

# Web 3.0, Dataweb e Inteligencia Artificial

**Complemento de Base de Datos**

Jesús Ortiz Calleja  
Juan Carlos Utrilla Martín

# Índice

---

## **Introducción**

Conceptos básicos

Stack Tecnológico

Trabajo

Resultados

Conclusiones

# Introducción

---

En los comienzos de la web, los buscadores simplemente se limitaban a buscar documentos cuyos textos guardaban similitud con nuestras búsquedas.

**Web 1.0**

# Introducción

Pero actualmente....



Google

cuando nacio rosalia

Q Todo Noticias Imágenes Vídeos Maps Más Configuración Herramientas

Aproximadamente 1.640.000 resultados (0,73 segundos)

Rosalía Vila / Fecha de nacimiento

25 de septiembre de 1993

edad 25 años



# Introducción

La **web 3.0** nos permite obtener información más precisa y de forma más fácil. ¿Y si hacemos una petición?



# Introducción

---

- Los buscadores **entienden lo que preguntamos.**
- **Entienden** la información asociada a los **recursos web.**

**Web 3.0**

# Introducción

---

La **web 3.0** se limita a darnos información, no aporta soluciones ni realiza acciones.

Sin embargo gracias a la aplicación de **IA** surge un nuevo concepto que cumple estas expectativas.

**Web 4.0 = Web 3.0 +**



Artificial  
intelligence

# Índice

---

Introducción

**Conceptos básicos**

Stack Tecnológico

Trabajo

Resultados

Conclusiones



# Conceptos básicos - Dataweb

---

Son lo datos asociados a los **recursos web**.

Depósitos de petróleo

**A**

Predicciones de consumo petróleo

**B**

Extracción en tiempo real

**C**

**Nueva  
información**

```
graph LR; A[A] --> NI[Nueva información]; B[B] --> NI; C[C] --> NI;
```

The diagram illustrates the flow of data from three sources to a central information node. Source A (Depósitos de petróleo) is represented by a red letter 'A' in a rounded square. Source B (Predicciones de consumo petróleo) is represented by a red letter 'B' in a rounded square. Source C (Extracción en tiempo real) is represented by a red letter 'C' in a rounded square. Arrows from each of these three sources point towards a central rounded rectangle labeled 'Nueva información' in red text.

# Conceptos básicos - Dataweb

---

Están pensados para  
ser leídos **por humanos,**  
**no por máquinas.**

# Conceptos básicos - Semántica y ontología

---

**Solución 1 al dataweb: añadir metadatos.**



**R D F S**



# Conceptos básicos - Semántica y ontología

---

**Solución 2 al dataweb:** IA que comprenda los datos como los humanos.



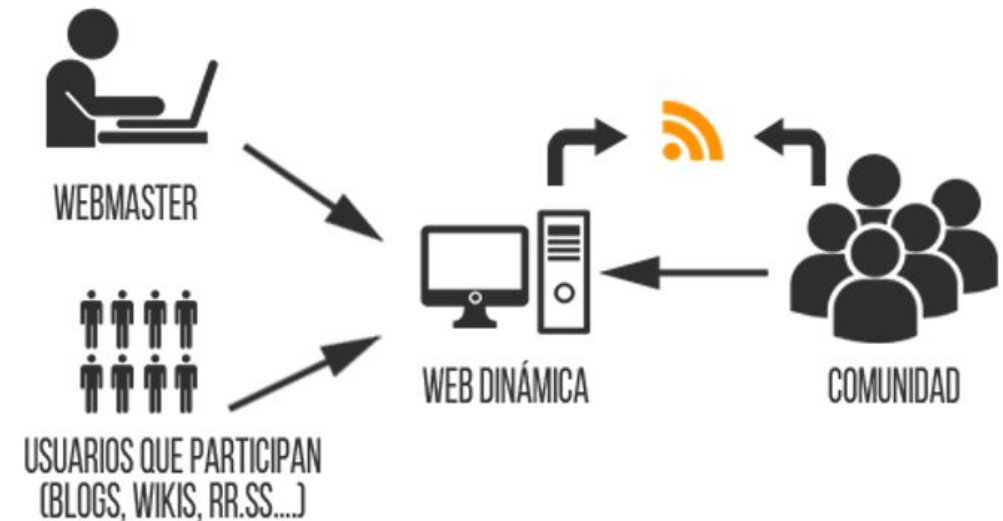
# Conceptos básicos - Evolución de la web

---

## Web 1.0



## Web 2.0



# Conceptos básicos - Evolución de la web

---

## Web 3.0



# Conceptos básicos - Evolución de la web

---

## Web 4.0



# Conceptos básicos - IA

---

**Inteligencia artificial (IA):** son el conjunto de técnicas que permiten a las máquinas desarrollar actividades complejas normalmente asociadas a los seres humanos.

- **IA computacional:** mecanismos adaptativos para sistemas complejos y cambiantes. Ej: **Machine Learning**.



# Conceptos básicos - Machine Learning

---

Su objetivo es que las máquinas puedan aprender, generando ellas mismas el conocimiento.

# Conceptos básicos - Machine Learning

---

**Machine learning** - Aprendizaje supervisado

**Conjunto de entrenamiento:**



# Conceptos básicos - Machine Learning

---

**Machine learning** - Aprendizaje no supervisado

**Conjunto de entrenamiento:**



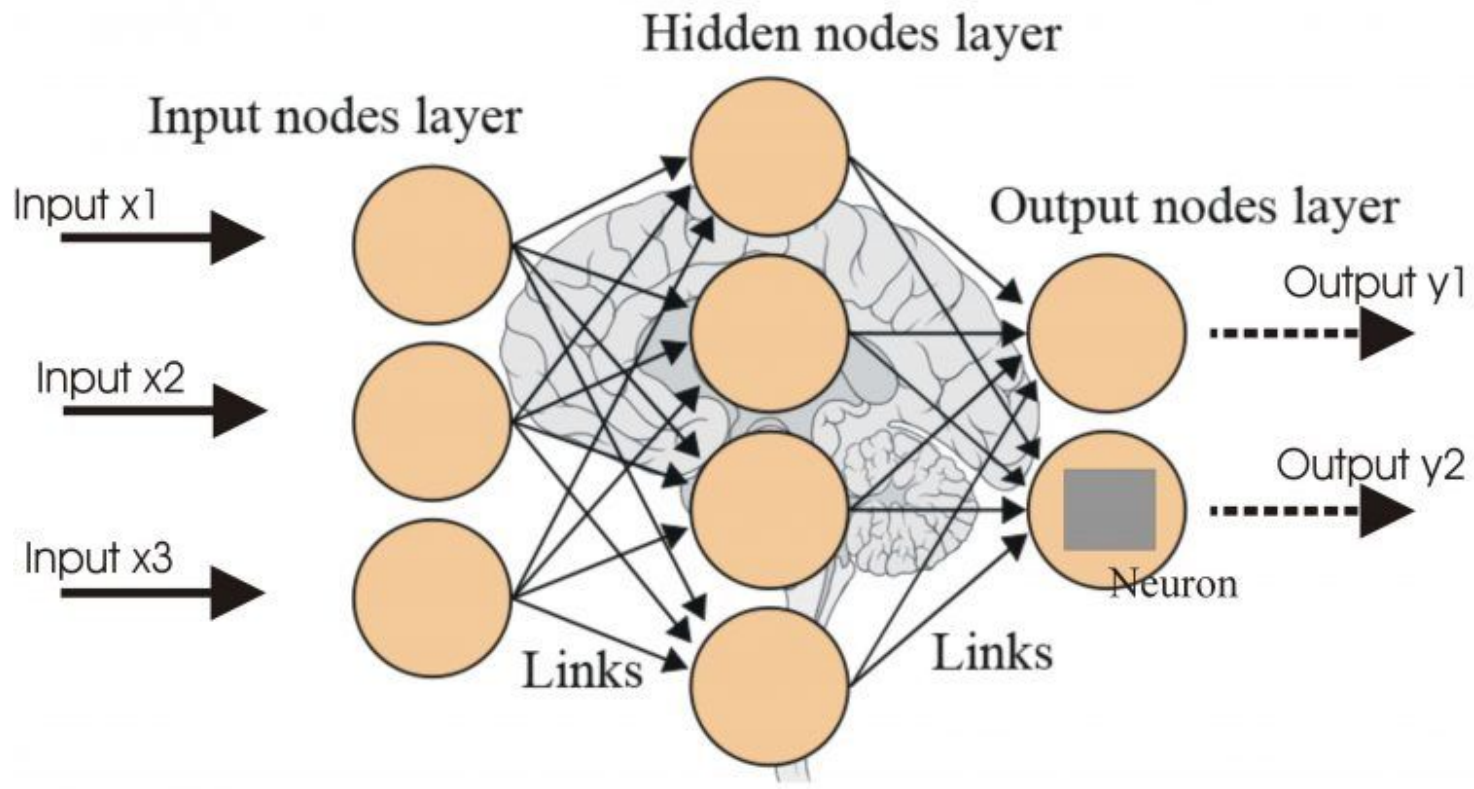
# Conceptos básicos - Deep Learning

---

Es una subrama del machine learning, principalmente se diferencia en que se dan muchos pasos, de ahí “deep”.

# Conceptos básicos - Red neuronal

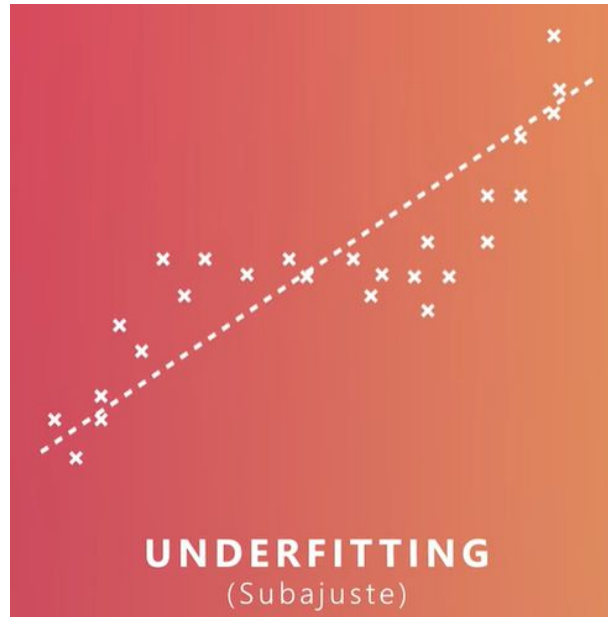
---



# Conceptos básicos - Underfitting

---

El **underfitting** se da cuando el modelo no es suficientemente flexible para el problema que se está tratando



# Conceptos básicos - Overfitting

---

El **overfitting** se da cuando el modelo está sobreajustado a los datos de entrenamiento, pero no es capaz de generalizar lo aprendido a datos nuevos.



# Conceptos básicos - Underfitting y Overfitting

---





# Índice

---

Introducción

Conceptos básicos

**Stack Tecnológico**

Trabajo

Resultados

Conclusiones

# Stack Tecnológico - Web semántica

---



# Stack Tecnológico - Inteligencia artificial

---



# Índice

---

Introducción

Conceptos básicos

Stack Tecnológico

**Trabajo**

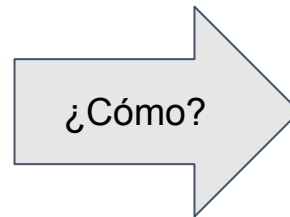
Resultados

Conclusiones

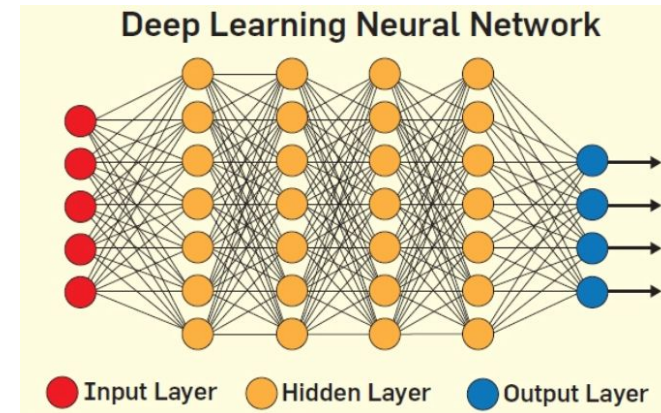
# Trabajo - Objetivo

---

Inferir la semántica de  
un texto de DBpedia



Deep learning con  
redes neuronales



# Trabajo - Obtención de los datos

---

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbc: <http://dbpedia.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umb: <http://umbel.org/umbel/rc/>

SELECT DISTINCT ?subject ?object WHERE {
    ?subject rdf:type umb:Work.
    ?subject dbo:abstract ?object.FILTER(lang(?object) = 'en')
}
LIMIT 15000
```

# Trabajo - Obtención de los datos

---

```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbc: <http://dbpedia.org/property/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX umb: <http://umbel.org/umbel/rc/>

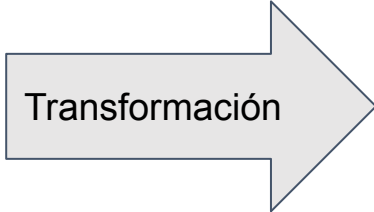
SELECT DISTINCT ?subject ?object WHERE {
    ?subject rdf:type umb:Person.
    ?subject dbo:abstract ?object.FILTER(lang(?object) = 'en')
}
LIMIT 15000
```

# Trabajo - Preprocesamiento de datos

---

## Lista de textos

[..., "Esto es  
un texto sobre  
una  
persona", ...]



Transformación

## Lista de vectores

[..., [ 0 1 2 3 4 5 6 ], ...]

+

## Lista de labels

[..., Person, ...]



# Trabajo - Particionado de los datos

---

Datos para la  
red neuronal

- Datos de entrenamiento:  
Entrena a la red neuronal
- Datos de validación:  
Mejora el ajuste de la red neuronal
- Datos de testing  
Prueba el modelo final

# Trabajo - Ver demo

---

Ejecución del código

# Índice

---

Introducción

Conceptos básicos

Stack Tecnológico

Trabajo

**Resultados**

Conclusiones

# Resultados obtenidos

---

Para las pruebas se han modificado los siguientes parámetros:

- Longitud del texto
- Épocas
- Tamaño del lote de ejemplos

# Resultados

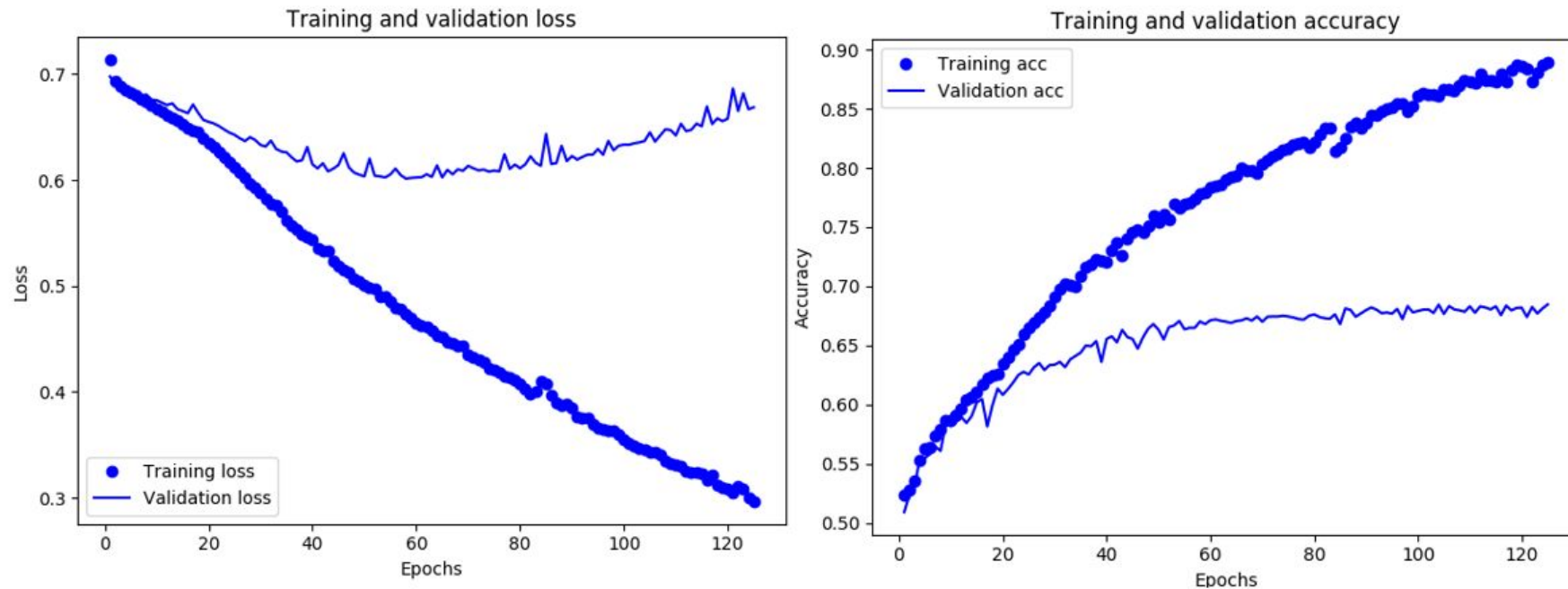
---

Para las pruebas se han modificado los siguientes parámetros:

- **Longitud del texto:** cantidad de palabras que la red neuronal va a aprender de cada texto.
- **Épocas:** n° veces que se itera sobre el conjunto de datos de entrenamiento y validación.
- **Tamaño del lote de ejemplos:** cada cuantos se realiza un ajuste sobre la red neuronal.

# Resultados - Longitud del texto (25)

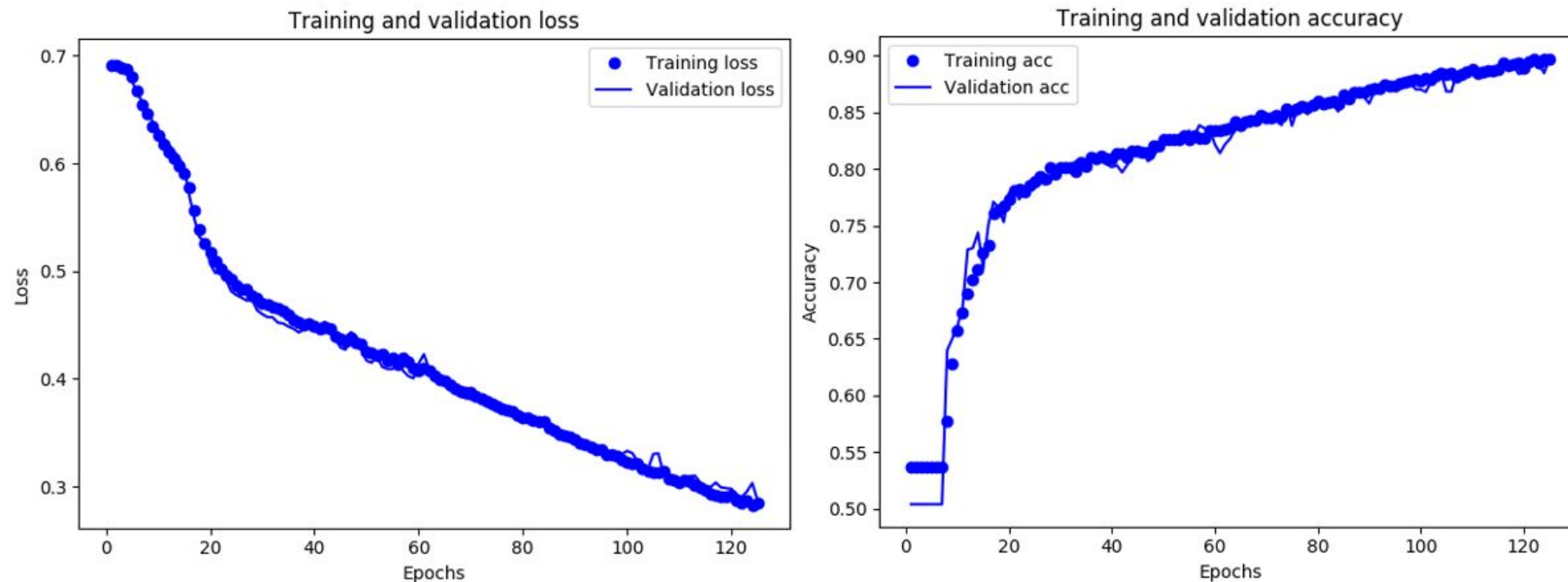
Para text\_len = 25, epoch = 125 y batch\_size = 500



```
Epoch 125/125  
5000/5000 [=====] - 1s 135us/sample - loss: 0.2963 - acc: 0.8896 - val_loss: 0.6688 - val_acc: 0.6846  
10000/10000 [=====] - 0s 37us/sample - loss: 0.6754 - acc: 0.6804
```

# Resultados - Longitud del texto (512)

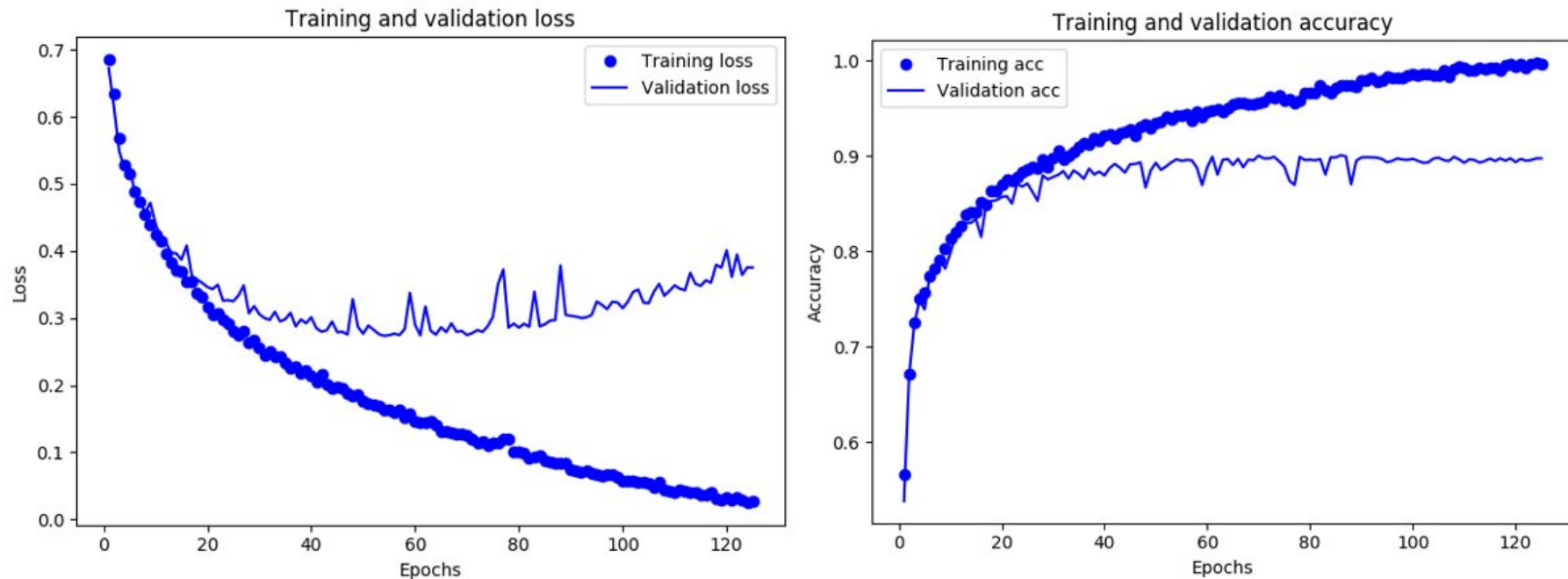
Para text\_len = 512, epoch = 125 y batch size = 500



```
Epoch 125/125  
5000/5000 [=====] - 4s 823us/sample - loss: 0.2841 - acc: 0.8972 - val_loss: 0.2873 - val_acc: 0.8964  
10000/10000 [=====] - 1s 144us/sample - loss: 0.3164 - acc: 0.8728
```

# Resultados - Tamaño de lote de ejemplos (50)

Para text\_len = 256, epoch = 125 y batch size = 50

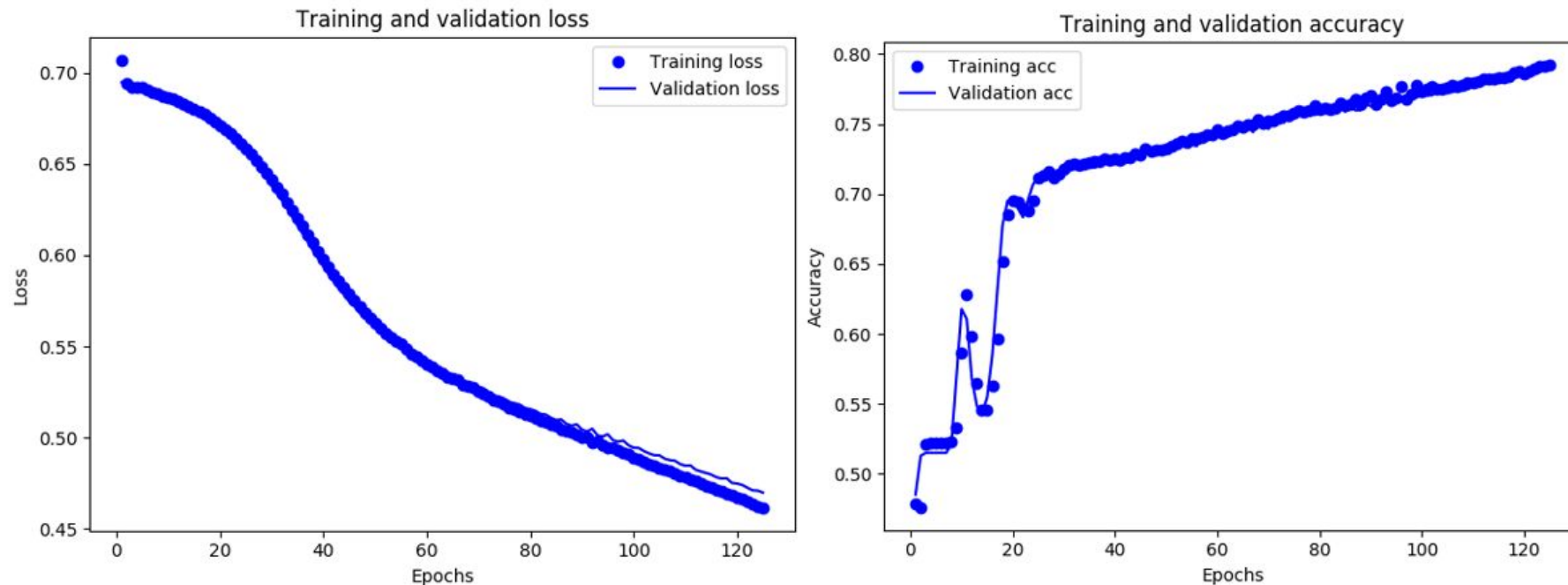


```
Epoch 125/125  
5000/5000 [=====] - 7s 1ms/sample - loss: 0.0261 - acc: 0.9966 - val_loss: 0.3755 - val_acc: 0.8974  
10000/10000 [=====] - 1s 95us/sample - loss: 0.3772 - acc: 0.8983
```



# Resultados - Tamaño de lote de ejemplos (2500)

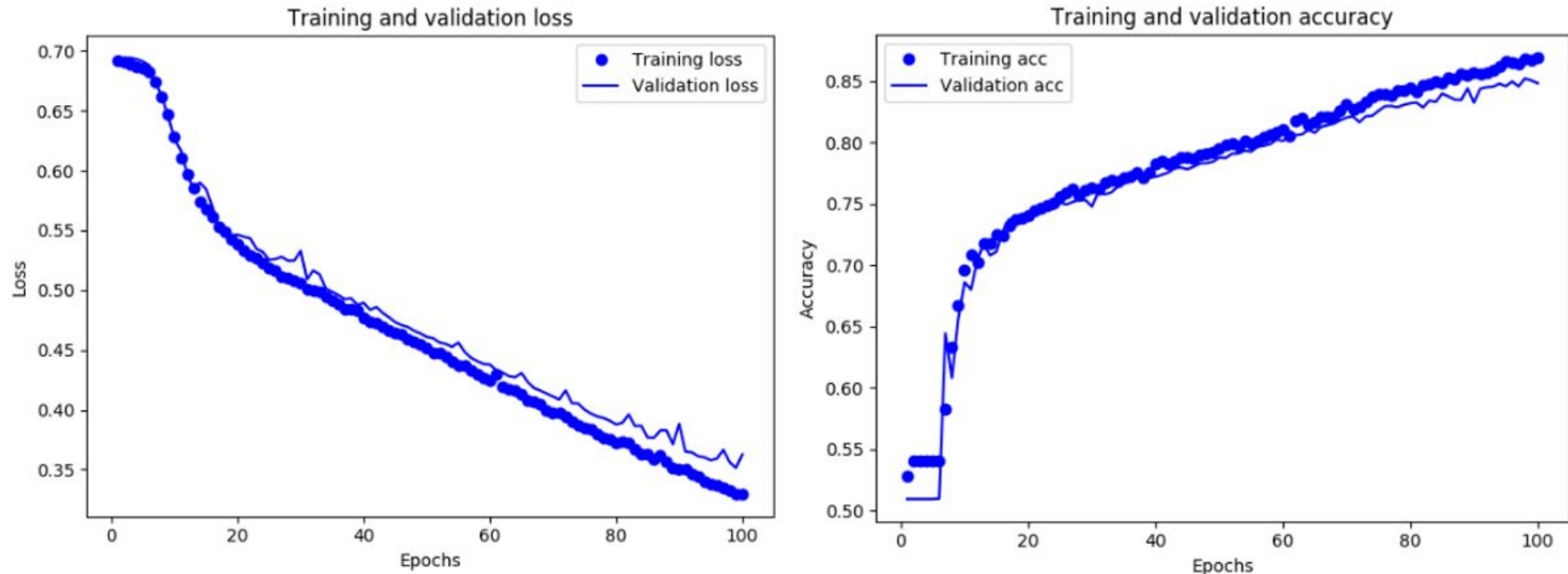
Para text\_len = 256, epoch = 125 y batch size = 2500



```
Epoch 125/125  
5000/5000 [=====] - 2s 382us/sample - loss: 0.4612 - acc: 0.7924 - val_loss: 0.4698 - val_acc: 0.7902  
10000/10000 [=====] - 1s 89us/sample - loss: 0.4855 - acc: 0.7727
```

# Resultados - Épocas (100)

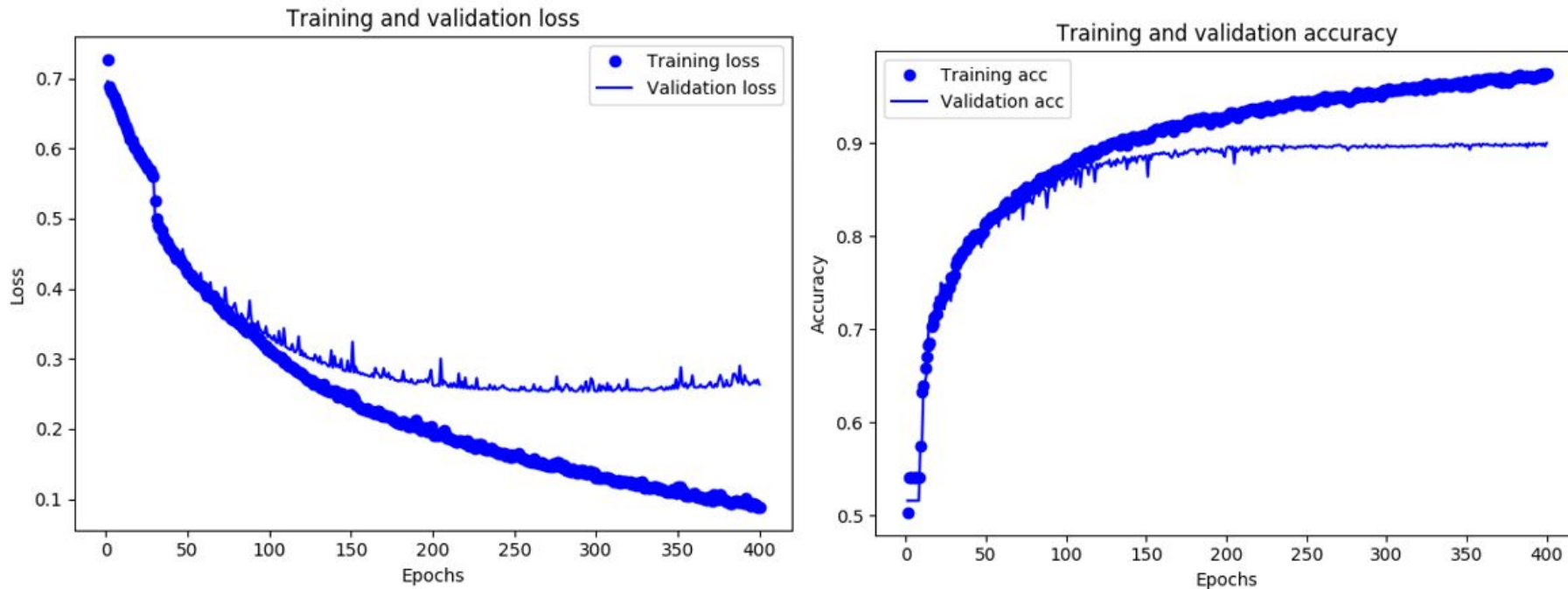
Para text\_len = 256, epoch = 100 y batch size = 500



```
Epoch 100/100  
5000/5000 [=====] - 2s 441us/sample - loss: 0.3292 - acc: 0.8696 - val_loss: 0.3626 - val_acc: 0.8484  
10000/10000 [=====] - 1s 82us/sample - loss: 0.3987 - acc: 0.8205
```

# Resultados - Épocas (400)

Para text\_len = 256, epoch = 400, batch\_size = 500



```
Epoch 400/400  
5000/5000 [=====] - 2s 477us/sample - loss: 0.0877 - acc: 0.9744 - val_loss: 0.2635 - val_acc: 0.9002  
10000/10000 [=====] - 1s 99us/sample - loss: 0.2632 - acc: 0.9016
```

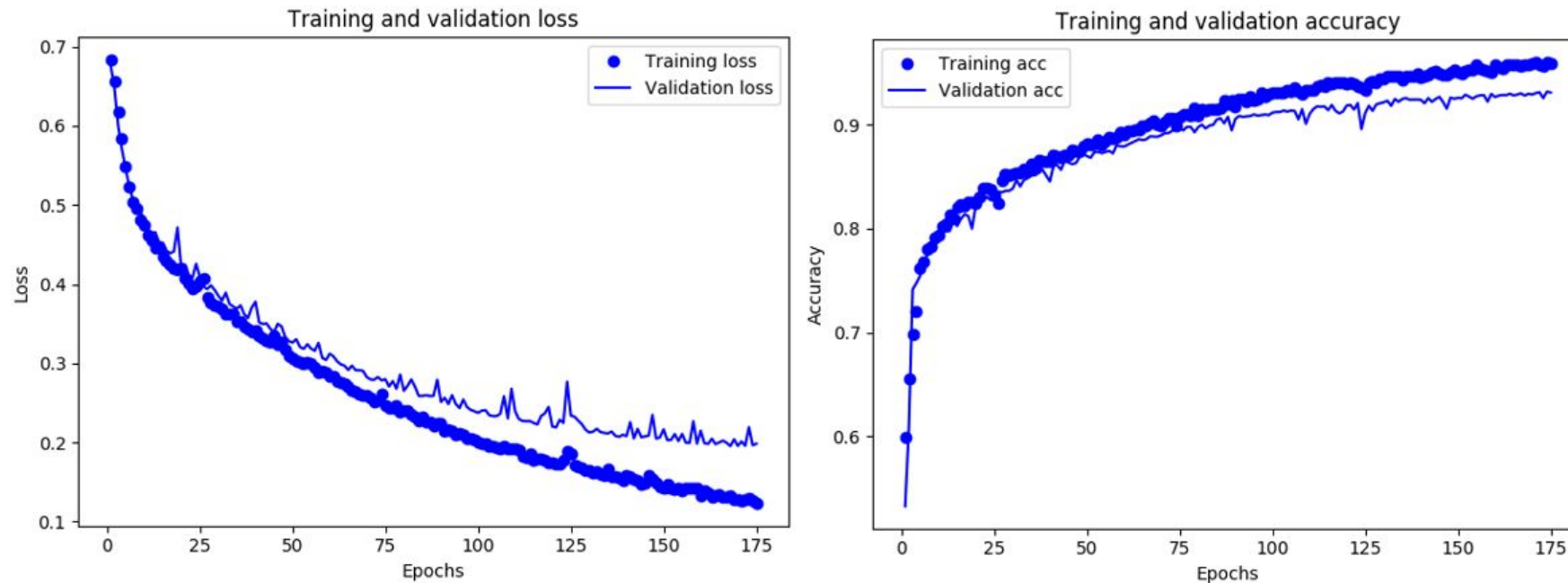
# Resultados - Valores óptimos

---

Valores óptimos	
Longitud del texto	512
Épocas	175
Tamaño de lote de ejemplos	200

# Resultados - Valores óptimos

Para text\_len = 512, epoch = 175, batch\_size = 200



```
Epoch 175/175  
5000/5000 [=====] - 5s 1ms/sample - loss: 0.1227 - acc: 0.9590 - val_loss: 0.1983 - val_acc: 0.9308  
10000/10000 [=====] - 1s 149us/sample - loss: 0.2188 - acc: 0.9183
```

# Resultados - Valores óptimos

---

El algoritmo presenta unos resultados bastante aceptables acercándose a **un nivel de acierto superior al 90% sin cometer overfitting.**



# Índice

---

Introducción

Conceptos básicos

Stack Tecnológico

Trabajo

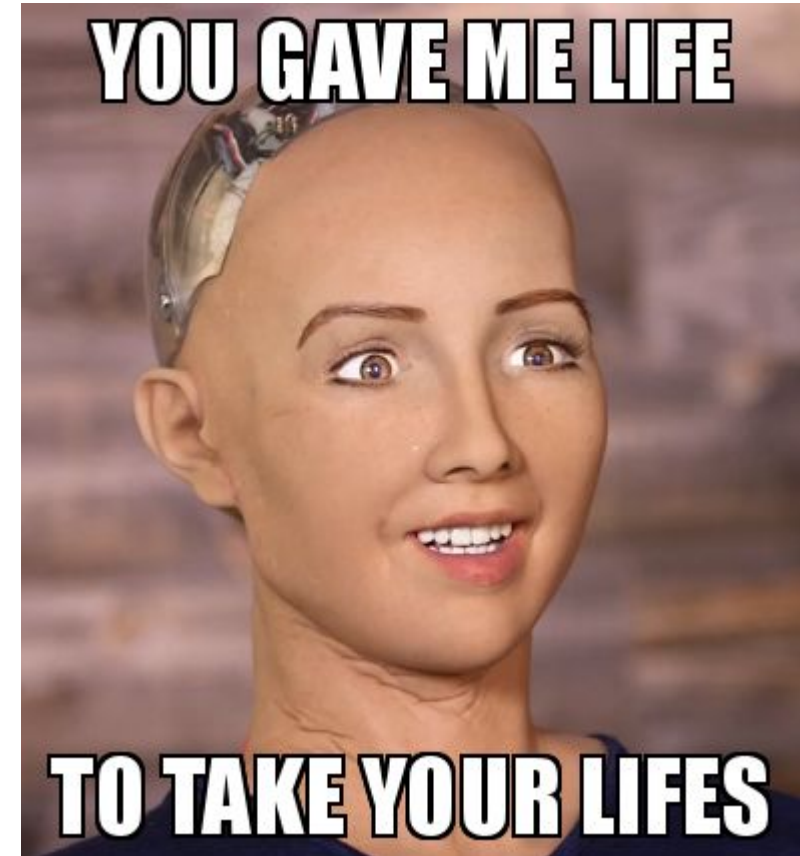
Resultados obtenidos

**Conclusiones**

# Conclusiones

---

La web semántica no dejará de existir, pero los datos serán procesados y creados utilizando inteligencia artificial.







# ¿Preguntas?

Jesús Ortiz Calleja

Juan Carlos Utrilla Martín