PROCESAMIENTO DE LENGUAJE NATURAL

ESTRUCTURA DE LA CLASE

INTRODUCCIÓN

Presentación del tema de la clase

Introducción al NLP

PROCESAMIENTO DEL LENGUAJE

Evolución del NLP

N-grams

Tokenización

Vectorización

Métricas ROUGE

ATENCIÓN

El problema de las redes recurrentes

Atención

Transformadores

Grandes modelos de lenguaje

GPTs

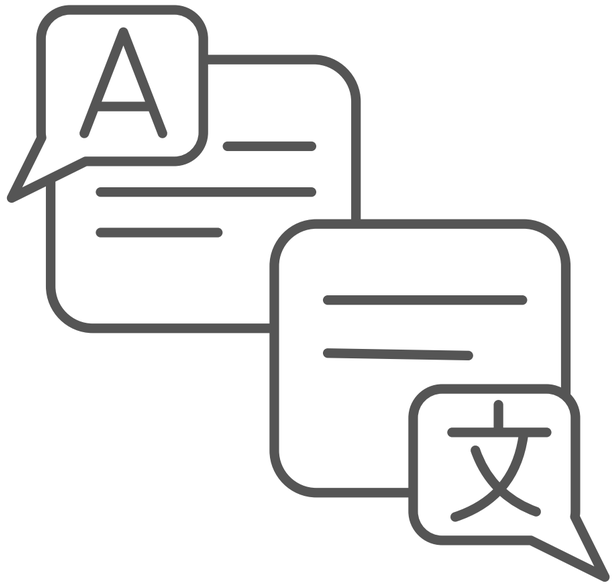
ACTIVIDAD PRÁCTICA

Traducción de inglés a español

CONCLUSIONES

Recapitulación de los puntos clave de la clase

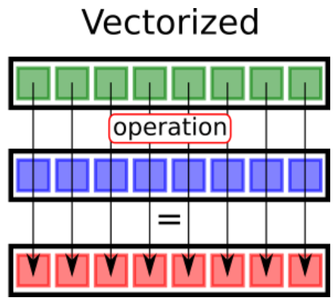
EL LENGUAJE

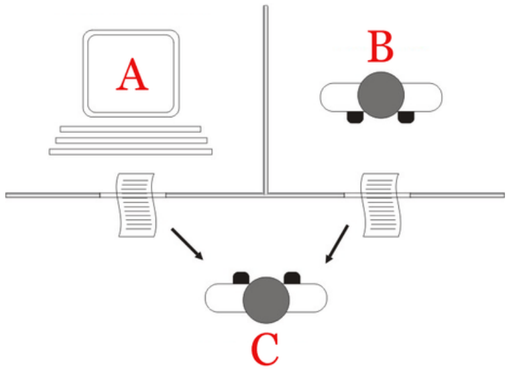


EVOLUCIÓN DEL NLP

Diccionarios y análisis sintáctico

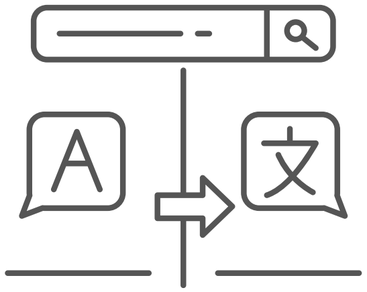
Vectorización





Test de Turing

Análisis estadístico



LLMs

N - GRAMAS

Esto	es	una	oración

N=1 Esto / es / una / oración

N=2 Esto es / es una / una oración

N=3 Esto es una / es una oración

Ya no se usan tanto como antes, pero siguen siendo una buena forma de análisis de estilo y métrica de evaluación

TOKENIZACIÓN

Esto	es	una	oración

Dividir la frase en una cantidad de partículas llamadas “tokens”, que pueden ser caracteres, palabras, subpalabras e incluso frases

TIPOS DE TOKENIZACIÓN

BASADA EN PALABRAS

Divide el texto en palabras separadas por espacios o signos de puntuación

["Hola", ",", "eres", “genial”, "."]

BASADA EN CARACTERES

Divide el texto carácter por carácter, incluyendo espacios y puntuación.

['H', 'o', 'l', 'a', ',', ' ', 'e', 'r', 'e', 's', ' ', 'g', 'e', 'n', 'i', 'a', 'l', '.']

BASADA EN SUBPALABRAS

Divide el texto en subpalabras o unidades más pequeñas que las palabras completas.

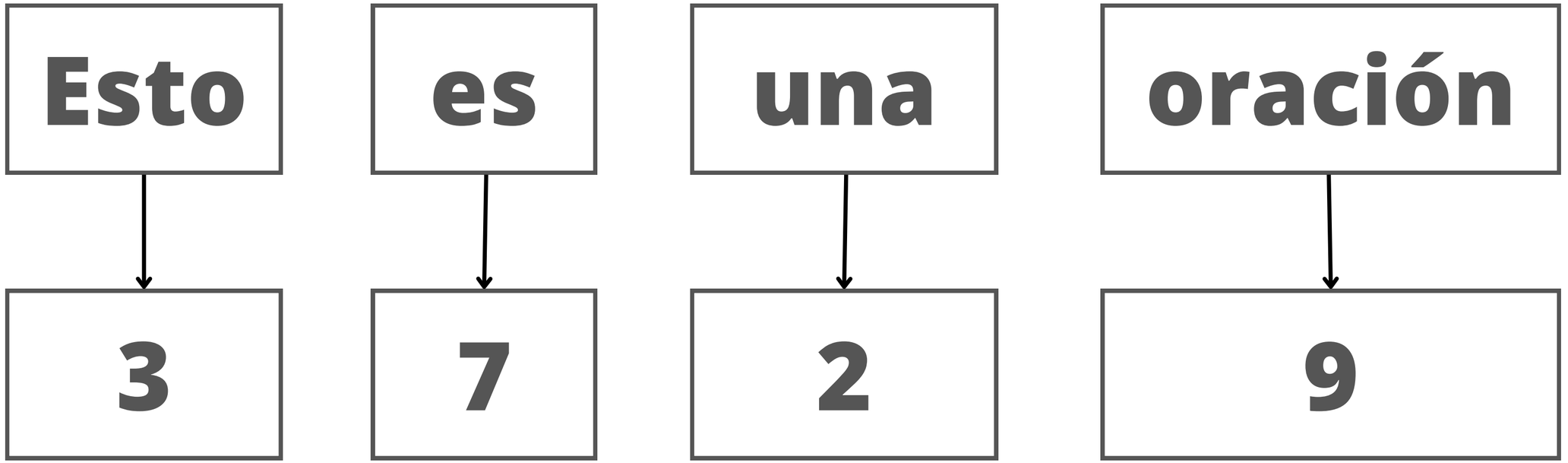
["des", "composición"]

BASADA EN FRASES

Divide el texto en frases o sentencias.

["Hola,", "eres genial."]

VECTORIZACIÓN



Convierten los tokens a un vector

TIPOS DE VECTORIZACIÓN

BOLSA DE PALABRAS (BOW)

Representa la frecuencia de cada token. Ignora su orden

TF-IDF

Calcula la importancia de una palabra en un documento, basada en su frecuencia en relación con todos los documentos

REPRESENTACIONES VECTORIALES

Convierte los tokens en vectores densos (Word embeddings) en un espacio de alta dimensión.

REPRESENTACIONES VECTORIALES CONTEXTUALES

Dependiendo del contexto un mismo token puede ser representado por otro vector

ESTRUCTURA DE UNA RED RECURRENTE

CELDAS

Unidad básica de procesamiento que guarda un estado oculto calculado de las salidas anteriores

CAPAS

Agrupaciones del mismo conjunto de celdas

FUNCIÓN DE ACTIVACIÓN

No linearidad aplicada para que la red aprenda patrones complejos

FUNCIÓN DE PÉRDIDA

Mide la diferencia entre la salida predicha de la red y la salida real.

ALGORITMO DE OPTIMIZACIÓN

Cómo se actualizan los pesos de la red

MÉTRICA DE ROUGE-N

Esto	es	una	oración

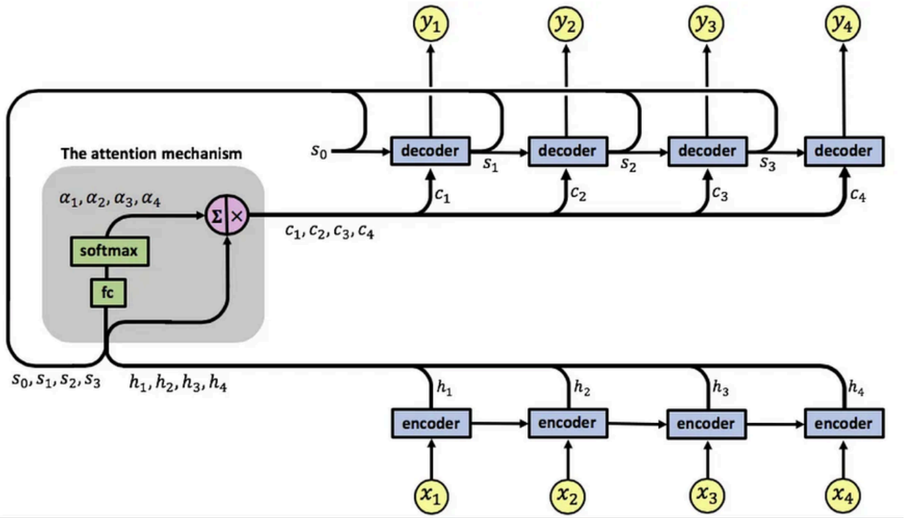
Esto	es	una	canción

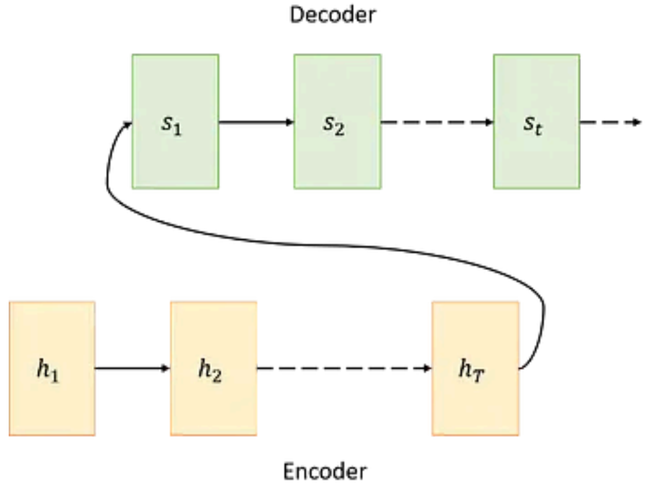
Mide la superposición de n-gramas

PROBLEMAS DE ATENCIÓN

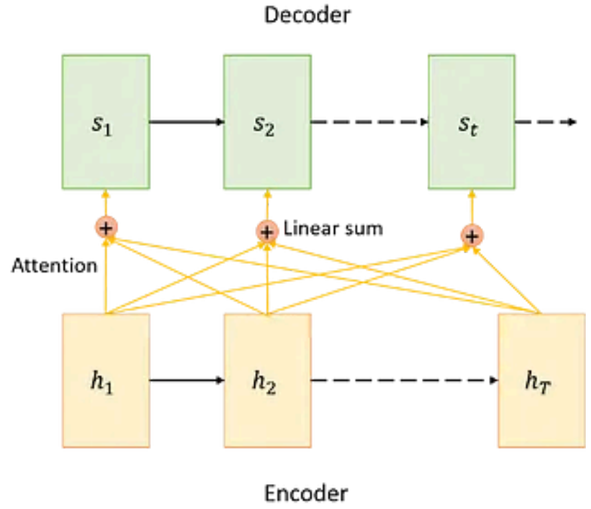
UN ELEFANTE SE BALANCEABA SOBRE LA TELA DE UNA ARAÑA

ATENCIÓN





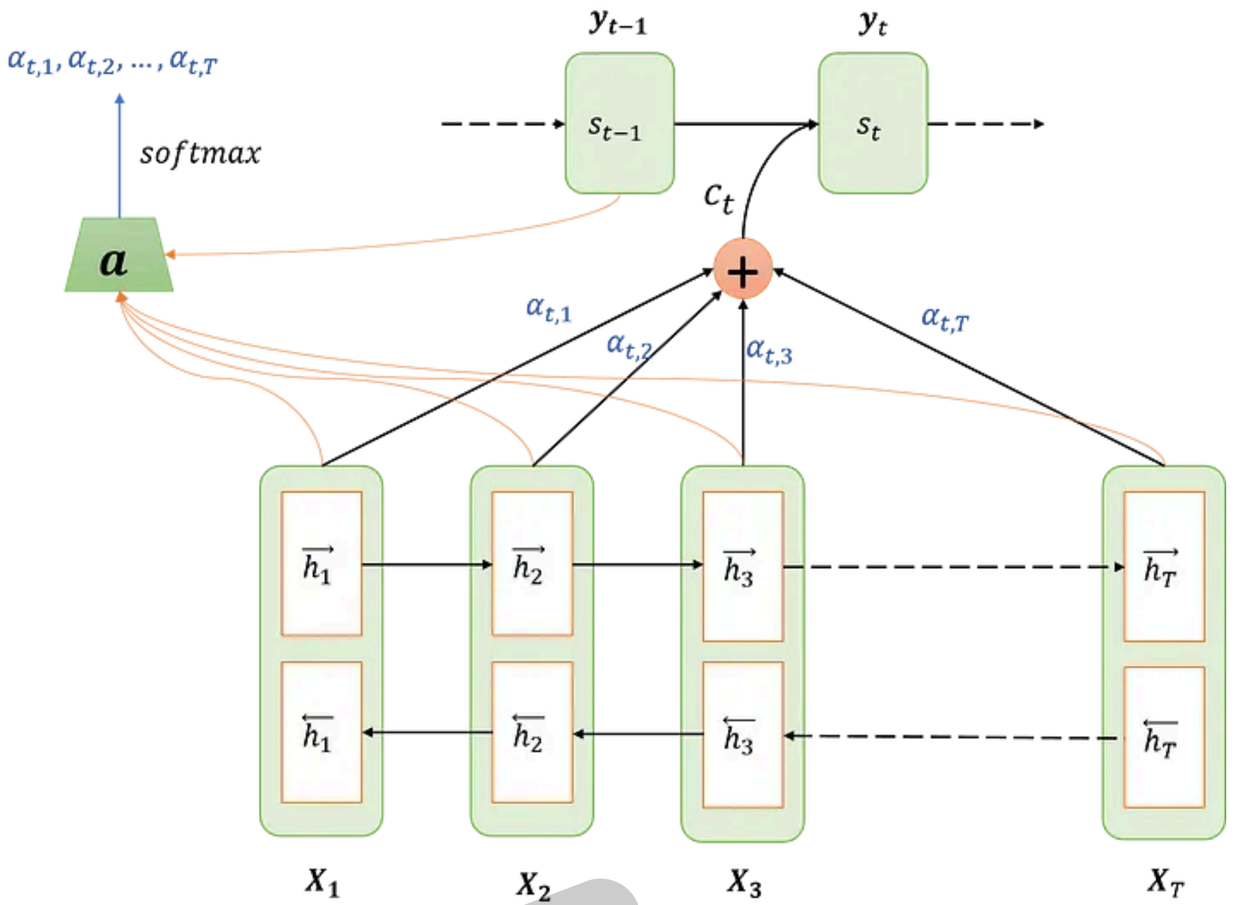
(a) Vanilla Encoder Decoder Architecture



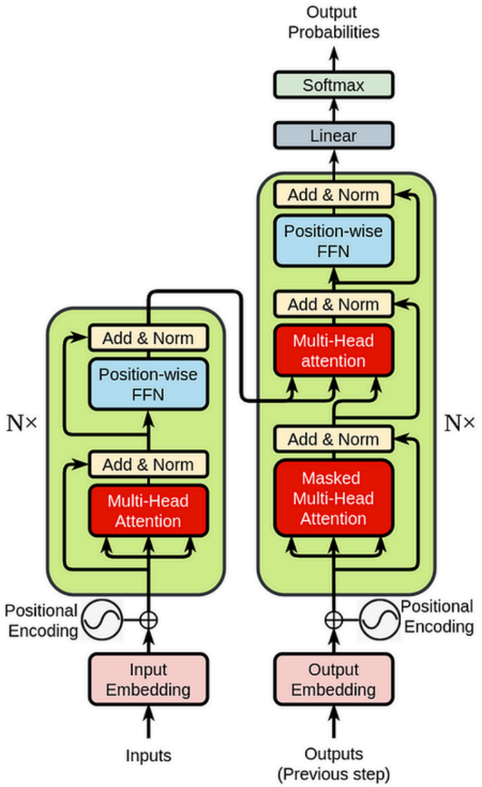
(b) Attention Mechanism

Pasa información del estado oculto del encoder al decoder y “aprende” la importancia de cada uno

ATENCIÓN EN RRN

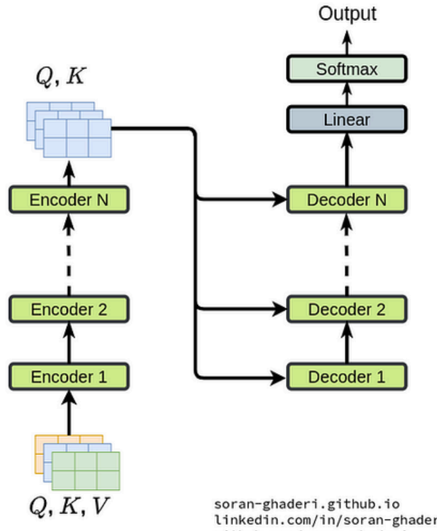


TRANSFORMERS



TRANSFORMER

```
pip install transformerx
github.com/tensorops/TransformerX
```



soran-ghaderi.github.io
linkedin.com/in/soran-ghaderi
github.com/soran-ghaderi

Sigue siendo un encoder decoder, pero se elimina la capa recurrente y se remplaza con capas de atención

La atención Multi-head a diferencia de las redes recurrentes permite el cómputo en paralelo

A su vez, la atención permite retener información en secuencias más largas

Al no ser secuenciales sino paralelas, pueden fallar en tareas de repetición, identificación de patrones y dependencias

LLMS

Son modelos entrenados en cantidades inmensurables de datos de texto

- **Generación de texto**
- **Análisis de texto**
- **Traducción de texto**
- **Chatbots**

Un transformer entrenado para la generación de texto se denomina GPT

CONCLUSIONES

TOKENIZACIÓN Y VECTORIZACIÓN

La tokenización separa el texto en unas unidades definidas y la vectorización convierte esas unidades a valores númericos entendidos por la computadora.

ATENCIÓN

Mecanismo que permite a las redes neurales poder tomar importancia a ciertos elementos de entrada

TRANSFORMERS

Modelo basado en atención, que remplaza la capa recurrente con una capa de autoatención que le permite operar de manera paralela