

DESARROLLO DE PROYECTO I

MAESTRÍA EN CIENCIAS DE LOS DATOS

CENTRO UNIVERSITARIO DE CIENCIAS ECONÓMICO-ADMINISTRATIVAS

LEONARDO DANIEL PANTOJA MORFIN
VÍCTOR TÉLLEZ GARCÍA

“NLTK (NATURAL LANGUAGE TOOLKIT)”

FECHA 1 DE NOVIEMBRE DE 2023

ÍNDICE

- Introducción
- Aplicaciones de Natural Language Processing
- Aplicaciones en Ciencia de Datos
- Librería NLTK
- NLTK vs spaCy

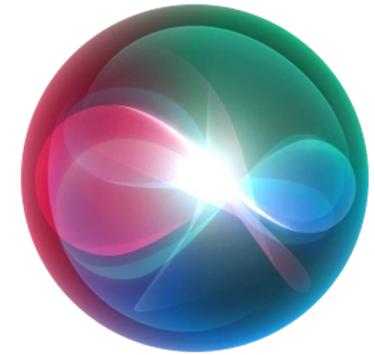


INTRODUCCIÓN

- NLP (Natural Language Processing), es una interdisciplina entre la computación y la lingüística. Se trata de darle a las computadoras la habilidad de manipular textos, usando algunas bases como reglas o probabilidad (estadística o redes neuronales), métodos de “machine learning”. El objetivo es que la computadora sea capaz de “entender” los contenidos de los documentos, extraer información de los documentos y categorizarla.
- Los principales retos que hay en ese subcampo es el reconocimiento de discurso, entendimiento del lenguaje natural y la generación de lenguaje natural.

APLICACIONES DE NLP

- Generación de texto
- Traducción
- Chatbots y Asistentes Virtuales
- Reconocimiento de Voz
- Análisis de Intención
- Resumen Automático



APLICACIONES EN CIENCIA DE DATOS

Extracción de Información y Entidades:

- Extracción de información de documentos no estructurados.
- Identificación de entidades clave (por ejemplo, nombres de personas, fechas, ubicaciones).

Clasificación y Categorización:

- Clasificación de documentos en categorías predefinidas.
- Categorización de temas o tópicos en grandes conjuntos de texto.

Análisis de Sentimiento y Opiniones:

- Análisis en comentarios y reseñas.
- Minería de opiniones en redes sociales y sitios web.
- Detección de Fraude y Seguridad

LIBRERÍA NLTK

NLTK es un conjunto de librerías y programas para el procesamiento de NLP en Python, se puede hacer con estas:

- Clasificación
- Etiquetado
- Análisis y razonamiento semántico.

Fue desarrollada por Steven Bird y Edward Loper en el departamento de Computación e Informática en la universidad de Pensilvania en 2001.

El objetivo de NLTK es dar soporte a investigadores y académicos en NLP y sus áreas relacionadas como lingüistas, ciencias cognitivas, inteligencia artificial y “machine learning”.

LIBRERÍA NLTK

“[Natural Language Processing with Python](#)” es el libro que crearon los autores para introducir de una manera práctica la programación para procesadores de idiomas, el libro habla de los fundamentos:

- Programar en python
- Categorizando texto
- Analizando estructura lingüística

NLTK

Search

NLTK Documentation

API Reference

Example Usage

Module Index

Wiki

FAQ

Open Issues

NLTK on GitHub

Installation

Installing NLTK

Installing NLTK Data

More

Release Notes

Contributing to NLTK

NLTK Team

Documentation

Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to **over 50 corpora and lexical resources** such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active **discussion forum**.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more. The online version of the book has been updated for Python 3 and NLTK 3. (The original Python 2 version is still available at https://www.nltk.org/book_1ed.)

Some simple things you can do with NLTK

Para más información [NLTK](https://www.nltk.org)

EJEMPLOS DE USO

1. Tokenización
2. Eliminación de palabra vacías
3. Etiquetado de partes de la oración (POS tagging)
4. Reconocimiento de Entidades Nombradas (NER)
5. Reducción de palabras y lematización
6. Clasificación de Texto
7. Concordancia y colocaciones
8. Acceso a corpus
9. Recursos de idioma
10. Visualización de texto

SPaCy VS NLTK

Ambas son librerías que tratan con NLP hechas para Python, pero difieren en el acercamiento que usan en su metodología, haciendo que ciertas tareas o procesos sean más sencillos que otros.

Siendo **NLTK** más antigua que **spaCy**, cuenta con una amplia gama de recursos y herramientas lingüísticas, lo cual resulta útil para investigaciones más especializadas. Sin embargo, esto también puede implicar una mayor dificultad para nuevos usuarios, lo que hace de **spaCy** una opción más accesible.

Adicionalmente, **spaCy** está diseñada para ser veloz y eficiente gracias al manejo de palabras basado en objetos. Pero, esto puede resultar contraproducente al momento de querer trabajar con los resultados.

spaCy

spaCy is an object-oriented library where we need to download packages related to English text.

It Provides the most efficient NLP algorithm for a given task. So if you care about only the end result, you can go with spaCy. It is quick and can perform every task easily, making it quite popular among many developers.

spaCy is a new library and has a very active user community amongst developers

NLTK

NLTK is mainly a string processing library.

NLTK provides access to many algorithms, and if you care about specific algorithms and customizations, then you can go with NLTK.

Usually, Researchers prefer to use NLTK in their work.

Extraído de [seaflex](#)

REFERENCIAS

- Bird, Steven, Edward Loper and Ewan Klein (2009).
Natural Language Processing with Python. O'Reilly Media Inc.
- <https://www.seaflux.tech/blogs/NLP-libraries-spaCy-NLTK-differences>
 - <https://ts2.space/en/nltk-vs-spacy-which-is-better-for-nlp/>

¡GRACIAS!