

# POS Tagging y Named Entity Recognition

Fundamentos de NLP

# Agenda

- 1. ¿Qué es POS Tagging?
- 2. Algoritmos para POS Tagging
- 3. ¿Qué es Named Entity Recognition?
- 4. Algoritmos para NER
- 5. Comparación POS vs NER
- 6. Aplicaciones Prácticas
- 7. Herramientas y Librerías

# ¿Qué es POS Tagging?

- Definición: Proceso de asignar categorías gramaticales a cada palabra
  - Objetivo: Identificar el rol sintáctico de cada palabra en una oración
  - Categorías: Sustantivo, Verbo, Adjetivo, Adverbio, etc.
  - Importancia: Base para análisis sintáctico y semántico
- 
- Ejemplo: 'El gato corre rápidamente'
  - El → Determiner (DT)
  - gato → Noun (NN)
  - corre → Verb (VBZ)
  - rápidamente → Adverb (RB)

# Desafíos del POS Tagging

- Ambigüedad Léxica: Una palabra puede tener múltiples roles
- Dependencia del Contexto: El significado cambia según el contexto
- Palabras Desconocidas: Vocabulario fuera del entrenamiento
- Variaciones Lingüísticas: Diferentes dialectos y estilos
- Ejemplos de Ambigüedad:
  - 'book' → Verb: 'Book a flight' vs Noun: 'Read a book'
  - 'running' → Verb: 'She is running' vs Adjective: 'Running shoes'

# Algoritmos para POS Tagging

- Hidden Markov Models (HMM)
  - Usa probabilidades de transición y emisión
  - Algoritmo Viterbi para decodificación
- Conditional Random Fields (CRF)
  - Modelo discriminativo
  - Considera características contextuales
- Redes Neuronales
  - BiLSTM-CRF
  - Transformer-based models

# ¿Qué es Named Entity Recognition?

- Definición: Identificar y clasificar entidades nombradas en texto
- Tipos de Entidades:
  - PERSON: Nombres de personas
  - ORG: Organizaciones, empresas
  - LOC/GPE: Lugares, países, ciudades
  - DATE: Fechas y expresiones temporales
  - MONEY: Valores monetarios
- Formato: Esquemas BIO, BILOU

# Algoritmos para NER

- Basados en Reglas:
  - Patrones predefinidos
  - Diccionarios y gazetteers
- Modelos Estadísticos:
  - CRF (Conditional Random Fields)
  - Support Vector Machines
- Deep Learning:
  - BiLSTM-CRF
  - Transformer-based (BERT, RoBERTa)
  - spaCy, Flair, Hugging Face

# Aplicaciones Prácticas

- Traducción Automática:
  - POS: Estructura gramatical
  - NER: Preservar nombres propios
- Búsqueda de Información:
  - POS: Comprensión de consultas
  - NER: Búsqueda por entidades
- Análisis de Sentimientos:
  - POS: Identificar adjetivos/adverbios
  - NER: Sentimiento específico de entidades
- Chatbots:
  - POS: Clasificación de intenciones
  - NER: Extracción de parámetros



# Herramientas y Librerías

- NLTK (Python):
  - `pos_tag()` para POS tagging
  - `ne_chunk()` para NER básico
- spaCy (Python):
  - Modelos pre-entrenados
  - Pipeline completo NLP
- Hugging Face Transformers:
  - Modelos BERT, RoBERTa
  - Fine-tuning personalizado
- Stanford CoreNLP (Java):
  - Suite completa NLP

# Tendencias Actuales

- Modelos Transformer:
  - BERT, RoBERTa, DistilBERT
  - Transfer learning
- Multilingual Models:
  - mBERT, XLM-R
  - Zero-shot cross-lingual
- Few-shot Learning:
  - GPT-3, ChatGPT
  - Prompt engineering
- giDomain Adaptation:
  - Especialización por dominio
  - Active learning