

Stemming

Stemming word ka ending part remove kar deta hai, aur root word bana deta hai.

Example:

- playing → play
- studies → studi
- running → run

❖ Problem: Kabhi kabhi incorrect root word ban jaata hai.

Lemmatization

Lemmatization dictionary based correct word return karta hai.

Example:

better → good

running → run

studies → study

❖ Lemmatization accurate hota hai but thoda slow hota hai.

🔥 STEP 2: POS Tagging (Part of Speech Tagging)

Sentence me word ka role batata hai:

Noun

Verb

Adjective

Example:

"I love AI"

I → Pronoun

love → Verb

AI → Noun

❖ Chatbot + grammar analysis me use hota hai.

Stemming Code

```
import nltk  
from nltk.stem import PorterStemmer
```

```
stemmer = PorterStemmer()
```

```
words = ["playing", "played", "plays", "studies", "running", "happiness"]
```

```
for word in words:
```

```
    print(word, "=>", stemmer.stem(word))
```

Output:

```
playing => play  
played => play  
plays => play  
studies => studi  
running => run  
happiness => happi
```

❖ Notice:

"studies" => "studi" ❌ (wrong word)

"happiness" => "happi" ❌

Lemmatization Code

```
import nltk  
from nltk.stem import WordNetLemmatizer  
  
nltk.download("wordnet")
```

```
nltk.download("omw-1.4")
```

```
lemmatizer = WordNetLemmatizer()
```

```
words = ["playing", "played", "plays", "studies", "running", "better"]
```

for word in words:

```
    print(word, "=>", lemmatizer.lemmatize(word))
```

❖ Output:

playing => playing

played => played

plays => play

studies => study

running => running

better => better

❖ Notice: Lemmatizer ne "running" ko "run" nahi banaya
kyunki POS tag nahi diya.

Lemmatization with POS

```
import nltk  
from nltk.stem import WordNetLemmatizer
```

```
nltk.download("wordnet")
```

```
nltk.download("omw-1.4")
```

```
lemmatizer = WordNetLemmatizer()
```

```
print("running =>", lemmatizer.lemmatize("running", pos="v"))
print("better =>", lemmatizer.lemmatize("better", pos="a"))
```

❖ **Output:**

running => run

better => good

