



Procesamiento de Lenguaje Natural

Santiago Pineda Quintero

Grupo de Control y Procesamiento Digital de Señales (GCPDS)
Dep. de Ing. Eléctrica Electrónica y Computación (DIEEC)
Facultad de Ingeniería y Arquitectura (FIA)
Universidad Nacional de Colombia sede Manizales

Contenido



UNIVERSIDAD
NACIONAL
DE COLOMBIA

- 1 Introducción
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones

Contenido



- 1 **Introducción**
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones

¿Qué es PLN?

Enseñar a las máquinas procesar y analizar lenguaje humano
(texto y audio).



¿Qué es un modelo de lenguaje amplio (LLM)?



Contenido



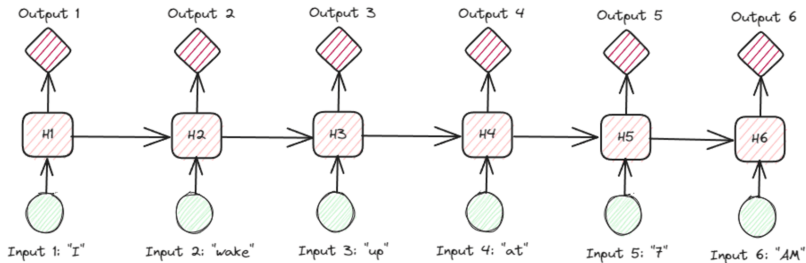
UNIVERSIDAD
NACIONAL
DE COLOMBIA

- 1 Introducción
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones

Redes Neuronales Recurrentes (RNN)



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Problemas de las RNN



UNIVERSIDAD
NACIONAL
DE COLOMBIA

- **Son Lentas:** procesan elemento por elemento de la secuencia de texto.
- **Pérdida de Memoria a Largo Plazo:** no pueden codificar relaciones a largo plazo entre los elementos de una secuencia.

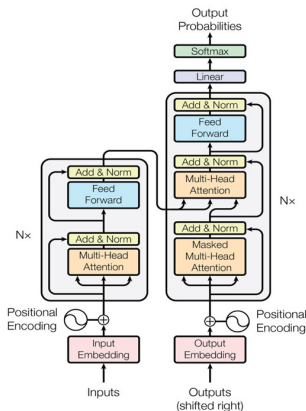


Arquitectura Transformer



UNIVERSIDAD
NACIONAL
DE COLOMBIA

¡La Arquitectura detrás de ChatGPT!

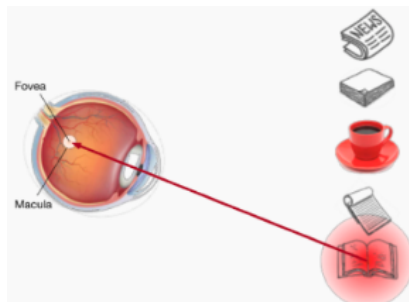


¡Atención!: La magia detrás de un Transformer



UNIVERSIDAD
NACIONAL
DE COLOMBIA

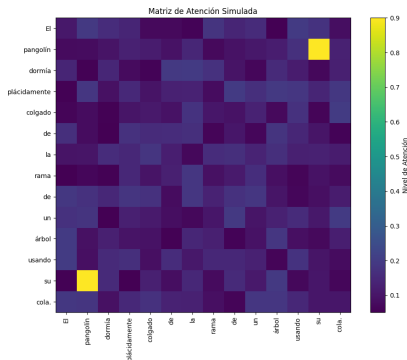
La atención es un mecanismo crítico en el aprendizaje profundo que le permite al modelo concentrarse en las partes más relevantes de la entrada para la tarea en cuestión.



Matriz de Atención



UNIVERSIDAD
NACIONAL
DE COLOMBIA



¡Los transformers a través de su mecanismo de atención resuelven los dos problemas de las RNN!

Contenido



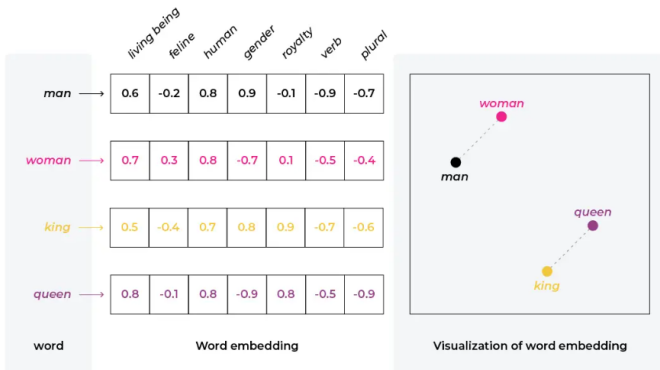
UNIVERSIDAD
NACIONAL
DE COLOMBIA

- 1 Introducción
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores**
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones

Embedding



UNIVERSIDAD
NACIONAL
DE COLOMBIA

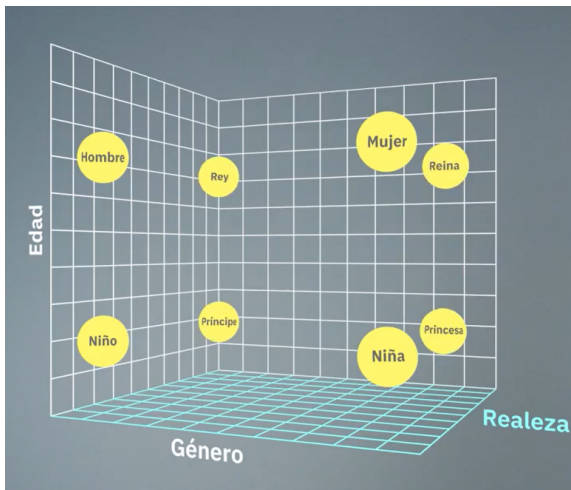


$$\mathbb{X} \in [0, 1]^d \quad (1)$$

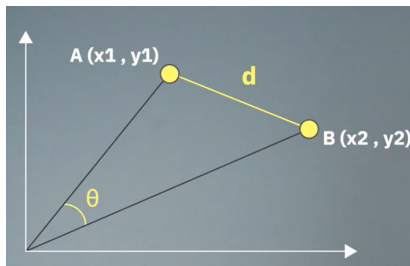
Espacio Dimensional



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Medidas de Similitud Semántica



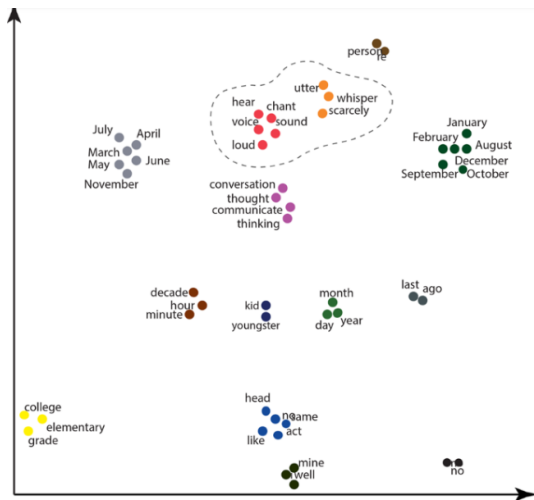
$$d = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \quad (2)$$

$$\theta = \cos^{-1} \left(\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \right) \quad (3)$$

Clusterización de Palabras



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Contenido



UNIVERSIDAD
NACIONAL
DE COLOMBIA

- 1 Introducción
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones

Tareas clásicas de NLP con LangChain



UNIVERSIDAD
NACIONAL
DE COLOMBIA



Contenido



UNIVERSIDAD
NACIONAL
DE COLOMBIA

- 1 Introducción
- 2 Arquitecturas de Deep Learning para el Procesamiento de Texto
- 3 Modelos Tokenizadores
- 4 Tareas Clásicas de Procesamiento de Lenguaje Natural
- 5 Conclusiones**

Conclusiones



- El procesamiento de lenguaje natural es clave para transformar datos textuales en conocimiento accionable mediante modelos avanzados como LLMs.
- Técnicas como embeddings y arquitecturas RAG facilitan la integración de recuperación y generación de texto para soluciones más precisas e inteligentes.
- Los agentes conversacionales y las tareas clásicas de PLN tienen aplicaciones prácticas en sectores como servicio al cliente, análisis de datos y automatización.



¡Gracias!

Santiago Pineda Quintero

email: spinedaq@unal.edu.co