Modelos de Inteligencia Artificial

Curso de Especialización de Inteligencia Artificial y Big Data
IES Gran Capitán 2024/25

UNIDAD 3.

Procesamiento del Lenguaje Natural y sus aplicaciones. Potencial y limitaciones.

Índice de contenidos

2. Potencial de las técnicas existentes de procesamiento del lenguaje. Limitaciones.

2.4. Traducción automática

Traducción automática

La traducción automática es la tarea de traducir texto de un idioma a otro. Por ejemplo, del inglés al español.

Los modelos se basan en la arquitectura secuencia a secuencia de **Transformer** [nlp-machine_translation4]

Modelos de traducción automática

Visita la página Hugging Face

Hugging Face

Traducción automática

```
sudo docker pull nvcr.io/nvidia/nemo:1.5.1

sudo docker run --runtime=nvidia -it -v /mnt/d/docker/
 nemo:/nemo/notebooks/host -p 8888:8888 -p 6006:6006 --shm-
 size=16g --ulimit memlock=-1 --ulimit stack=67108864 nvcr.
 io/nvidia/nemo:1.5.1



from nemo.collections.nlp.models import MTEncDecModel

# Se obtiene el listado de modelos preentrenados disponibles

MTEncDecModel.list_available_models()

# Se descarga el modelo de inglés a español

model = MTEncDecModel.from_pretrained("nmt_en_es_transfor-
 mer12x2")

# Se ejecutra sobre la muy conocida frase "Hola Mundo"

translations = model.translate(["Hello World!"],
 source_lang="en", target_lang="es")

print(translations)
```

Traducción automática

```
>>> translations = model.translate(["Hello World!"], source_lang="en", target_lang="es")
>>> print(translations)
['¡Hola mundo!']
>>>
```

Una prueba análoga sobre el acto I de Hamlet arroja el siguiente resultado:

```
>>> translations = model.translate(["Bernardo climbed the stairs to the castle's ramparts. It was a bitter
ly cold night. He made his way carefully through the freezing fog to relieve Francisco of his guard duty.
He saw a dim figure and challenged him. 'Who's there?' 'No, you answer me!' It was Francisco's voice. 'Sto
p and identify yourself!' Bernardo stopped. 'Long live the King!'"], source_lang="en", target_lang="es")
(translations)>>> print(translations)
["Bernardo subió las escaleras hasta las murallas del castillo. Era una noche amargamente fría. Se abrió p
aso cuidadosamente a través de la helada niebla para liberar a Francisco de su deber de guardia. Vio una f
igura tenue y le desafió. '¿Quién está ahí?' '¡No, me respondes!' Era la voz de Francisco. '¡Deténgase e i
dentifíquese!' Bernardo se detuvo. '¡Viva el rey!'"]
>>>
```

2. Potencial de las técnicas existentes de procesamiento del lenguaje. Limitaciones.

2.5. Similitud de textos

Similitud de textos

Algunos modelos (normalmente de tamaño considerable) han sido entrenados de tal forma que las palabras son etiquetadas de manera que aluden a conceptos semánticos.

Ej: un perro es un mamífero y a su vez es un animal.
Un etiquetado correcto situará en lugares más próximos entre sí a un perro y a un gato (dos mamíferos) que un perro y un salmón.
Manazana es una fruta, que es a la vez comida, en una buena lógica sería más "parecido" semánticamente una manzana y una naranja que a un muslo de pollo.

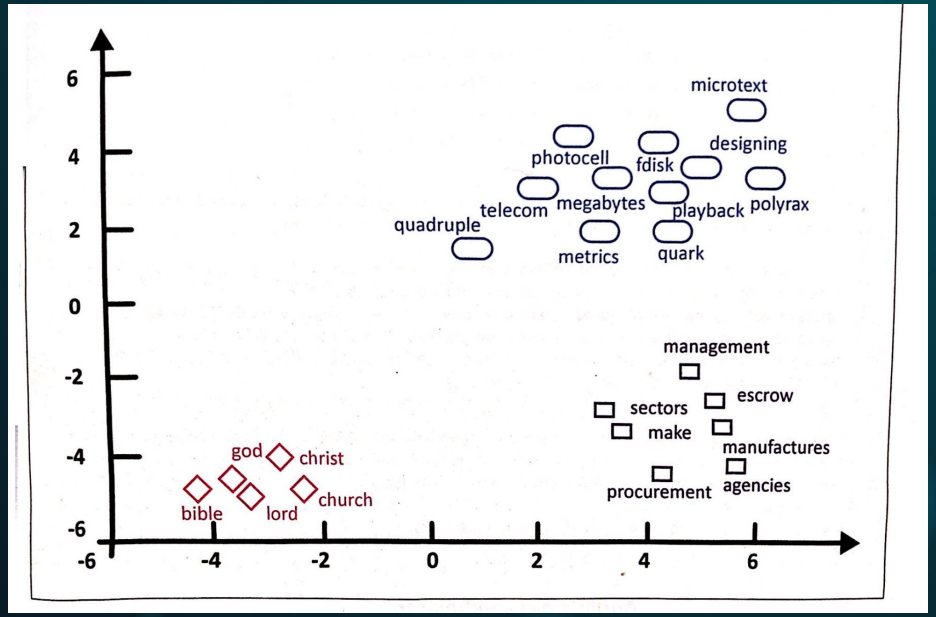
Vectorización de textos

Similitud de textos

Técnica → ***Vectorización y cálculo del coseno***

(dependerá del set de datos que haya sido empleado para llevar el entrenamiento)

Vectorización de textos



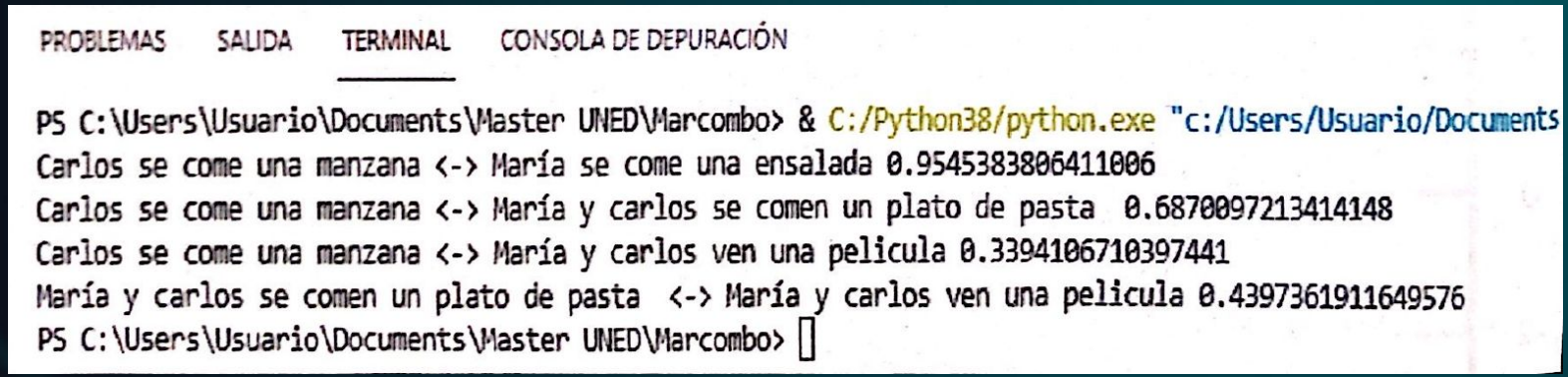
Similitud de textos

Ej. realizado con Spacy, empleando el diccionario español de mayor tamaño, que ha de ser previamente descargado (es_core_news_lg)

```
import spacy
from spacy import displacy
from collections import Counter


import es_core_news_lg
nlp = es_core_news_lg.load()
doc1 = nlp('Carlos se come una manzana')
doc2 = nlp('María se come una ensalada')
doc3 = nlp('María y carlos se comen un plato de pasta ‹)
doc4 = nlp('María y carlos ven una película')
print(doc1, "<->", doc2, doc1.similarity(doc2))
print(doc1, "<->", doc3, doc1.similarity(doc3))
print(doc1, "<->", doc4, doc1.similarity(doc4))
print(doc3, "<->", doc4, doc3.similarity(doc4))
```

Similitud de textos



Las frases 1 y 2 son muy parecidas (95%).

La frase 4 hay una similitud del (43,9%) porque en las dos están María y Carlos pero realizan acciones distintas.

La frase 3 es la que menos similitud tiene con un (33,9%) ya que sólo está Carlos.