

Stemmatizacao x Lemmatizacao

Comparação lado a lado

Palavra original	Lematização (spaCy)	Stemming (NLTK)	
correndo	correr	corr	
correu	correr	corr	
corrida	corrida	corr	
alunos	aluno	alun	
estudando	estudar	estud	
foi	ir	fo	
melhores	bom	melh	

LEMATIZAÇÃO com spaCy

```
python
CopiarEditar
import spacy
nlp = spacy.load("pt_core_news_sm")

palavras = ["correndo", "correu", "corrida", "alunos", "estudando", "foi", "m
elhores"]
doc = nlp(" ".join(palavras))
lemas = [token.lemma_ for token in doc]
print(lemas)
```

Resultado:

```
CSS
CopiarEditar
```

```
['correr', 'correr', 'corrida', 'aluno', 'estudar', 'ir', 'bom']
```

➡ Resultado **limpo e gramaticalmente correto**.

STEMMING com NLTK (RSLPStemmer)

```
python
CopiarEditar
import nltk
from nltk.stem import RSLPStemmer
nltk.download('rslp')

stemmer = RSLPStemmer()
stems = [stemmer.stem(p) for p in palavras]
print(stems)
```

 **Resultado:**

```
css
CopiarEditar
['corr', 'corr', 'corr', 'alun', 'estud', 'fo', 'melh']
```

Qual usar?

Situação	Melhor escolha
Análise de sentimentos, NLP avançado	Lematização
Busca por radicais semelhantes	Stemming
Alta precisão gramatical	Lematização
Alta performance e simplicidade	Stemming