

Tokenización y lematización de un texto con Spacy

Código

```
import spacy

nlp = spacy.load('es_core_news_sm')

texto = 'Quizás te ha pasado que cuando una persona te habla de cómo  
funcionan las criptomonedas, su idioma se transforma súbitamente en algo  
completamente incomprensible. En cuanto utiliza conceptos como "cadena de  
bloques" o "minería", las cosas se complican. Y cuando el tema del momento  
es "la fusión" de Ethereum, considerada como la mayor revolución de las  
criptomonedas en los últimos años, el desafío por entender qué está  
pasando se vuelve más complejo.'  
doc = nlp(texto)

lista_tokens = [token.lemma_ for token in doc]  
print(lista_tokens)
```

Salida

```
['quizás', 'tú', 'haber', 'pasar', 'que', 'cuando', 'uno', 'persona', 'tú', 'hablar', 'de', 'cómo', 'funcionar', 'el', 'criptom  
oneda', ',', 'su', 'idioma', 'él', 'transformar', 'súbitamente', 'en', 'algo', 'completamente', 'incomprensible', '.', 'en', 'c  
uanto', 'utilizar', 'concepto', 'como', '"', 'cadena', 'de', 'bloque', '"', 'o', '"', 'mineer', '"', ',', 'el', 'cosa', 'él',  
'complicar', '.', 'y', 'cuando', 'el', 'tema', 'del', 'momento', 'ser', '"', 'el', 'fusión', '"', 'de', 'Ethereum', ',', 'consi  
derado', 'como', 'el', 'mayor', 'revolución', 'de', 'el', 'criptomonedas', 'en', 'el', 'último', 'año', ',', 'el', 'desafío', 'p  
or', 'entender', 'qué', 'estar', 'pasar', 'él', 'volver', 'más', 'complejo', '.']
```