

✓ Instrucciones de Actividad

Realice 1 ejemplo comparando lemmatization y stemming

```
# Descargando el paquete
!python -m spacy download es_core_news_sm
```

```
Collecting es-core-news-sm==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/es\_core\_news\_sm-3.8.0/es\_core\_news\_sm-3.8.0-py3-none-any.whl (12.9 MB)
    12.9/12.9 MB 64.0 MB/s eta 0:00:00
Installing collected packages: es-core-news-sm
Successfully installed es-core-news-sm-3.8.0
✓ Download and installation successful
You can now load the package via spacy.load('es_core_news_sm')
⚠ Restart to reload dependencies
If you are in a Jupyter or Colab notebook, you may need to restart Python in
order to load all the package's dependencies. You can do this by selecting the
'Restart kernel' or 'Restart runtime' option.
```

```
# Librerias
import nltk
nltk.download('punkt_tab')
from nltk.stem import SnowballStemmer
import spacy
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Package punkt_tab is already up-to-date!
```

```
# Texto de ejemplo
texto = """Tras acceder a la Universidad, Hinako se muda a un pueblo de la costa. Allí se dedica a coger olas, pues se siente invencible c
```

```
# STEMMING
stemmer = SnowballStemmer("spanish") # Usamos el stemmer español
tokens = texto.lower().split() # Convertimos el texto a minúsculas y lo separamos en palabras
stems = [stemmer.stem(token) for token in tokens] # Obtenemos el stem de cada palabra
```

```
print("Stemming:")
print(list(zip(tokens, stems))) # Mostramos palabra original y su stem
```

```
Stemming:
[('tras', 'tras'), ('acceder', 'acced'), ('a', 'a'), ('la', 'la'), ('universidad,', 'universidad,'), ('hinako', 'hinak'), ('se', 'se'), ('muda', 'mud'), ('a', 'a'), ('
```

```
# LEMMATIZATION
nlp = spacy.load("es_core_news_sm") # Cargamos el modelo de spaCy para español
doc = nlp(texto)
lemmas = [(token.text, token.lemma_) for token in doc] # Obtenemos el lema de cada palabra
```

```
print("\nLemmatization:")
print(lemmas) # Mostramos palabra original y su lema
```

```
└─ Lemmatization:
    [('Tras', 'tras'), ('acceder', 'acceder'), ('a', 'a'), ('la', 'el'), ('Universidad', 'Universidad'), ('', ', ', ', '), ('Hinako', 'Hinako'), ('se', 'él'), ('muda', 'mudar'),
```

Conclusion

En este ejemplo se compararon dos técnicas de procesamiento de lenguaje natural: stemming y lemmatización. Ambas se aplicaron a un mismo texto en español. El stemming, realizado con NLTK, mostró versiones recortadas de las palabras sin considerar su estructura gramatical ni su significado real. Por otro lado, la lemmatización, ejecutada con spaCy, devolvió las formas base correctas de las palabras, respetando su contexto gramatical.