

Stemming

Lematization