In[1]:

```
import nltk
```

In[2]:

```
from nltk.tokenize import word_tokenize
from nltk.tokenize import sent_tokenize
```

In[50]:

```
a = "Hello and welcome friends to NLP workshop My name is shridhar mankar I will be teaching you NLP from scratch"
```

In[51]:

```
A = word_tokenize(a)
A
```

In[52]:

```
S = sent_tokenize(a)
S
```

Type, Length and Frequency Checking

In[53]:

```
type(A),len(A)
```

In[7]:

```
from nltk.probability import FreqDist
frequency = FreqDist()


# In[8]:


for i in A:
 frequency[i] = frequency[i]+1

frequency


# # Stemming

# In[9]:


from nltk.stem import PorterStemmer
pst = PorterStemmer()


# In[10]:


pst.stem('Making')


# In[11]:


for i in A:
    print(pst.stem(i))


# In[12]:


pst.stem('universal')


# In[13]:
```

```
pst.stem('universe')
```

In[14]:

```
pst.stem('university')
```

In[15]:

```
pst.stem('alumni')
```

In[16]:

```
pst.stem('alumnus')
```

Lemmatization

In[17]:

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

In[18]:

```
lemmatizer = WordNetLemmatizer()
```

In[19]:

```
pst.stem('trouble')
```

In[20]:

```
lemmatizer.lemmatize('trouble')
```

In[21]:

```
for i in A:
    print(lemmatizer.lemmatize(i))
```

In[22]:

```
lemmatizer.lemmatize('alumnus')
```

In[23]:

```
lemmatizer.lemmatize('alumni')
```

In[24]:

```
lemmatizer.lemmatize('universe')
```

In[25]:

```
lemmatizer.lemmatize('university')
```

pos_tag

In[26]:

```
nltk.download('averaged_perceptron_tagger')
```

In[27]:

```
for i in A:
 print(nltk.pos_tag([i]))
```

Named entity recognition

In[28]:

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
```

In[29]:

```
text= "'Harry Lives in New York"'
words= word_tokenize(text)
postags=pos_tag(words)
```

In[30]:

```
tree = nltk.ne_chunk(postags)
print(tree)
```

In[60]:

```
text= 'John wants a new Samsung device from Pune'
words= word_tokenize(text)
postags=pos_tag(words)
```

In[61]:

```
tree = nltk.ne_chunk(postags)
print(tree)


# # Stopwords

# In[33]:


from nltk.corpus import stopwords


# In[34]:


stop_words = set(stopwords.words('english'))


# In[35]:


stop_words


# In[36]:


msg = "My name is shridhar mankar, I love making videos and watching kdrama. My speciality
is making things easy"

words = word_tokenize(msg)

filtered_sentence = []

for w in words:
    if w not in stop_words:
        filtered_sentence.append(w)

print(words)
print(filtered_sentence)
```