Spacy

Ahora haremos uso de Spacy para tokenizar textos , en lugar del nltk del ejercicio anterior.

```
# Importamos las librerías
import os        # Para manejar rutas y carpetas en el so
import spacy     # Librería para procesamiento de lenguaje

# Agarramos la carpeta base donde está instalado spacy
basedir = os.path.dirname(spacy.__path__[0])

# Vemos si el modelo está instalado
# Lo buscamos dentro de la carpeta base + el nombre del modelo
if not os.path.isdir(os.path.join(basedir, 'es_core_news_sm')):
    # Si no está instalado, importamos la función para descargar modelos de spacy
    from spacy.cli import download
    # Y se descarga autoamticamente
    download('es_core_news_sm')
```

```
# Cargamos el modelo
nlp = spacy.load('es_core_news_sm')
```

```
# Cargo un texto defecto (El mismo de la pelicula de "Memorias de un Caracol" usado en la de procesado usando nltk)
text = "Australia, años 70. Grace Pudel es una niña solitaria e inadaptada, aficionada a coleccionar figuras decorativas de caracoles y co
```

```
doc = nlp(text) # Aplicamos el modelo
for idx, token in enumerate(doc): # Recorremos y mostramos los primeros 4 tokens
    print(token.text, "→", token.pos_)
    if idx==3:
        break
```

```
Australia → PROPN
, → PUNCT
años → NOUN
70 → NUM
```

Conclusion

Este código verifica si el modelo de español de spaCy está instalado, si no, lo descarga automáticamente. Luego carga el modelo para

Este código verifica si el modelo de español de spaCy está instalado, si no, lo descarga automáticamente. Luego carga el modelo para procesar texto en español. A continuación, aplica el modelo a un texto de ejemplo y muestra los primeros cuatro tokens, que pueden ser palabras o signos de puntuación, para comprobar que el texto se ha segmentado correctamente.