

ANÁLISIS DE TEXTO Y NUBE DE PALABRAS

Examen Tema 3 - MIA

Autor:	Julio García
Asignatura:	MIA - Modelos de Inteligencia Artificial
Tema:	Examen Tema 3
Fecha:	29 de enero de 2026

1. INTRODUCCIÓN

Este informe presenta los resultados del análisis de frecuencia de palabras realizado sobre un texto de dominio público descargado desde Project Gutenberg. El objetivo es identificar los términos más utilizados y representarlos visualmente mediante una nube de palabras.

Objetivos:

- Descargar un texto de dominio público (Historia Natural y Moral).
- Limpiar y procesar el texto usando Python y NLTK.
- Calcular la frecuencia de aparición de las palabras.
- Generar una visualización de nube de palabras.

2. METODOLOGÍA

Fuente de datos: Se utilizó el texto "Historia Natural y Moral de las Indias" obtenido de Project Gutenberg.

Sitio web de Project Gutenberg:

The screenshot shows the Project Gutenberg homepage with a navigation bar at the top. Below it, a breadcrumb trail indicates the path: Project Gutenberg > 77,621 free eBooks > 2 by José de Acosta. The main content area displays the title 'Historia natural y moral de las Indias (vol. 1 of 2) by José de Acosta'. To the left is a thumbnail of the book cover, which is dark green with gold lettering. Below the cover are social sharing icons (Twitter, Facebook, etc.) and a QR code. The right side of the page contains a table titled 'Read or download for free' with various file formats and their sizes.

How to read ?	Size	?	?	?
Read now!	750 kB			
EPUB3 (E-readers incl. Send-to-Kindle)	795 kB			
EPUB (older E-readers)	802 kB			
EPUB (no images, older E-readers)	497 kB			
Kindle	912 kB			
older Kindles	775 kB			
Plain Text UTF-8	625 kB			
Download HTML (zip)	591 kB			
There may be more files related to this item.				

Descarga del archivo de texto:

Index of /files/70219

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory		-	
 70219-0.txt	2023-03-06 12:34	625K	
 70219-0.zip	2023-03-06 12:34	223K	
 70219-h.zip	2023-03-08 06:48	590K	
 70219-h/	2023-03-08 07:20	-	

Apache Server at www.gutenberg.org Port 80

Procesamiento del texto:

- Conversión a minúsculas.
- Eliminación de signos de puntuación y números.
- Tokenización (división en palabras).
- Filtrado de stopwords (palabras vacías) en español.
- Filtrado adicional de palabras cortas (longitud <= 3).

Herramientas utilizadas:

- Python 3.x
- NLTK (Natural Language Toolkit) para limpieza y stopwords
- WordCloud para generar la visualización
- Matplotlib para gráficos
- Collections.Counter para conteo de frecuencias

3. RESULTADOS

3.1. Visualización: Nube de Palabras

A continuación se presenta la nube de palabras generada, donde el tamaño de cada palabra es proporcional a su frecuencia de aparición en el texto.



3.2. Análisis de Frecuencias

El análisis de frecuencias revela los términos más importantes del texto. Se han eliminado artículos, preposiciones y otras palabras comunes para resaltar el contenido semántico.

4. CONCLUSIONES

El análisis de frecuencia de palabras ha permitido identificar los conceptos centrales del documento. La visualización mediante nube de palabras ofrece una forma rápida e intuitiva de comprender la temática del texto sin necesidad de leerlo completo. Las librerías de Python utilizadas demostraron ser eficientes para el procesamiento de lenguaje natural y la generación de gráficos.