

# OpenCourseWare

## Procesamiento de Lenguaje Natural con Aprendizaje Profundo, Máster en Ciencia y Tecnología Informática

### Tema 1. Introducción

# Agenda

- Qué es Procesamiento de Lenguaje Natural (PLN)?.
- Por qué PLN es importante?
- Principales aplicaciones de PLN.
- Qué es aprendizaje profundo?
- Aprendizaje automático vs Aprendizaje Profundo.
- Aprendizaje profundo aplicado a PLN.
- Ejemplos de arquitecturas profundas para PLN.

# Procesamiento de Lenguaje Natural

- Técnicas computacionales para analizar y representar el lenguaje humano de forma automática
- Campo multidisciplinar: informática, lingüística, matemáticas, psicología, etc.

# Por qué PLN?



- Crecimiento exponencial de los datos
  - 2013, 3.5 ZB
  - 2022, 40 ZB
  - 2025, 180 ZB
- > 80% en formato no estructurado, principalmente texto.

\*1 ZB = 1 trillion GB

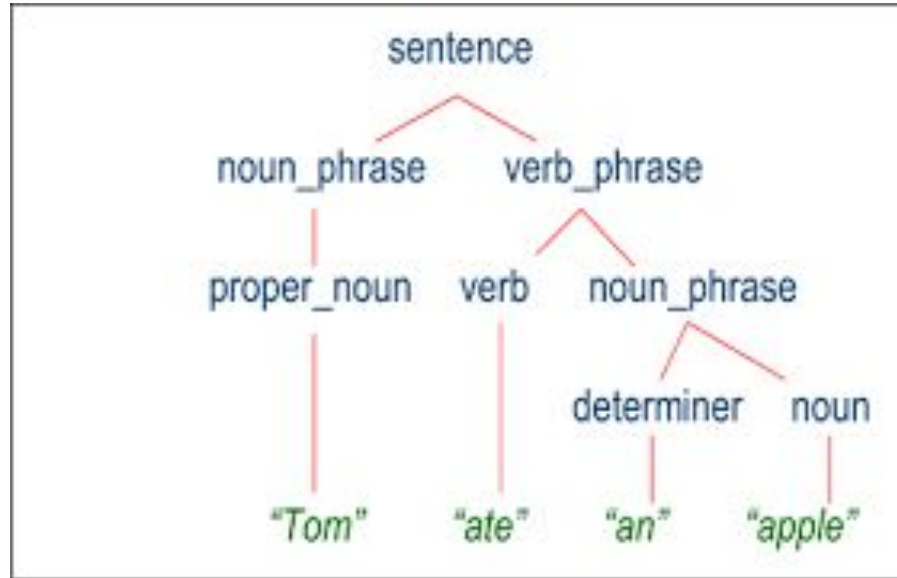
# Aplicaciones PLN

# Análisis Morfosintáctico (Part-of-speech (PoS) tagging)



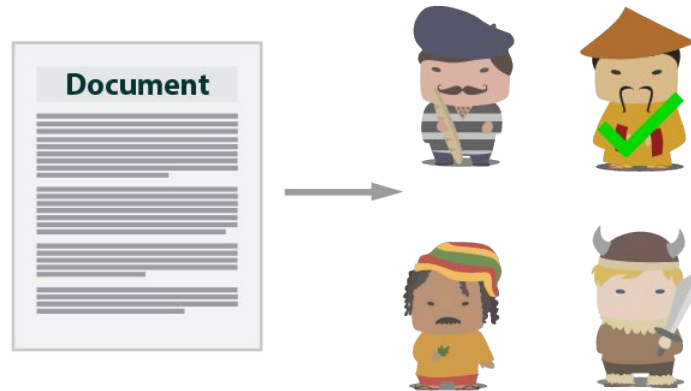
<https://github.com/dhirajhr/POS-Tagging>

# Análisis Sintáctico (Parsing)



<https://forum.huawei.com/>

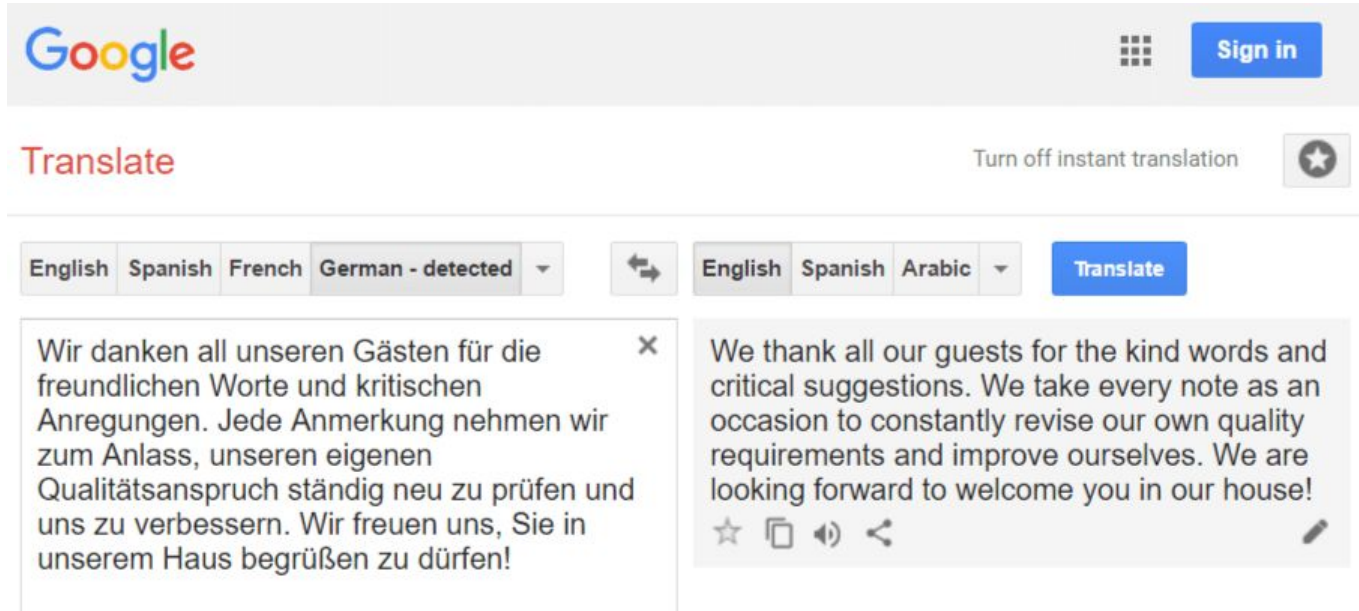
# Detección del idioma



<https://aylien.com/text-api/language-detection/>



# Traducción automática



The screenshot displays the Google Translate web interface. At the top, the Google logo is on the left, and a 'Sign in' button is on the right. Below the logo, the word 'Translate' is written in red. To the right of 'Translate' is a link that says 'Turn off instant translation' and a star icon in a circle. The main translation area has two language selection bars. The left bar shows 'English', 'Spanish', 'French', and 'German - detected' (selected). The right bar shows 'English', 'Spanish', and 'Arabic' (selected). A blue 'Translate' button is positioned between the two bars. Below the language bars, the German text 'Wir danken all unseren Gästen für die freundlichen Worte und kritischen Anregungen. Jede Anmerkung nehmen wir zum Anlass, unseren eigenen Qualitätsanspruch ständig neu zu prüfen und uns zu verbessern. Wir freuen uns, Sie in unserem Haus begrüßen zu dürfen!' is shown in a box with a close button (X). To the right, the English translation 'We thank all our guests for the kind words and critical suggestions. We take every note as an occasion to constantly revise our own quality requirements and improve ourselves. We are looking forward to welcome you in our house!' is shown in a box with a star, copy, and share icons, and a pencil icon for editing.

Google

Sign in

Translate

Turn off instant translation

English Spanish French German - detected

English Spanish Arabic

Translate

Wir danken all unseren Gästen für die freundlichen Worte und kritischen Anregungen. Jede Anmerkung nehmen wir zum Anlass, unseren eigenen Qualitätsanspruch ständig neu zu prüfen und uns zu verbessern. Wir freuen uns, Sie in unserem Haus begrüßen zu dürfen!

We thank all our guests for the kind words and critical suggestions. We take every note as an occasion to constantly revise our own quality requirements and improve ourselves. We are looking forward to welcome you in our house!

# Recuperación de Información

The screenshot shows a Google search for 'BERT'. The search bar at the top contains the word 'BERT'. Below the search bar, the results are displayed. The first result is from 'inboundcycle.com' with the title 'Guía avanzada de Google BERT: qué es, cómo funciona y en ...'. Below this, there is a section titled 'Otras preguntas de los usuarios' with four questions: '¿Qué es Bert algoritmo?', '¿Que se significa Bert?', '¿Cómo funciona la IA de Google?', and '¿Qué inteligencia artificial usa Google?'. To the right of the search results is a knowledge panel for 'BERT', which is described as a 'Modelo de lenguaje'. The panel includes a brief explanation of BERT as a Bidirectional Encoder Representations from Transformers model and a list of related search terms: GPT-3, Word2vec, PyTorch, and fastText.

Recibidos (1) - isegura@inf.uci... X CursoNLP - Google Drive X 1 - PLN - Intro - Presentacione... X 0-Intro - Presentaciones de G... X BERT - Buscar con Google X +

google.com/search?q=BERT&rlz=1C5CHFA\_enES833ES833&oq=BERT&aqs=chrome..69i57j46i433i2j0i433j0j46i433.697j0j15&sourceid=chrome&ie=UTF-8

Aplicaciones Curso: MAG. Estru... Mi unidad - Googl... Firmar Clases imp... GITHUB EDA Cambridge Evernote Web Overleaf Portafirmas Traducir Cookidoo\* - la pla...

Google BERT

Todo Imágenes Vídeos Noticias Maps Más Configuración Herramientas

Aproximadamente 424.000.000 resultados (0,32 segundos)

<https://www.inboundcycle.com> > google-bert-que-es-co... >

**Guía avanzada de Google BERT: qué es, cómo funciona y en ...**

21 nov 2019 — BERT es el acrónimo para Bidirectional Encoder Representations from Transformers (Representaciones de Codificador Bidireccional de ...

Otras preguntas de los usuarios

¿Qué es Bert algoritmo?

¿Que se significa Bert?

¿Cómo funciona la IA de Google?

¿Qué inteligencia artificial usa Google?

Enviar comentarios

<https://www.bbc.com> > mundo > noticias-50223408 >

**Google: cómo funciona BERT, la mayor actualización del ...**

29 oct 2019 — Sistema neuronal. De acuerdo a Google, BERT es una red neuronal de código abierto que ha sido entrenada para procesar el lenguaje natural.

<https://www.arsys.es> > blog > algoritmo-bert >

**Así funciona Bert, el nuevo algoritmo de Google - Blog de ...**

16 dic 2019 — BERT significa Bidirectional Encoder Representations from Transformers y lo que hace es analizar el contexto de las palabras de la búsqueda ...

<https://conectasoftware.com> > marketing-online > algoritm...

**Nuevo algoritmo Google BERT: ¿Qué es? ¿Cómo ayuda a ...**

8 may 2020 — Bidireccional: BER es un algoritmos que funciona de manera bidireccional, es decir, analiza los textos ya sea de derecha a izquierda como de ...

Google BERT

Más imágenes

BERT

Modelo de lenguaje

BERT o Representación de Codificador Bidireccional de Transformadores es una técnica basada en redes neuronales para el pre-entrenamiento del procesamiento del lenguaje natural desarrollada por Google. BERT fue creado y publicado en 2018 por Jacob Devlin y sus colegas de Google. [Wikipedia](#)

También se buscó Ver 10 más

GPT-3 Word2vec PyTorch fastText

Sugerencias

# Extracción de Información

## Text in

**Brazil** ranks number 5 in the list of countries by population.

The term "**Ibu Negara**" (Lady/Mother of the State) is used for **wife** of the President of Indonesia.

**Game of Thrones** is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels. It ranks **fourth** among the **IMDB Top Rated TV Shows**

## Data out

### THE COUNTRIES WITH THE LARGEST POPULATION

China	1	1,388,232,693
India	2	1,342,512,706
Unites States	3	326,474,013
Indonesia	4	263,510,146
<b>Brasil</b>	5	174,315,386

### THE COUNTRY'S' FIRST LADIES

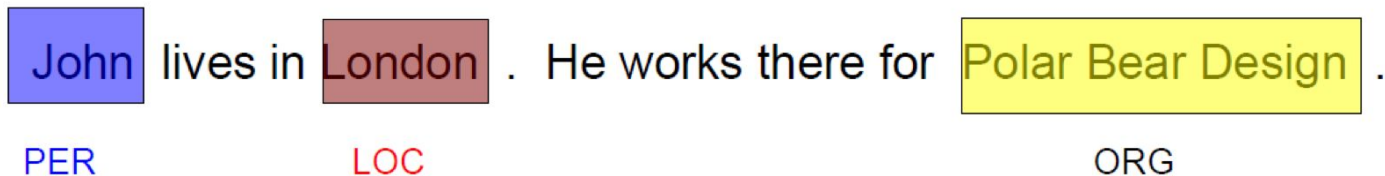
Brigitte Macron  
- Spouse: Emmanuel Macron, President of France (2017 - )  
Melania Trump  
- Spouse: Donald J. Trump, U.S. President (2017 - )  
**Iriana Widodo**  
- Spouse: Joko Widodo, **President of Indonesia** (2014 - )  
- Also known as: "**Ibu Negara**" (Lady/Mother of the State)

### IMDB TOP RATED TV SHOWS

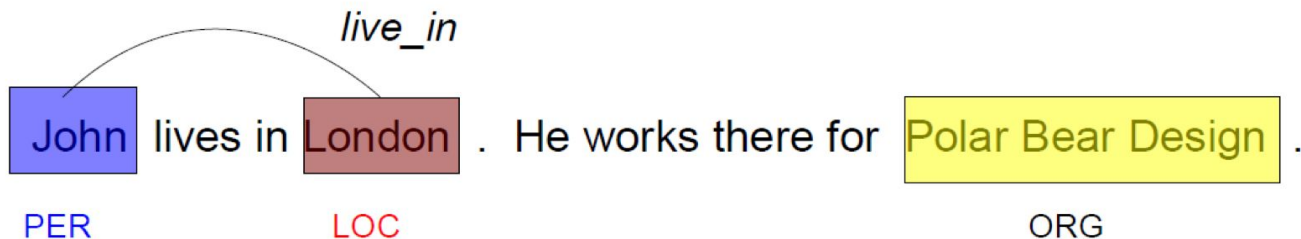
- 1 Planet Earth II (2016) 9.6.
- 2 Band of Brothers (2001) 9.5.
- 3 Planet Earth (2006) 9.5.
- 4 **Game of Thrones** (2011) 9.4.
- 5 Breaking Bad (2008) 9.4.

# Extracción de Información. Tareas:

## 1) Reconocimiento de Entidades Nombradas. (Named Entity Recognition (NER))

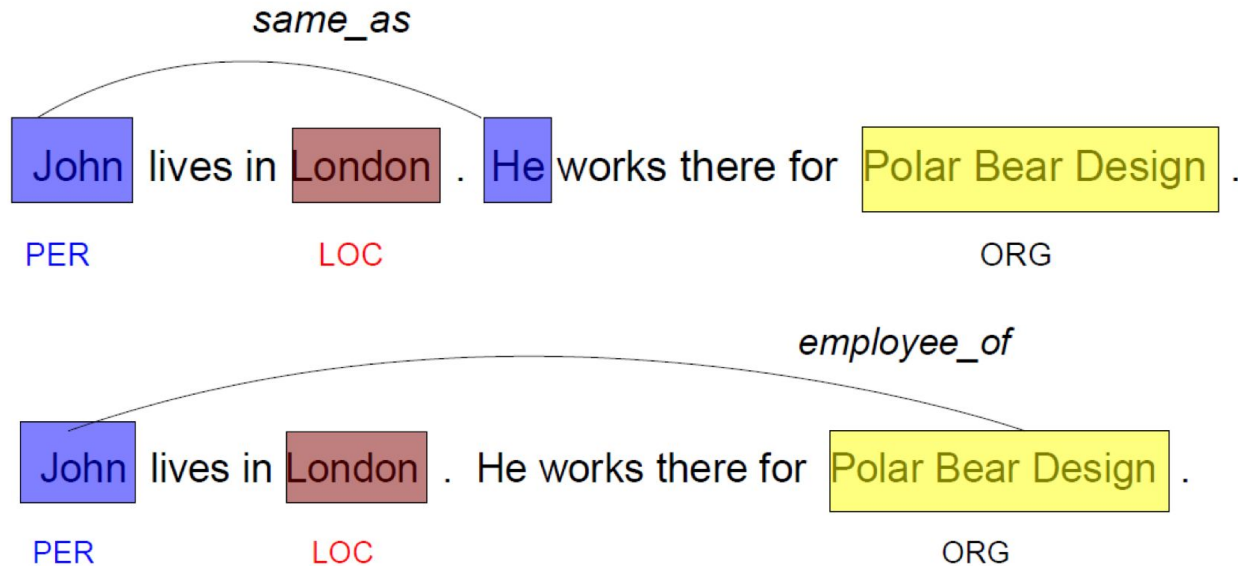


## 2) Extracción de relaciones



# Extracción de Información. Tareas:

## 3) Resolución de correferencias.





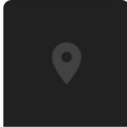



# Sistemas de Búsqueda de Respuestas (Question Answering)

Google

cuál es el edificio más alto del mundo

Todo Imágenes Noticias Vídeos Maps Más Configuración Herramientas

Mundo / Edificios (más altos)

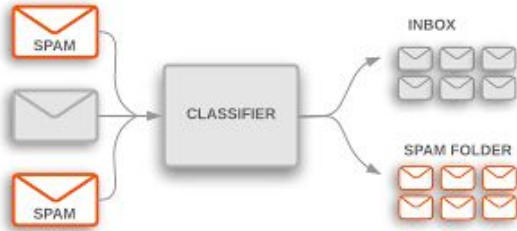
					
Burj Khalifa 828 m	Torre de Shanghai 632 m	Makkah Clock Tower 601 m	Ping An Finance Cen... 555 m	Lotte World Tower 555 m	One World Trade Center 541,3 m

¿Cuál es el edificio más alto del mundo? | Plataforma ...

<https://www.plataformaarquitectura.cl> › ArchDaily › Artículos ▼

23 ene. 2019 - En la actualidad existen instituciones especializadas que establecen los parámetros para definir objetivamente cuánto mide un edificio.

# Clasificación de Textos



# Generación de Resúmenes

**Source Text:** Peter and Elizabeth took a taxi to attend the night party in the city.

While in the party, Elizabeth collapsed and was rushed to the hospital.

**Summary:** Elizabeth was hospitalized after attending a party with Peter.





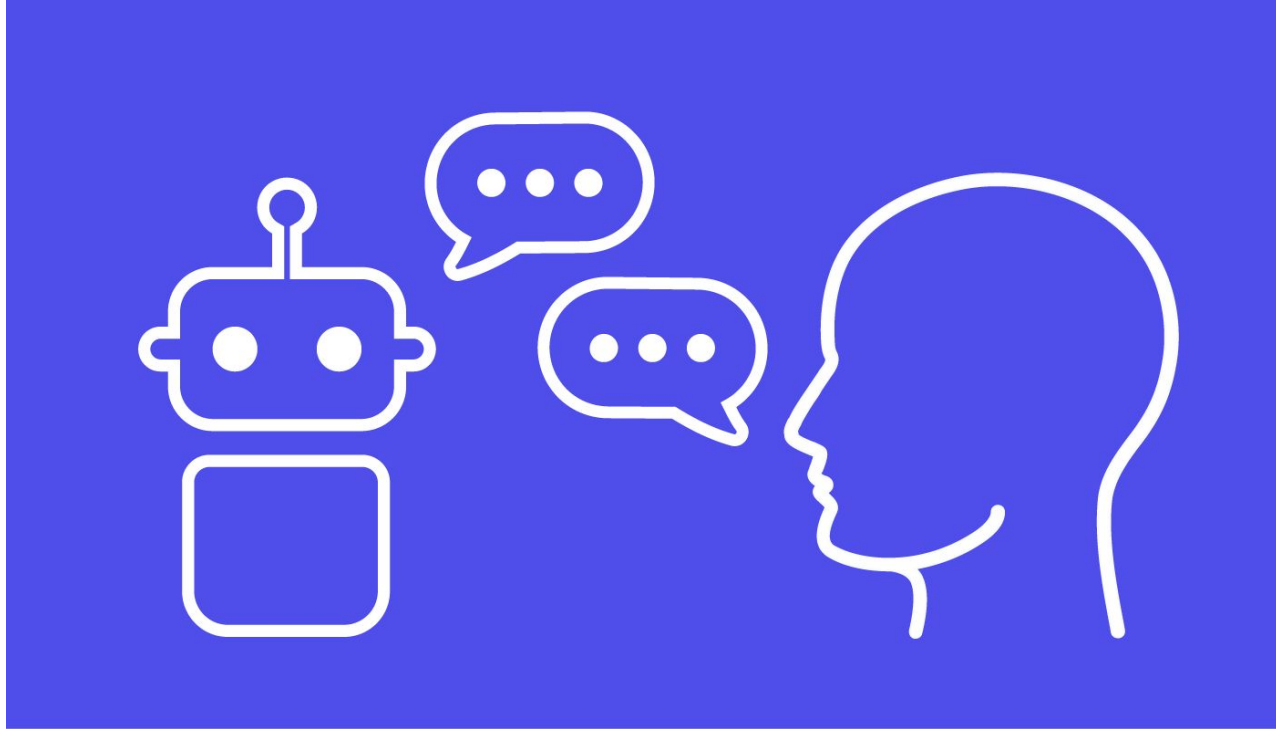
# Simplificación de Textos

---

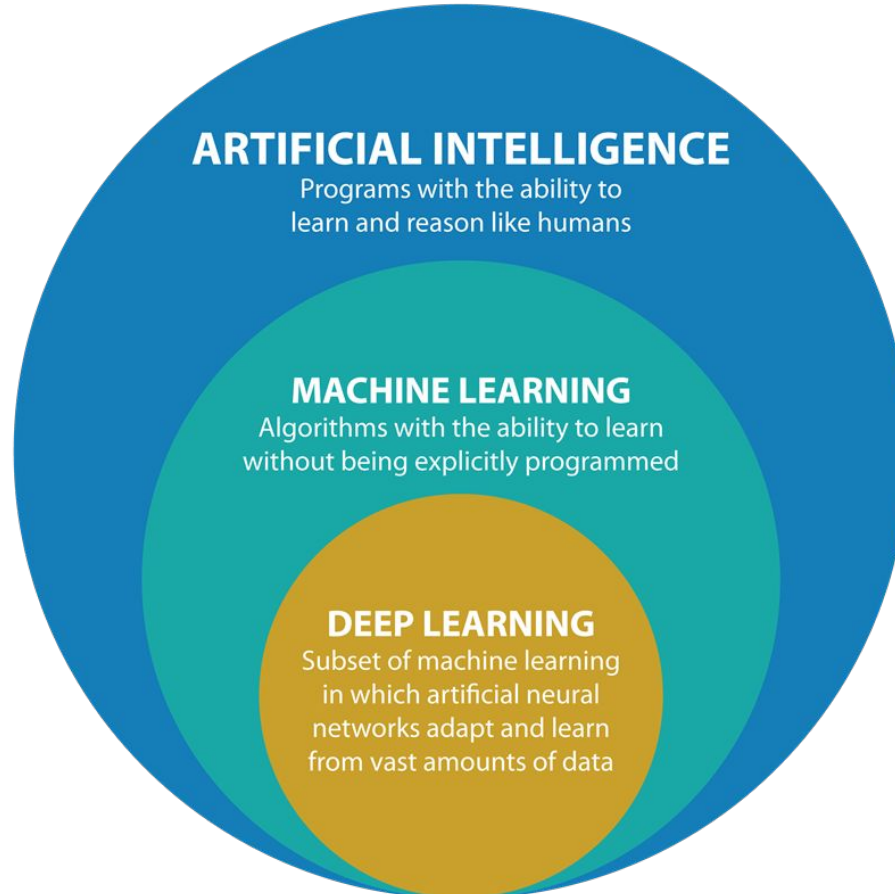
Normal: Alfonso Perez *Munoz, usually referred to as Alfonso*, is a former Spanish *footballer, in the striker position*.

Simple: Alfonso Perez is a former Spanish *football player*.

# Asistentes Conversacionales (Chatbots)

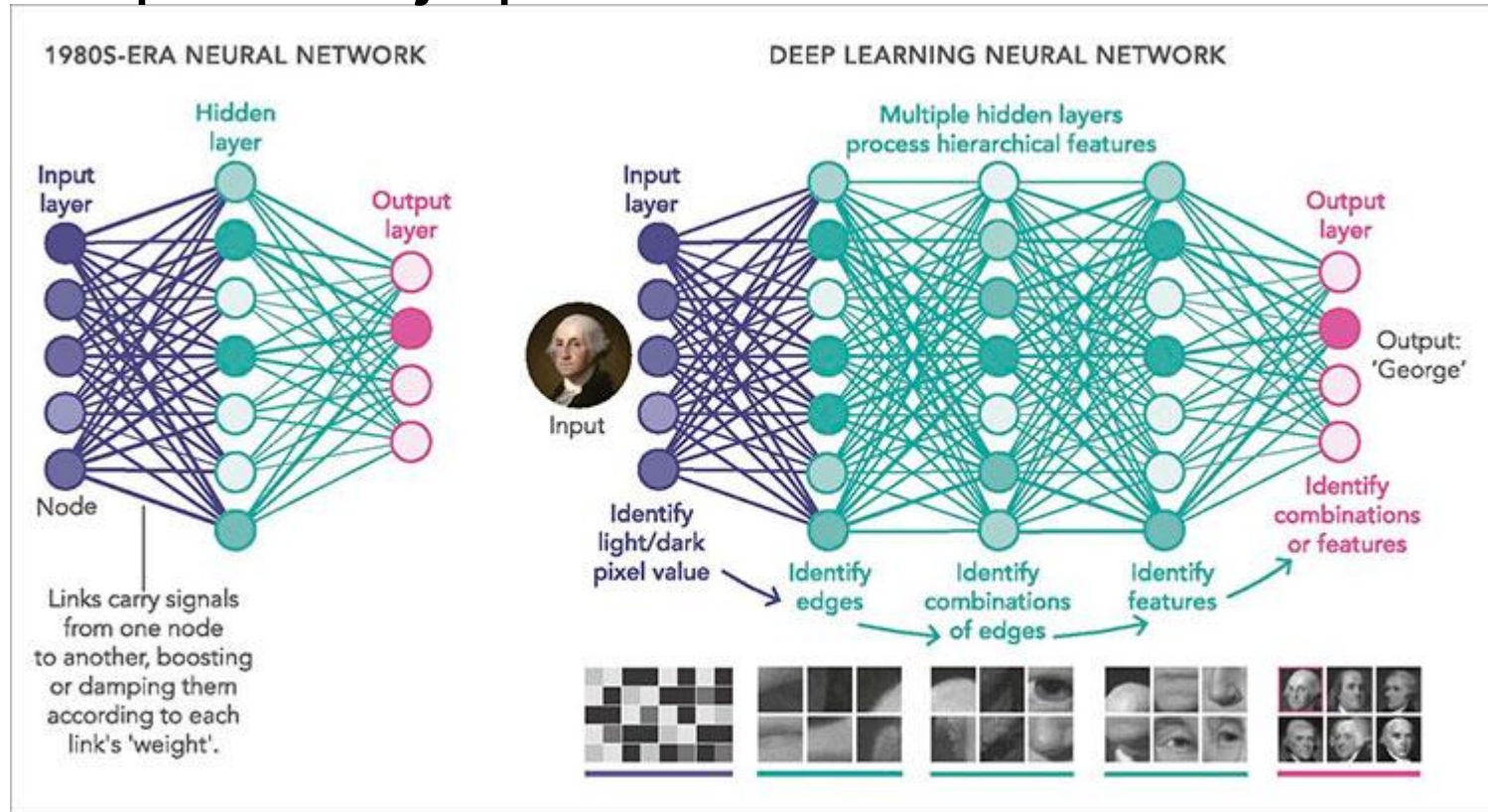


# Qué es aprendizaje profundo?



# Qué es aprendizaje profundo?

Employ multiple processing layers to learn hierarchical representations of data



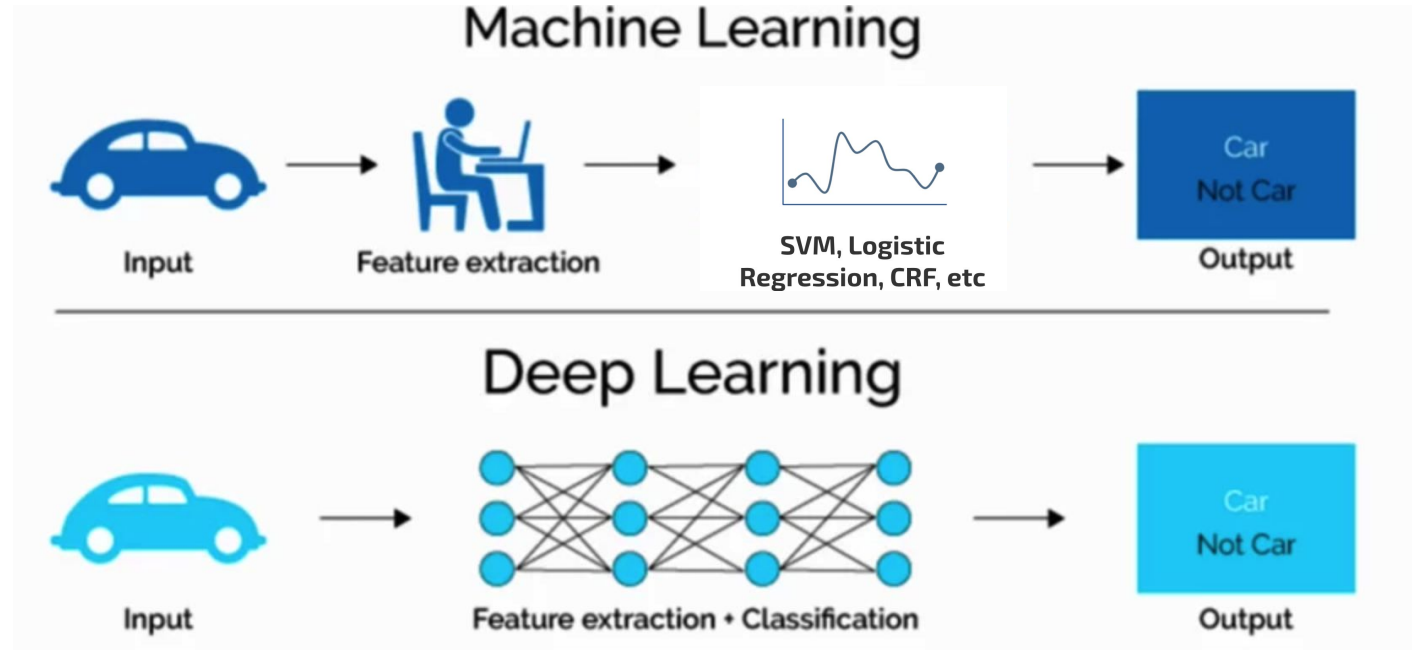
<https://www.futurespace.es/redes-neuronales-y-deep-learning-capitulo-1-preludio/>

Produce state-of-the-art results in many domains  
(such as computer vision, pattern recognition)

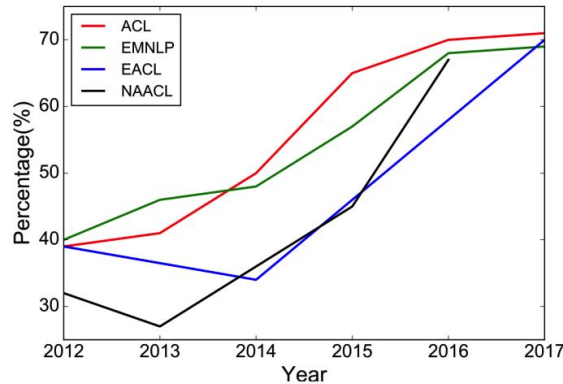
# Aprendizaje automático vs Aprendizaje Profundo

Traditional ML algorithms for NLP tasks heavily require on hand-crafted features. Feature engineering is a time-consuming process.

Moreover, it is not usually robust enough (low recall)



# Aprendizaje profundo para PLN



## Deep Learning in NLP

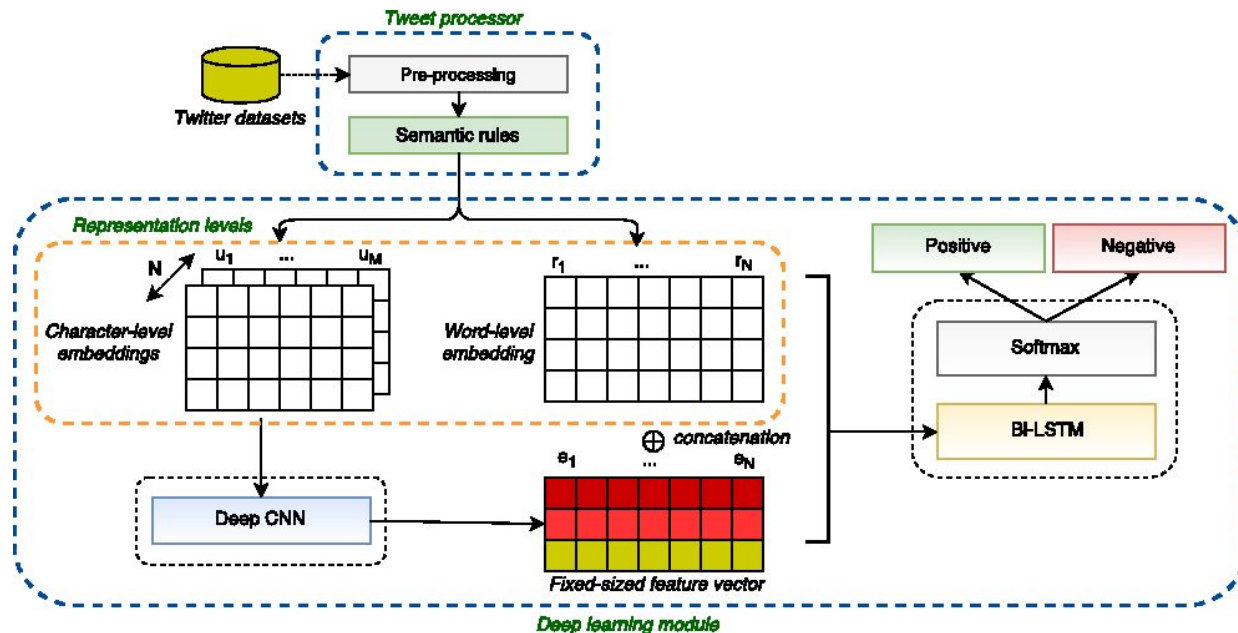
In the last years, NLP research has also exploited deep learning techniques

Produce state-of-the-art results in many NLP applications (machine translation, IE, text summarization, etc)

## Results



# Ejemplo: BiLSTM for Text Classification (Sentiment Analysis of Tweets)



# Resumen

- **PLN:** entender y generar el lenguaje humano de forma automática.
- **Aprendizaje profundo** son redes neuronales con múltiples capas.
- Selección de características es una tarea crucial en los enfoques tradicionales de aprendizaje automático.
- **El aprendizaje profundo evita la tarea de selección de características** porque la red aprende la mejor representación de los datos.



# Resumen

- El **aprendizaje profundo** ha revolucionado el campo del NLP, obteniendo los mejores resultados en muchas de las aplicaciones de PLN (clasificación de textos, traducción automática, etc).
- Las técnicas más avanzadas son:
  - word embeddings -> CNN -> BiLSTM -> Transformers -> BERT

Thank you  
Question time!!!

[isegura@inf.uc3m.es](mailto:isegura@inf.uc3m.es)

<https://hulat.inf.uc3m.es/nosotros/miembros/isegura>

<https://github.com/isegura>