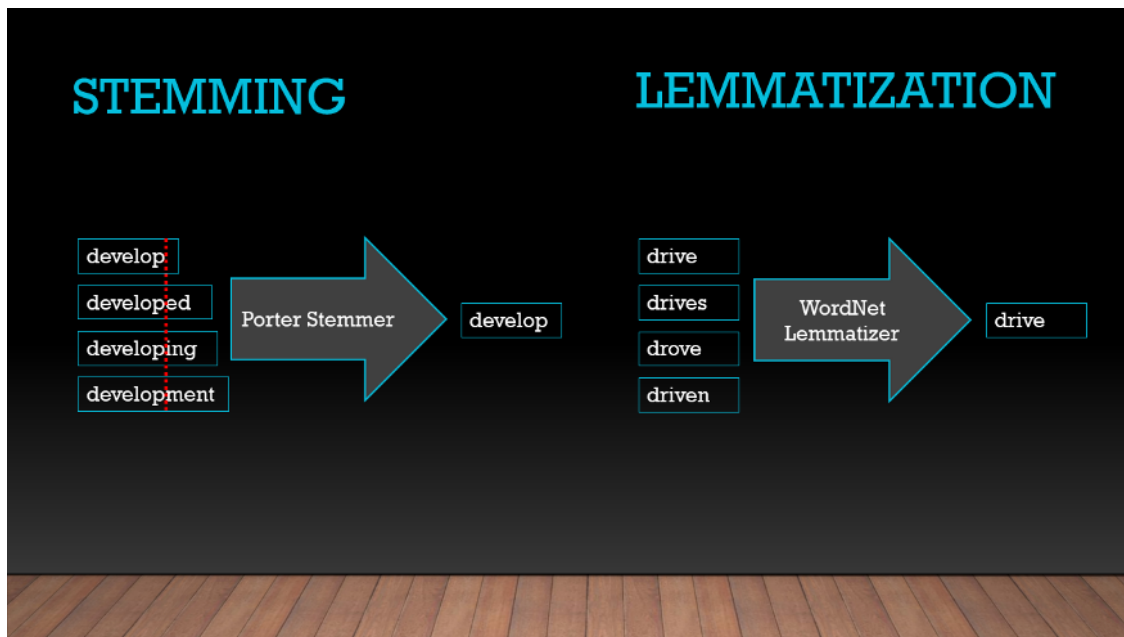


03 Lemmatization

December 20, 2025

1 Lemmatization

- **Base Form Reduction:** Lemmatization reduces words to their meaningful base form(e.g., “running” to “run”)
- **Enhances NLP Accuracy:** It normalizes words variations, improving the accuracy of NLP tasks by treating different word forms as the same.



```
[1]: import spacy
from prettytable import PrettyTable
nlp = spacy.load('en_core_web_sm')
```

```
[2]: text = 'run ran running'
doc = nlp(text)

table = PrettyTable(field_names=['token', 'lemma']) # here lemma means root_
↳ words

for token in doc:
    table.add_row([token.text, token.lemma_])
```

```
print(table)
```

token	lemma
run	run
ran	run
running	run

```
[3]: text = 'The scientists discover new species every year, Last year, they_
↳discovered an ancient artifact. they are discovering new techniques with_
↳their recent discovery'
doc = nlp(text)

table = PrettyTable(field_names=['token', 'lemma', 'pos']) # here lemma means_
↳root words

for token in doc:
    table.add_row([token.text, token.lemma_, token.pos_])

print(table)
```

token	lemma	pos
The	the	DET
scientists	scientist	NOUN
discover	discover	VERB
new	new	ADJ
species	specie	NOUN
every	every	DET
year	year	NOUN
,	,	PUNCT
Last	last	ADJ
year	year	NOUN
,	,	PUNCT
they	they	PRON
discovered	discover	VERB
an	an	DET
ancient	ancient	ADJ
artifact	artifact	NOUN
.	.	PUNCT
they	they	PRON
are	be	AUX
discovering	discover	VERB
new	new	ADJ

	techniques		technique		NOUN	
	with		with		ADP	
	their		their		PRON	
	recent		recent		ADJ	
	discovery		discovery		NOUN	
+-----+-----+-----+						