Saloni Bhingardive Roll No. 23 CSE (DS) NLP PRACTICAL No. 03

```
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
from nltk.stem.porter import PorterStemmer
```

```
text="Martin Luther King Jr. led many demonstrations against racism. He delivered his message in a non-violent manner. Some members of his mo
```

```
PS=PorterStemmer()
```

```
words=word_tokenize(text)
for w in words:
  print(w+ ":"+ PS.stem(w))
```

```
people:peopl
.:.
During:dure
Nelson:nelson
Mandela:mandela
’:’
s:s
best:best
known:known
speech:speech
in:in
1994:1994
,:,
he:he
recited:recit
“:“
Our:our
Deepest:deepest
Fear:fear
”:”
,:,
an:an
inspirational:inspir
poem:poem
by:by
Marianne:mariann
Williamson:williamson
.:.
Mandela:mandela
initially:initi
avoided:avoid
violence:violenc
but:but
ended:end
up:up
resorting:resort
to:to
it:it
following:follow
the:the
massacre:massacr
of:of
unarmed:unarm
black:black
Africans:african
by:by
the:the
government:govern
.:.
Martin:martin
Luther:luther
King:king
Jr.:jr.
was:wa
assassinated:assassin
in:in
1968:1968
.:.
```

```
my_word="Determination"
print(PS.stem(my_word))
```

```
determin
```

```
one="Alumns"
print(PS.stem(one))
```

```
alumn
```

```
two="alumni"
print(PS.stem(two))
```

```
alumni
```

```
three="alumne"
print(PS.stem(three))
```

```
alumn
```

```
four="university"
print(PS.stem(four))
```

```
univers
```

```
five="universal"
print(PS.stem(five))
```

```
univers
```

```
s="universe"
print(PS.stem(s))
```

```
univers
```

```
d="Determined"
print(PS.stem(d))
```

```
determin
```

```
d1="Determining"
print(PS.stem(d1))
```

```
determin
```

```
d2="Determination"
print(PS.stem(d2))
```

```
determin
```

```
h="Hyperlink"
print(PS.stem(h))
```

```
hyperlink
```

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

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
True
```

```
lm=WordNetLemmatizer()
```

```
text="I was running to become an athlete and then I went home"
```

```
words= word_tokenize(text)
for w in words:
  print(w+":"+lm.lemmatize(w))
```

```
I:I
was:wa
running:running
to:to
become:become
an:an
athlete:athlete
and:and
then:then
I:I
went:went
home:home
```

```
w1="chocolates"
print(lm.lemmatize(w1))
```

```
chocolate
```

```
print("hats :", lm.lemmatize("hats"))
print("algae :", lm.lemmatize("algae"))
 # a denotes adjective in "pos"
print("better :", lm.lemmatize("better", pos="a"))
```

```
hats : hat
algae : algae
better : good
```

```
import nltk
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize
from nltk.corpus import wordnet

# Downloading necessary NLTK resources
nltk.download('wordnet')
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')


lm = WordNetLemmatizer()


text = "I was running to become an athlete and then I went home"

# Tokenize the text into words
words = word_tokenize("Danny was running late, he was supposed to reach school by 9, but it was already 9:15. I was running to become an athl

# Function to convert NLTK POS tags to WordNet POS tags
def get_wordnet_pos(tag):
    if tag.startswith('J'):
        return wordnet.ADJ
    elif tag.startswith('V'):
        return wordnet.VERB
    elif tag.startswith('N'):
        return wordnet.NOUN
    elif tag.startswith('R'):
        return wordnet.ADV
    else:
        return wordnet.NOUN

# Getting POS tags for each word
pos_tags = nltk.pos_tag(words)

# Lemmatizing each word with its POS tag
for word, tag in pos_tags:
    wordnet_pos = get_wordnet_pos(tag)
    lemmatized_word = lm.lemmatize(word, pos=wordnet_pos)
    print(f"{word}:{lemmatized_word}")
```

```
Danny:Danny
was:be
running:run
late:late
```

```
,:,
he:he
was:be
supposed:suppose
to:to
reach:reach
school:school
by:by
9:9
,:,
but:but
it:it
was:be
already:already
9:15:9:15
.:.
I:I
was:be
running:run
to:to
become:become
an:an
athlete:athlete
and:and
then:then
I:I
went:go
home:home
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
```

IMP NOTE : Limitation of Tokenization : Tokenization is unable to capture the meaning of the sentence hence, results in ambiguity.