

Introducción al Procesado de Lenguaje Natural



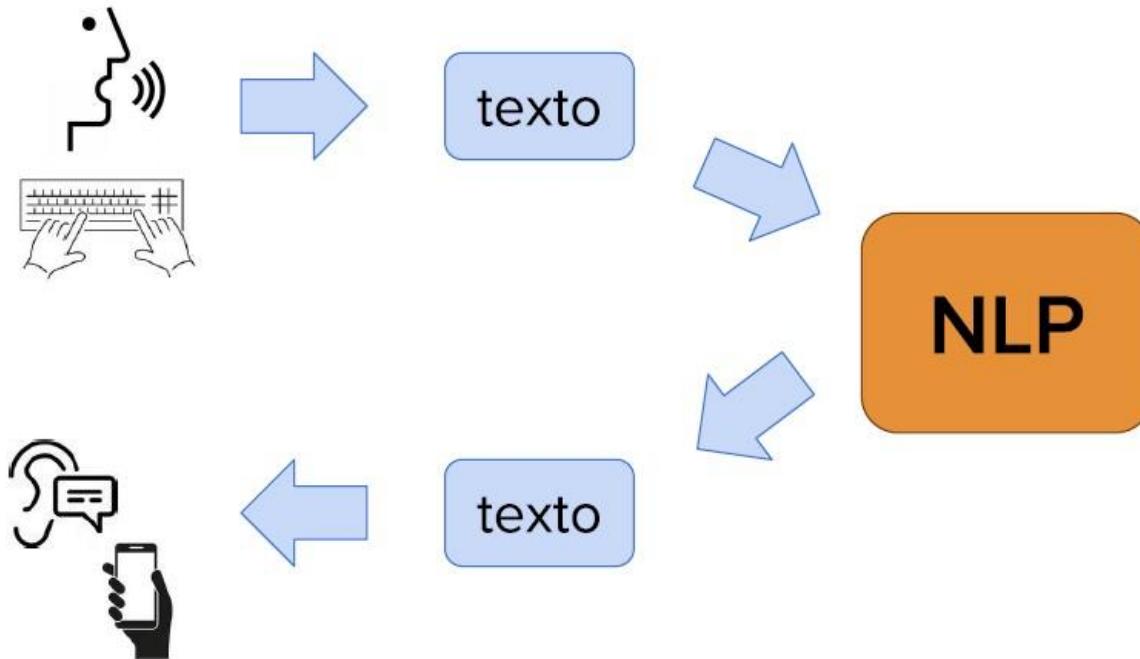
¿Qué es el NLP?

- Natural Language Processing (NLP)
- Intersección entre la IA, la lingüística y las ciencias de la computación
- Casi siempre se trabaja con texto, aunque también voz
- Es keyword matching, es normalización de textos, es ML, es clustering
- ...
- Entendimiento humano-máquina

Perfiles en NLP

- Lingüistas
- Matemáticos
- Ingenieros
- Estadísticos,.....

Cadena típica de NLP

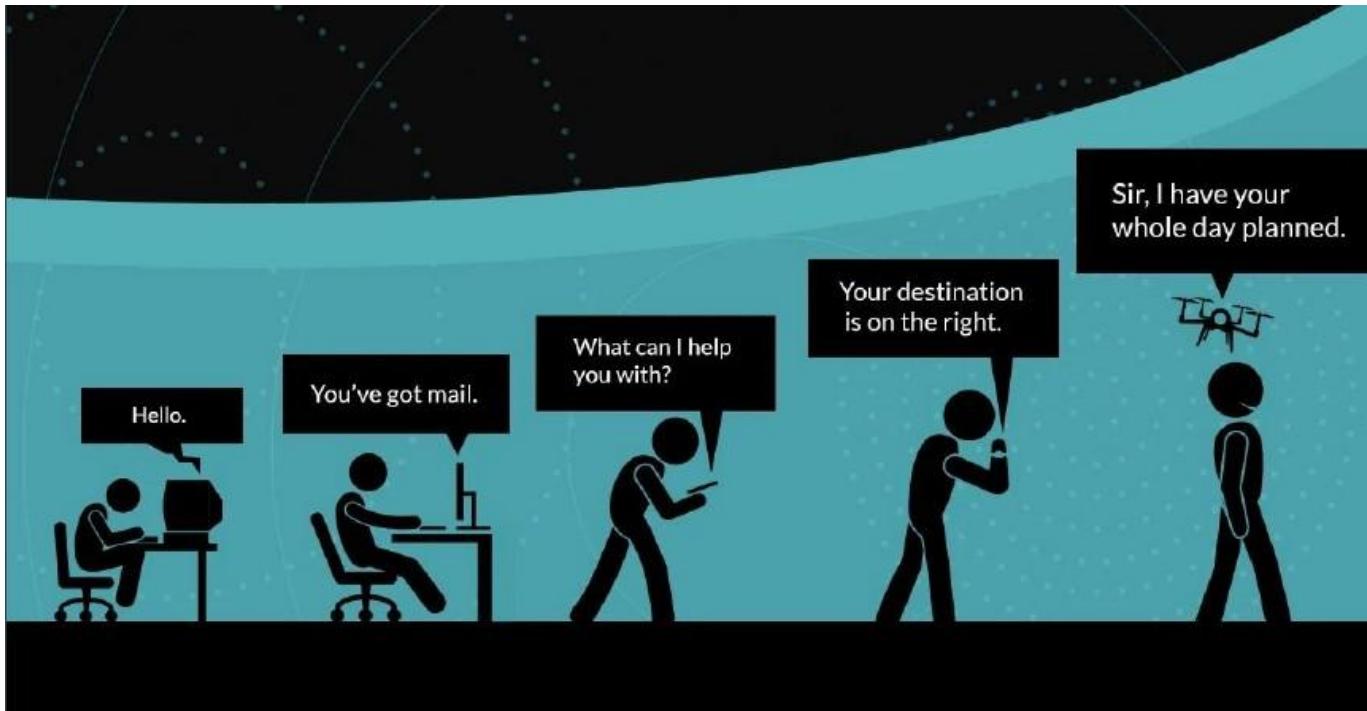


Tareas en NLP

- Dominio concreto
 - Clasificación de documentos
 - Análisis de sentimiento en comentarios sobre marcas, productos, mercados, políticos,....
 - Búsqueda de información en un buscador
 - Detección de los temas de moda en redes sociales
- End-to End
 - Agentes conversacionales
 - Traducción bidireccional en tiempo real
 - Módulo de entendimiento de lenguaje (NLU) completo



Estado del arte

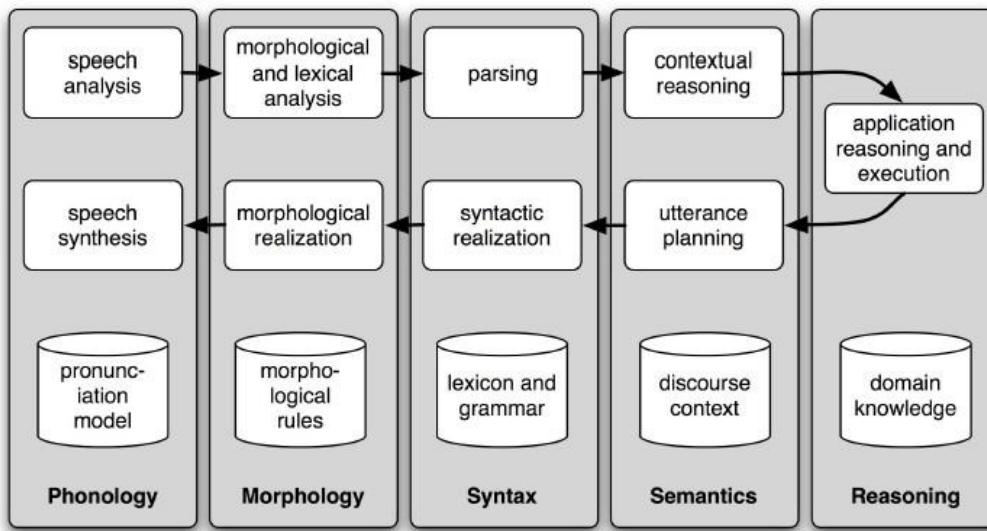


Estado del arte

Prácticamente Resueltas	<ul style="list-style-type: none">• Detección de spam• PoS Tagging• “Named Entity Recognition”
Progresando adecuadamente	<ul style="list-style-type: none">• Sentiment Analysis• Machine Translation• Desambiguación• Extraccion de Informacion
	<ul style="list-style-type: none">• Diálogo• Parafrasear• Preguntas y Respuestas• Resumir texto• Detección de sarcasmo

Tareas en NLP

- Algunas tareas de NLP en un asistente virtual

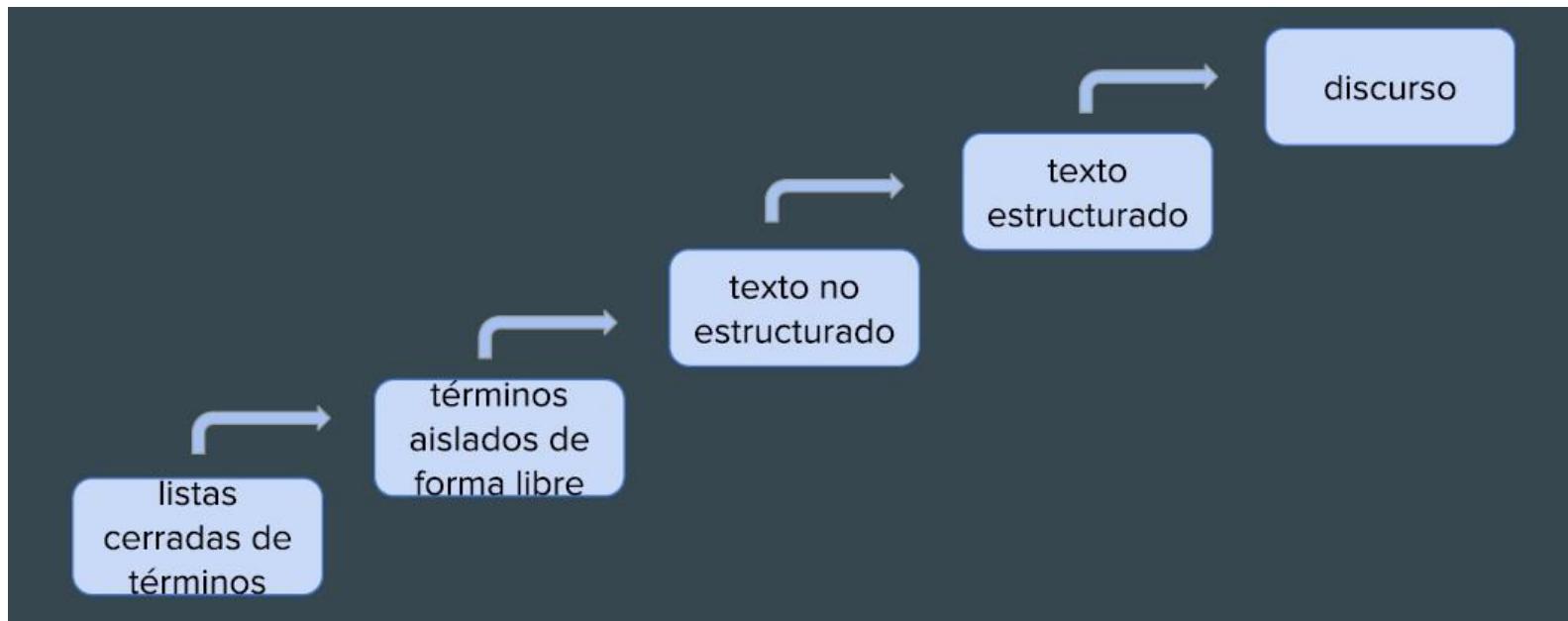


Estado del arte

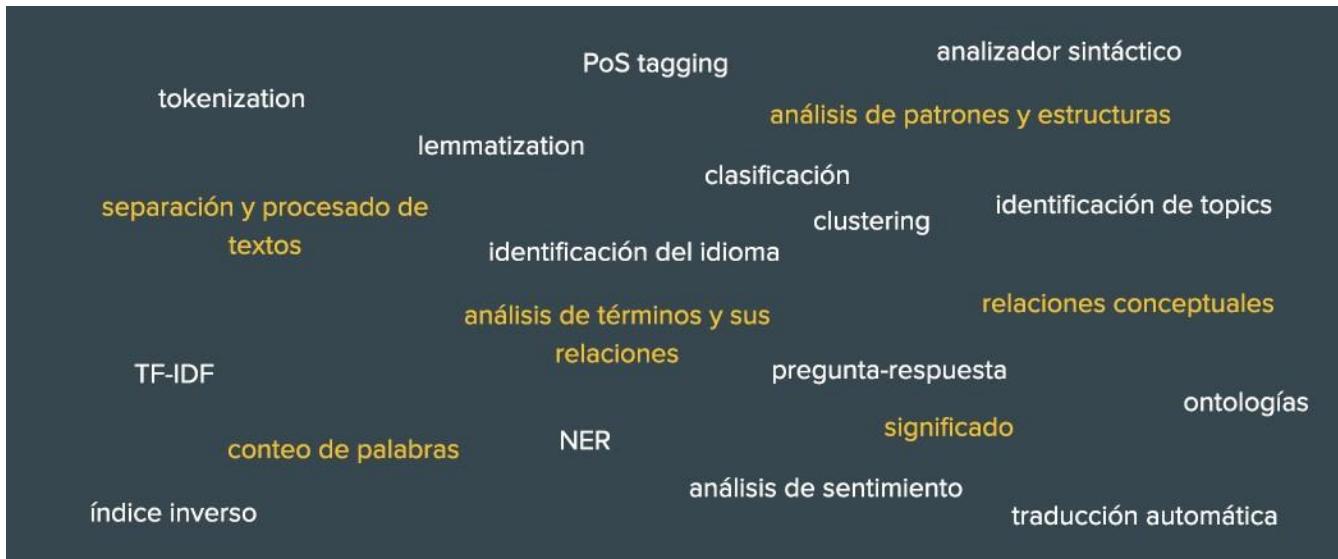


Procesado de texto

¿Qué es un texto?



Procesado de texto



Algunos retos

Entre lo pienso,
lo que quiero decir,
lo que creo decir,
lo que digo,
lo que quieres oír,
lo que oyes,
lo que crees entender,
lo que quieres entender,
lo que entiendes

Existen 9 posibilidades de no entenderse



Algunos retos

Regla de la comunicación del 7-38-55 % de Mehrabian



Algunos retos

Mapear palabras y conceptos

- **Homonimia: misma forma, distinto significado, distinta etimología**
 - Homografía: misma escritura
 - Nada (nadar) y nada, vela (velar) y vela, bota (calzado) y bota (vino)
 - Homofonía: misma pronunciación
 - Basta y vasta, sabia y savia, botar y votar
- **Polisemia: misma palabra (misma etimología), distintos significados**
 - Cabo (escalón militar) y cabo (accidente geográfico), gato (animal) y gato (herramienta)
- **Sinonimia: relación de igualdad entre el significado de +2 palabras/enunciados**
 - Barato y económico, demente y loco, idioma y lengua,.....



Algunos retos

- **Typos**
 - Ejemplos de respuestas en un formulario (¿dónde reside usted?): Illescas, illescas, ILLESCAS, ILL
- **Ground truth**
 - ¿Con qué entrenamos? ¿Con qué validamos?
- **Tropos**
 - Ironía, Sarcasmo, Metáfora, Alegoría, Oxímoron,...
- **Contexto**
 - P: ¿Qué temperatura hace en Roma?
 - R: 27 grados centígrados
 - P: ¿Y en Lima?
 - R:
- **Idioma**
- Puede llegar cualquier texto (sobre todo en asistentes virtuales)

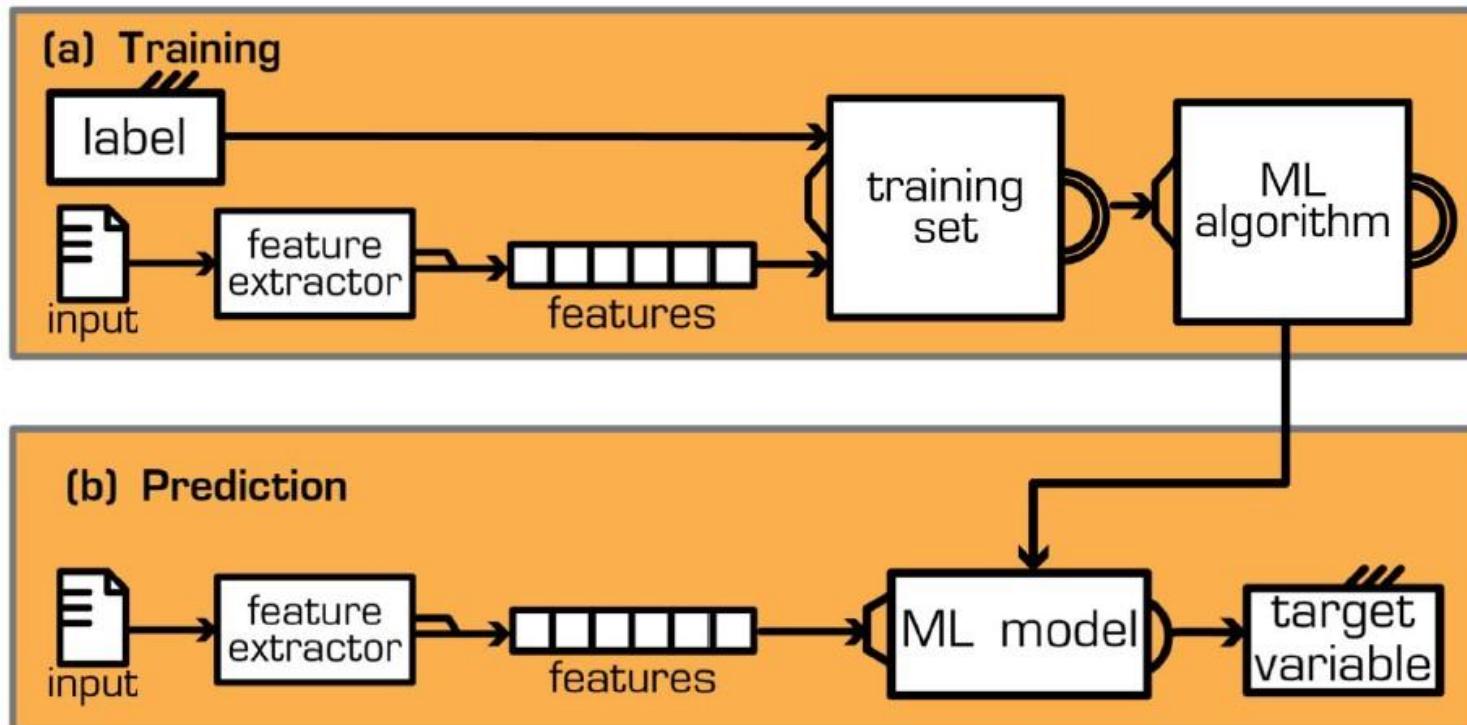


NLP vs Computer Vision

- En general, los modelo de CV performance mejor
- Modelos pre-entrenados
- Debate
 - Retos en modelos de Visión Artificial
 - Retos en modelo de NLP



Pipeline de NLP





KEEPCODING

Tech School

Madrid | Barcelona | Bogotá

Datos de contacto