TASK 6 : CHINESE SEGMENTATION USING JIEBA

In [1]:

```python
import jieba
seg = jieba.cut("私の名前はガリガッタチャイタニアです", cut_all = True)
print(" ".join(seg))
```

```
Building prefix dict from the default dictionary ...
Dumping model to file cache C:\Users\maste\AppData\Local\Temp\jieba.cache
Loading model cost 0.793 seconds.
Prefix dict has been built successfully.
```

私 の 名 前 はガリガッタチャイタニアです

TASK 7 : BASIC TEXT PROCESSING PIPELINE

In [2]:

```python
import nltk
sent = 'Become an expert in NLP'
words = nltk.word_tokenize(sent)
print(words)
```

['Become', 'an', 'expert', 'in', 'NLP']

In [9]:

```python
texts = ''' Born in Nine Mile, British Jamaica, Marley began his professional musical career in 19
63, after forming Bob Marley and the Wailers. The group released its debut studio album The Wailin
g Wailers in 1965, which contained the single "One Love/People Get Ready"; the song was popular wo
rldwide, and established the group as a rising figure in reggae.[8] The Wailers subsequently relea
sed eleven further studio albums; while initially employing louder instrumentation and singing, th
e group began engaging in rhythmic-based song construction in the late 1960s and early 1970s, whic
h coincided with the singer's conversion to Rastafarianism. During this period Marley relocated to
London, and the group typified their musical shift with the release of the album The Best of The W
ailers (1971).[9]
The group attained international success after the release of the albums Catch a Fire and Burnin'
(both 1973), and forged a reputation as touring artists. Following the disbandment of the Wailers
a year later, Marley went on to release his solo material under the band's name.[10] His debut stu
dio album Natty Dread (1974) received positive reception, as did its follow-up Rastaman Vibration
(1976). A few months after the album's release Marley survived an assassination attempt at his hom
e in Jamaica, which prompted him to permanently relocate to London. During his time in London he r
ecorded the album Exodus (1977); it incorporated elements of blues, soul, and British rock, enjoye
d widespread commercial and critical success.
In 1977, Marley was diagnosed with acral lentiginous melanoma; he died as a result of the illness
in 1981. His fans around the world expressed their grief, and he received a state funeral in Jamai
ca. The greatest hits album Legend was released in 1984, and became the best-selling reggae album
of all time.[11] Marley also ranks as one of the best-selling music artists of all time, with esti
mated sales of more than 75 million records worldwide.[12] He was posthumously honored by Jamaica
soon after his death with a designated Order of Merit by his nation. In 1994, he was inducted into
the Rock and Roll Hall of Fame. Rolling Stone ranked him No. 11 on its list of the 100 Greatest Ar
tists of All Time.'''


word = nltk.word_tokenize(texts)
print(word)
```

['Born', 'in', 'Nine', 'Mile', ',', 'British', 'Jamaica', ',', 'Marley', 'began', 'his',
'professional', 'musical', 'career', 'in', '1963', ',', 'after', 'forming', 'Bob', 'Marley',
'and', 'the', 'Wailers', '.', 'The', 'group', 'released', 'its', 'debut', 'studio', 'album', 'The'
, 'Wailing', 'Wailers', 'in', '1965', ',', 'which', 'contained', 'the', 'single', '``', 'One',
'Love/People', 'Get', 'Ready', "''", ';', 'the', 'song', 'was', 'popular', 'worldwide', ',',
'and', 'established', 'the', 'group', 'as', 'a', 'rising', 'figure', 'in', 'reggae', '.', '[', '8'
, ']', 'The', 'Wailers', 'subsequently', 'released', 'eleven', 'further', 'studio', 'albums', ';',
'while', 'initially', 'employing', 'louder', 'instrumentation', 'and', 'singing', ',', 'the',

'group', 'began', 'engaging', 'in', 'rhythmic-based', 'song', 'construction', 'in', 'the', 'late',
'1960s', 'and', 'early', '1970s', ',', 'which', 'coincided', 'with', 'the', 'singer', "'s", 'conve
rsion', 'to', 'Rastafarianism', '.', 'During', 'this', 'period', 'Marley', 'relocated', 'to',
'London', ',', 'and', 'the', 'group', 'typified', 'their', 'musical', 'shift', 'with', 'the', 'rel
ease', 'of', 'the', 'album', 'The', 'Best', 'of', 'The', 'Wailers', '(', '1971', ')', '.', '[', '9
', ']', 'The', 'group', 'attained', 'international', 'success', 'after', 'the', 'release', 'of', '
the', 'albums', 'Catch', 'a', 'Fire', 'and', 'Burnin', "'", '(', 'both', '1973', ')', ',', 'and',
'forged', 'a', 'reputation', 'as', 'touring', 'artists', '.', 'Following', 'the', 'disbandment', '
of', 'the', 'Wailers', 'a', 'year', 'later', ',', 'Marley', 'went', 'on', 'to', 'release', 'his',
'solo', 'material', 'under', 'the', 'band', "'s", 'name', '.', '[', '10', ']', 'His', 'debut', 'st
udio', 'album', 'Natty', 'Dread', '(', '1974', ')', 'received', 'positive', 'reception', ',', 'as'
, 'did', 'its', 'follow-up', 'Rastaman', 'Vibration', '(', '1976', ')', '.', 'A', 'few', 'months',
'after', 'the', 'album', "'s", 'release', 'Marley', 'survived', 'an', 'assassination', 'attempt',
'at', 'his', 'home', 'in', 'Jamaica', ',', 'which', 'prompted', 'him', 'to', 'permanently',
'relocate', 'to', 'London', '.', 'During', 'his', 'time', 'in', 'London', 'he', 'recorded', 'the',
'album', 'Exodus', '(', '1977', ')', ';', 'it', 'incorporated', 'elements', 'of', 'blues', ',', 's
oul', ',', 'and', 'British', 'rock', ',', 'enjoyed', 'widespread', 'commercial', 'and',
'critical', 'success', '.', 'In', '1977', ',', 'Marley', 'was', 'diagnosed', 'with', 'acral',
'lentiginous', 'melanoma', ';', 'he', 'died', 'as', 'a', 'result', 'of', 'the', 'illness', 'in',
'1981', '.', 'His', 'fans', 'around', 'the', 'world', 'expressed', 'their', 'grief', ',', 'and', '
he', 'received', 'a', 'state', 'funeral', 'in', 'Jamaica', '.', 'The', 'greatest', 'hits',
'album', 'Legend', 'was', 'released', 'in', '1984', ',', 'and', 'became', 'the', 'best-selling', '
reggae', 'album', 'of', 'all', 'time', '.', '[', '11', ']', 'Marley', 'also', 'ranks', 'as',
'one', 'of', 'the', 'best-selling', 'music', 'artists', 'of', 'all', 'time', ',', 'with',
'estimated', 'sales', 'of', 'more', 'than', '75', 'million', 'records', 'worldwide', '.', '[',
'12', ']', 'He', 'was', 'posthumously', 'honored', 'by', 'Jamaica', 'soon', 'after', 'his',
'death', 'with', 'a', 'designated', 'Order', 'of', 'Merit', 'by', 'his', 'nation', '.', 'In',
'1994', ',', 'he', 'was', 'inducted', 'into', 'the', 'Rock', 'and', 'Roll', 'Hall', 'of', 'Fame',
'.', 'Rolling', 'Stone', 'ranked', 'him', 'No', '.', '11', 'on', 'its', 'list', 'of', 'the',
'100', 'Greatest', 'Artists', 'of', 'All', 'Time', '.']

In [7]:

```
tagged = nltk.pos_tag(word)
print(tagged)
```

[('11', 'CD'), ('on', 'IN'), ('its', 'PRP$'), ('list', 'NN'), ('of', 'IN'), ('the', 'DT'), ('100',
'CD'), ('Greatest', 'JJS'), ('Artists', 'NNS'), ('of', 'IN'), ('All', 'DT'), ('Time', 'NNP'), ('.'
, '.')]

In [ ]: