

In [24]:

```
import nltk
nltk.download('punkt')
from nltk.tokenize import sent_tokenize
text="""Dogs ,Canis lupus familiaris, are domesticated mammals,
not natural wild animals. They were originally bred from
wolves. They have been bred by humans for a long time, and
were the first animals ever to be domesticated.
Today, some dogs are used as pets, others are used to help
humans do their work. They are a popular pet because they
are usually playful, friendly, loyal and listen to humans.
Thirty million dogs in the United States are registered as
pets. Dogs eat both meat and vegetables, often mixed
together and sold in stores as dog food. Dogs often have
jobs, including as police dogs, army dogs, assistance dogs,
fire dogs, messenger dogs, hunting dogs, herding dogs, or
rescue dogs.
They are sometimes called "canines" from the Latin word for dog
- canis. Sometimes people also use "dog" to describe other
canids, such as wolves. A baby dog is called a pup or puppy.
A dog is called a puppy until it is about one year old.
Dogs are sometimes referred to as "man's best friend" because
they are kept as domestic pets and are usually loyal and
like being around humans."""
tokenized_text=sent_tokenize(text)
print(tokenized_text)
```

```
[nltk_data] Downloading package punkt to
[nltk_data]      /Users/thanakornpasangthien/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
['Dogs ,Canis lupus familiaris, are domesticated mammals,\nnot natur
al wild animals.', 'They were originally bred from\nwolves.', 'They
have been bred by humans for a long time, and\nwere the first animal
s ever to be domesticated.', 'Today, some dogs are used as pets, oth
ers are used to help\nhumans do their work.', 'They are a popular pe
t because they\nare usually playful, friendly, loyal and listen to h
umans.', 'Thirty million dogs in the United States are registered as
\npets.', 'Dogs eat both meat and vegetables, often mixed\ntogether
and sold in stores as dog food.', 'Dogs often have\njobs, including
as police dogs, army dogs, assistance dogs,\nfire dogs, messenger do
gs, hunting dogs, herding dogs, or\nrescue dogs.', 'They are sometim
es called "canines" from the Latin word for dog\n- canis.', 'Sometim
es people also use "dog" to describe other\ncanids, such as wolve
s.', 'A baby dog is called a pup or puppy.', 'A dog is called a pupp
y until it is about one year old.', 'Dogs are sometimes referred to
as "man's best friend" because\nthey are kept as domestic pets and a
re usually loyal and\nlike being around humans.']
```

In [25]:

```
from nltk.tokenize import word_tokenize
tokenized_word=word_tokenize(text)
print(tokenized_word)
```

```
['Dogs', ',', 'Canis', 'lupus', 'familiaris', ',', 'are', 'domestica
ted', 'mammals', ',', 'not', 'natural', 'wild', 'animals', '.', 'The
y', 'were', 'originally', 'bred', 'from', 'wolves', '.', 'They', 'ha
ve', 'been', 'bred', 'by', 'humans', 'for', 'a', 'long', 'time',
',', 'and', 'were', 'the', 'first', 'animals', 'ever', 'to', 'be',
'domesticated', '.', 'Today', ',', 'some', 'dogs', 'are', 'used', 'a
s', 'pets', ',', 'others', 'are', 'used', 'to', 'help', 'humans', 'd
o', 'their', 'work', '.', 'They', 'are', 'a', 'popular', 'pet', 'bec
ause', 'they', 'are', 'usually', 'playful', ',', 'friendly', ',', 'l
oyal', 'and', 'listen', 'to', 'humans', '.', 'Thirty', 'million', 'd
ogs', 'in', 'the', 'United', 'States', 'are', 'registered', 'as', 'p
ets', '.', 'Dogs', 'eat', 'both', 'meat', 'and', 'vegetables', ',',
'often', 'mixed', 'together', 'and', 'sold', 'in', 'stores', 'as',
'dog', 'food', '.', 'Dogs', 'often', 'have', 'jobs', ',', 'includin
g', 'as', 'police', 'dogs', ',', 'army', 'dogs', ',', 'assistance',
'dogs', ',', 'fire', 'dogs', ',', 'messenger', 'dogs', ',', 'huntin
g', 'dogs', ',', 'herding', 'dogs', ',', 'or', 'rescue', 'dogs',
',', 'They', 'are', 'sometimes', 'called', '`', 'canines', '"', 'f
rom', 'the', 'Latin', 'word', 'for', 'dog', '-', 'canis', '.', 'Some
times', 'people', 'also', 'use', '`', 'dog', '"', 'to', 'describ
e', 'other', 'canids', ',', 'such', 'as', 'wolves', '.', 'A', 'bab
y', 'dog', 'is', 'called', 'a', 'pup', 'or', 'puppy', '.', 'A', 'do
g', 'is', 'called', 'a', 'puppy', 'until', 'it', 'is', 'about', 'on
e', 'year', 'old', '.', 'Dogs', 'are', 'sometimes', 'referred', 't
o', 'as', '`', 'man', ',', 's', 'best', 'friend', '"', 'because',
'they', 'are', 'kept', 'as', 'domestic', 'pets', 'and', 'are', 'usua
lly', 'loyal', 'and', 'like', 'being', 'around', 'humans', '.']
```

In [26]:

```
from nltk.probability import FreqDist
fdist = FreqDist(tokenized_word)
print(fdist)
```

<FreqDist with 120 samples and 229 outcomes>

In [27]:

```
fdist
```

Out[27]:

```
FreqDist({' ': 18, '.': 13, 'are': 10, 'dogs': 10, 'as': 7, 'and':
6, 'to': 5, 'dog': 5, 'Dogs': 4, 'They': 4, ...})
```

In [28]:

```
fdist.most_common(10)
```

Out[28]:

```
[(',', 18),
 ('.', 13),
 ('are', 10),
 ('dogs', 10),
 ('as', 7),
 ('and', 6),
 ('to', 5),
 ('dog', 5),
 ('Dogs', 4),
 ('They', 4)]
```

In [29]:

```
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
nltk_stop_words=set(stopwords.words("english"))
print(nltk_stop_words)
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/thanakornpasangthien/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
{'had', 'an', 'i', 'been', 'but', 'who', 'just', "shouldn't", 'll',
'for', 'herself', "aren't", 'own', 'whom', 'of', 'some', 'your',
'm', 'them', 'that', "don't", 'which', "you've", 'wasn', 'these', 'n
ot', 'same', 'shouldn', 'and', 'have', 'shan', "you'll", 'more', 'ha
dn', 't', 'or', "won't", 'ourselves', 'themselves', 'there', 'agai
n', 'ma', 'weren', 'haven', 'what', 'he', 'do', 'should', 'under',
'isn', 'theirs', "you'd", "you're", 'yourself', 'few', "that'll", 'a
bove', 'having', "wouldn't", "couldn't", 'about', 'was', "doesn't",
"hasn't", 'by', 'to', 'hasn', 'couldn', 'are', 'over', 'each', 'othe
r', "weren't", 'wouldn', 'those', 'you', 'ours', 'very', 'into',
'o', 'am', 'during', 'aren', 'such', 'were', 'where', "should've",
'the', "haven't", 'they', 'does', "isn't", 'myself', 'doing', 'my',
'she', 'is', 'being', 'up', 'why', 'too', 'needn', 'out', 'we', 'the
ir', 'now', 're', "needn't", 'both', 'through', 'can', 've', 'does
n', "shan't", 'all', 'until', 'when', "mustn't", 'from', 'ain', 'onc
e', 'its', "didn't", 'a', 'here', 's', "wasn't", 'on', 'don', 'nor',
'below', 'has', 'won', 'him', 'than', 'it', 'as', "it's", 'while',
"mightn't", 'then', 'himself', 'no', 'y', 'didn', 'me', 'mustn', 'ho
w', 'mightn', 'before', 'after', 'so', 'be', 'with', 'off', 'her',
'this', 'only', 'against', "she's", 'between', 'itself', 'at', 'dow
n', 'yourselves', 'because', 'further', 'will', 'if', 'his', 'in',
'hadn't', 'd', 'most', 'yours', 'hers', 'any', 'did', 'our'}
```

In [31]:

```
filtered_words=[]
for w in tokenized_word:
    if w not in nltk_stop_words:
        filtered_words.append(w)

print("Tokenized Sentence:",tokenized_word)
print("\n")
print("Filterd Sentence:",filtered_words)
```

```
Tokenized Sentence: ['Dogs', ',', 'Canis', 'lupus', 'familiaris',
',', 'are', 'domesticated', 'mammals', ',', 'not', 'natural', 'wil
d', 'animals', '.', 'They', 'were', 'originally', 'bred', 'from', 'w
olves', '.', 'They', 'have', 'been', 'bred', 'by', 'humans', 'for',
'a', 'long', 'time', ',', 'and', 'were', 'the', 'first', 'animals',
'ever', 'to', 'be', 'domesticated', '.', 'Today', ',', 'some', 'dog
s', 'are', 'used', 'as', 'pets', ',', 'others', 'are', 'used', 'to',
'help', 'humans', 'do', 'their', 'work', '.', 'They', 'are', 'a', 'p
opular', 'pet', 'because', 'they', 'are', 'usually', 'playful', ',',
'friendly', ',', 'loyal', 'and', 'listen', 'to', 'humans', '.', 'Thi
rty', 'million', 'dogs', 'in', 'the', 'United', 'States', 'are', 're
gistered', 'as', 'pets', '.', 'Dogs', 'eat', 'both', 'meat', 'and',
'vegetables', ',', 'often', 'mixed', 'together', 'and', 'sold', 'i
n', 'stores', 'as', 'dog', 'food', '.', 'Dogs', 'often', 'have', 'jo
bs', ',', 'including', 'as', 'police', 'dogs', ',', 'army', 'dogs',
',', 'assistance', 'dogs', ',', 'fire', 'dogs', ',', 'messenger', 'd
ogs', ',', 'hunting', 'dogs', ',', 'herding', 'dogs', ',', 'or', 're
scue', 'dogs', '.', 'They', 'are', 'sometimes', 'called', '``', 'can
ines', '``', 'from', 'the', 'Latin', 'word', 'for', 'dog', '-', 'can
is', '.', 'Sometimes', 'people', 'also', 'use', '``', 'dog', '``',
'to', 'describe', 'other', 'canids', ',', 'such', 'as', 'wolves',
',', 'A', 'baby', 'dog', 'is', 'called', 'a', 'pup', 'or', 'puppy',
',', 'A', 'dog', 'is', 'called', 'a', 'puppy', 'until', 'it', 'is',
'about', 'one', 'year', 'old', '.', 'Dogs', 'are', 'sometimes', 'ref
erred', 'to', 'as', '``', 'man', '``', 's', 'best', 'friend', '``',
'because', 'they', 'are', 'kept', 'as', 'domestic', 'pets', 'and',
'are', 'usually', 'loyal', 'and', 'like', 'being', 'around', 'human
s', '.']
```

```
Filterd Sentence: ['Dogs', ',', 'Canis', 'lupus', 'familiaris', ',',
'domesticated', 'mammals', ',', 'natural', 'wild', 'animals', '.',
'They', 'originally', 'bred', 'wolves', '.', 'They', 'bred', 'human
s', 'long', 'time', ',', 'first', 'animals', 'ever', 'domesticated',
',', 'Today', ',', 'dogs', 'used', 'pets', ',', 'others', 'used', 'h
elp', 'humans', 'work', '.', 'They', 'popular', 'pet', 'usually', 'p
layful', ',', 'friendly', ',', 'loyal', 'listen', 'humans', '.', 'Th
irty', 'million', 'dogs', 'United', 'States', 'registered', 'pets',
',', 'Dogs', 'eat', 'meat', 'vegetables', ',', 'often', 'mixed', 'to
gether', 'sold', 'stores', 'dog', 'food', '.', 'Dogs', 'often', 'job
s', ',', 'including', 'police', 'dogs', ',', 'army', 'dogs', ',', 'a
ssistance', 'dogs', ',', 'fire', 'dogs', ',', 'messenger', 'dogs',
',', 'hunting', 'dogs', ',', 'herding', 'dogs', ',', 'rescue', 'dog
s', '.', 'They', 'sometimes', 'called', '``', 'canines', '``', 'Lati
n', 'word', 'dog', '-', 'canis', '.', 'Sometimes', 'people', 'also',
'use', '``', 'dog', '``', 'describe', 'canids', ',', 'wolves', '.',
'A', 'baby', 'dog', 'called', 'pup', 'puppy', '.', 'A', 'dog', 'call
ed', 'puppy', 'one', 'year', 'old', '.', 'Dogs', 'sometimes', 'refer
red', '``', 'man', '``', 'best', 'friend', '``', 'kept', 'domestic',
'pets', 'usually', 'loyal', 'like', 'around', 'humans', '.']
```

In [32]:

```
from nltk.stem import PorterStemmer  
stem = PorterStemmer()  
word = "connection"  
stem.stem(word)
```

Out[32]:

```
'connect'
```

In [33]:

```
from nltk.stem import PorterStemmer
from nltk.tokenize import sent_tokenize, word_tokenize
stem = PorterStemmer()
stemmed_words=[]
for w in filtered_words:
    stemmed_words.append(stem.stem(w))
print("Filtered Sentence:",filtered_words)
print("\n")
print("Stemmed Sentence:",stemmed_words)
```

```
Filtered Sentence: ['Dogs', ',', 'Canis', 'lupus', 'familiaris',
',', 'domesticated', 'mammals', ',', 'natural', 'wild', 'animals',
',', 'They', 'originally', 'bred', 'wolves', ',', 'They', 'bred', 'h
umans', 'long', 'time', ',', 'first', 'animals', 'ever', 'domesticat
ed', ',', 'Today', ',', 'dogs', 'used', 'pets', ',', 'others', 'use
d', 'help', 'humans', 'work', ',', 'They', 'popular', 'pet', 'usuall
y', 'playful', ',', 'friendly', ',', 'loyal', 'listen', 'humans',
',', 'Thirty', 'million', 'dogs', 'United', 'States', 'registered',
'pets', ',', 'Dogs', 'eat', 'meat', 'vegetables', ',', 'often', 'mix
ed', 'together', 'sold', 'stores', 'dog', 'food', ',', 'Dogs', 'ofte
n', 'jobs', ',', 'including', 'police', 'dogs', ',', 'army', 'dogs',
',', 'assistance', 'dogs', ',', 'fire', 'dogs', ',', 'messenger', 'd
ogs', ',', 'hunting', 'dogs', ',', 'herding', 'dogs', ',', 'rescue',
'dogs', ',', 'They', 'sometimes', 'called', '``', 'canines', '"',
'Latin', 'word', 'dog', '-', 'canis', ',', 'Sometimes', 'people', 'a
lso', 'use', '``', 'dog', '"', 'describe', 'canids', ',', 'wolves',
',', 'A', 'baby', 'dog', 'called', 'pup', 'puppy', ',', 'A', 'dog',
'called', 'puppy', 'one', 'year', 'old', ',', 'Dogs', 'sometimes',
'referred', '``', 'man', ',', 'best', 'friend', '"', 'kept', 'domes
tic', 'pets', 'usually', 'loyal', 'like', 'around', 'humans', ',']
```

```
Stemmed Sentence: ['dog', ',', 'cani', 'lupu', 'familiari', ',', 'do
mest', 'mammal', ',', 'natur', 'wild', 'anim', ',', 'they', 'origi
n', 'bred', 'wolv', ',', 'they', 'bred', 'human', 'long', 'time',
',', 'first', 'anim', 'ever', 'domest', ',', 'today', ',', 'dog', 'u
se', 'pet', ',', 'other', 'use', 'help', 'human', 'work', ',', 'the
y', 'popular', 'pet', 'usual', 'play', ',', 'friendli', ',', 'loya
l', 'listen', 'human', ',', 'thirti', 'million', 'dog', 'unit', 'sta
te', 'regist', 'pet', ',', 'dog', 'eat', 'meat', 'veget', ',', 'ofte
n', 'mix', 'togeth', 'sold', 'store', 'dog', 'food', ',', 'dog', 'of
ten', 'job', ',', 'includ', 'polic', 'dog', ',', 'armi', 'dog', ',',
'assist', 'dog', ',', 'fire', 'dog', ',', 'messeng', 'dog', ',', 'hu
nt', 'dog', ',', 'herd', 'dog', ',', 'rescu', 'dog', ',', 'they', 's
ometim', 'call', '``', 'canin', '"', 'latin', 'word', 'dog', '-',
'cani', ',', 'sometim', 'peopl', 'also', 'use', '``', 'dog', '"',
'describ', 'canid', ',', 'wolv', ',', 'A', 'babi', 'dog', 'call', 'p
up', 'puppi', ',', 'A', 'dog', 'call', 'puppi', 'one', 'year', 'ol
d', ',', 'dog', 'sometim', 'refer', '``', 'man', ',', 'best', 'frien
d', '"', 'kept', 'domest', 'pet', 'usual', 'loyal', 'like', 'aroun
d', 'human', ',']
```

In [35]:

```
import nltk
nltk.download('wordnet')
from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()
word = "connecting"
lem.lemmatize(word, pos="v")
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]      /Users/thanakornpasangthien/nltk_data...
[nltk_data]   Unzipping corpora/wordnet.zip.
```

Out[35]:

'connect'

In [36]:

```
from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()
word = "flying"
print(lem.lemmatize(word, pos="v"))
print(lem.lemmatize(word, pos="n"))
```

fly
flying

In [38]:

```
import nltk
nltk.download('averaged_perceptron_tagger')
sent = "I am studying at ICT"
tokens=nltk.word_tokenize(sent)
print(tokens)
nltk.pos_tag(tokens)
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]      /Users/thanakornpasangthien/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
['I', 'am', 'studying', 'at', 'ICT']
```

Out[38]:

```
[('I', 'PRP'),
 ('am', 'VBP'),
 ('studying', 'VBG'),
 ('at', 'IN'),
 ('ICT', 'NNP')]
```