Demo of Spacy Package

```python
import spacy
```

```python
spacy.about
```

```
<module 'spacy.about' from 'D:\\iu-training\\Introduction-to-AL-and-ML\\Source-code AI\\spacy-package\\myenv\\Lib\\site-packages\\spacy\\about.py'>
```

```python
spacy.__version__
```

```
'3.8.7'
```

```python
spacy.load('en_core_web_sm')
```

```
<spacy.lang.en.English at 0x1eee7accf20>
```

```python
### Our Steps of Tokensization are : Corpus > documnets > words > vocubalarly
```

```python
corpus = ''' Mr. Shahrukh Likes the Vadapav of Mumbai. And Mr. Salman Khan Like Chaat of the New Delhi'''
```

```python
print('My Corpus:',corpus)
```

```
My Corpus:  Mr. Shahrukh Likes the Vadapav of Mumbai. And Mr. Salman Khan Like Chaat of the New Delhi
```

```python
nlp = spacy.load('en_core_web_sm') # using english model in spacy for nlp. : Object it is a class.
```

```python
documents = nlp(corpus)
```

```python
print('documents or sentences :',documents)
```

```
documents or sentences :  Mr. Shahrukh Likes the Vadapav of Mumbai. And Mr. Salman Khan Like Chaat of the New Delhi
```

```python
print('documents or sentences :',documents.sents)
```

```
documents or sentences : <_cython_3_1_1.generator object at 0x000001EEE9B69990>
```

```python
for sentence in documents.sents:
    print(sentence)
```

```
 Mr. Shahrukh Likes the Vadapav of Mumbai.
And Mr. Salman Khan Like Chaat of the New Delhi
```

```python
sentences_list = [sentence for sentence in documents.sents]
```

```python
print(sentences_list)
```

```
[ Mr. Shahrukh Likes the Vadapav of Mumbai., And Mr. Salman Khan Like Chaat of the New Delhi]
```

```python
words = []
for sentence in documents.sents:
    #print(sentence)
    for word in sentence:
        #print(word)
```

```
Mr.
Shahrukh
Likes
the
Vadapav
of
Mumbai
.
And
Mr.
Salman
Khan
Like
Chaat
of
the
New
Delhi
```

```python
words = []
for sentence in documents.sents:
    #print(sentence)
    for word in sentence:
        words.append(str(word).lower())
```

```python
print('words:',words)
```
words: [' ', 'mr.', 'shahrukh', 'likes', 'the', 'vadapav', 'of', 'mumbai', '.', 'and', 'mr.', 'salman', 'khan', 'like', 'chaat', 'of', 'the', 'new', 'delhi']

```python
print('type of words',type(words[0]))
```
type of words <class 'str'>

```python
print('type of words',type(words[0]))
```
type of words <class 'str'>

```python
print('len of words :',len(words))
```
len of words : 19

```python
vocabs =  set(words)
```

```python
print('vocabs or unique word',vocabs)
```
vocabs or unique word {'likes', 'salman', 'vadapav', 'chaat', 'khan', ' ', 'mumbai', 'new', 'shahrukh', 'mr.', '.', 'and', 'like', 'delhi', 'of', 'the'}

```python
print('len of vocabs:',len(vocabs))
```
len of vocabs: 16