

AI_LAB_24_Sep_24

October 25, 2024

#Write a Python program using NLTK to tokenize a sentence, filter out stopwords, and apply stemming ?

```
[1]: import nltk
      from nltk.tokenize import word_tokenize
      from nltk.corpus import stopwords
      from nltk.stem import PorterStemmer
```

```
[2]: nltk.download('punkt')
      nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\manas\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\manas\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[2]: True
```

```
[3]: stop_words = set(stopwords.words('english'))
      stemmer = PorterStemmer()
```

```
[4]: sentence = "This is an example sentence, my name is Manas Kumar Manna I am a_
      ↪student of NIT Hamirpur."
```

```
[5]: tokens = word_tokenize(sentence)
      print("Tokens:", tokens)
```

```
Tokens: ['This', 'is', 'an', 'example', 'sentence', ',', 'my', 'name', 'is',
'Manas', 'Kumar', 'Manna', 'I', 'am', 'a', 'student', 'of', 'NIT', 'Hamirpur',
'.']
```

```
[6]: filtered_tokens = [word for word in tokens if word.lower() not in stop_words]
      print("Filtered Tokens:", filtered_tokens)
```

```
Filtered Tokens: ['example', 'sentence', ',', 'name', 'Manas', 'Kumar', 'Manna',
'student', 'NIT', 'Hamirpur', '.']
```

```
[7]: stemmed_tokens = [stemmer.stem(word) for word in filtered_tokens]
print("Stemmed Tokens:", stemmed_tokens)
```

```
Stemmed Tokens: ['exempl', 'sentenc', ',', 'name', 'mana', 'kumar', 'manna',
'student', 'nit', 'hamirpur', '.']
```

```
[ ]:
```

```
[ ]:
```