```
#Words stripping out punctuations
def CountWordsPunctuations(text):
   # Dictionaries to Store individual words and punctuation
   wordDict = {}
   punctuationDict = {}
   wordArr = text.split()
   # Look into each word and remove the trailing punctuation (ex: Hello, -> Hello)
   for w in wordArr:
       # get rid of punctuations from leading and trailing
       w = w.strip(".,!?;:-'\"()[]{\}").lower()
       # Update word frequency dictionary
       wordDict[w] = wordDict.get(w, 0) + 1

   # Now look into each character and count the punctuations
   for char in text:
       # Check for punctuation character
       if char in ".,!?;:-'\"()[]{\}":
           # Update punctuation frequency dictionary
           punctuationDict[char] = punctuationDict.get(char, 0) + 1

   return wordDict, punctuationDict



#words with punctuations. in that case, same words with different punctuations will be counted
as different words
#ex: Hello, Hello! Hello? Hello. Hello; Hello: Hello' Hello" Hello( Hello) Hello[ Hello]
Hello{ Hello}
def CountWordsNoPunctuations(text):
   # Dictionaries to Store individual words and punctuations
   wordDict = {}
   punctuationDict = {}
   wordArr = text.split()
   # Look into each word and remove the trailing punctuation (ex: Hello, -> Hello)
   for w in wordArr:
       # get rid of punctuations from leading and trailing
       w = w.lower()
       # Update word frequency dictionary
       wordDict[w] = wordDict.get(w, 0) + 1

   # Now look into each character and count the punctuations
   for char in text:
       # Check for punctuation character
       if char in ".,!?;:-'\"()[]{\}":
           # Update punctuation frequency dictionary
           punctuationDict[char] = punctuationDict.get(char, 0) + 1
```

```
    return wordDict, punctuationDict

#Test the code

input_text = open('input.txt', 'r').read()

#stripping out punctuations. Ex: 'Hello', 'Hello!' both will be 'hello' and count as same word
wordDict1, punctuationDict1 = CountWordsPunctuations(input_text)
wordDict1 = {k:v for k, v in sorted(wordDict1.items(), key=lambda x: x[1], reverse=True)}
punctuationDict1 = {k:v for k, v in sorted(punctuationDict1.items(), key=lambda x: x[1],
reverse=True)}

#with punctuation. Ex: 'Hello', 'Hello!' both will be counted as different words
wordDict2, punctuationDict2 = CountWordsNoPunctuations(input_text)
wordDict2 = {k:v for k, v in sorted(wordDict2.items(), key=lambda x: x[1], reverse=True)}
punctuationDict2 = {k:v for k, v in sorted(punctuationDict2.items(), key=lambda x: x[1],
reverse=True)}
print("Without punctuation:")
print("Word  ==»  Freq:")
for w, c in wordDict1.items():
    print(f" {w}  ==»  {c}")

print("\nPunctuation ==» Freq:")
for p, c in punctuationDict1.items():
    print(f" {p}  ==»   {c}")

print("\n\n\n\nWith punctuation:")
print("Word  ==»  Freq:")
for w, c in wordDict2.items():
    print(f" {w}  ==»  {c}")

print("\nPunctuation ==» Freq:")
for p, c in punctuationDict2.items():
    print(f" {p}  ==»   {c}")
```

Input File Text:

Natural Language Processing (NLP) represents the {intricate} {intersection} [of] (artificial) [intelligence] and linguistics, deploying advanced algorithms and statistical models to unravel the subtle nuances embedded within human language.
Its overarching goal is to surmount the formidable challenges posed by linguistic diversity, context ambiguity, and semantic intricacies, thereby fostering seamless interaction between humans and machines.
NLP is an ever-evolving domain, delving into tasks of immense complexity such as sentiment analysis, named entity recognition, machine translation, and abstractive text summarization.

The field's trajectory has been substantially influenced by the advent of sophisticated pre-trained language models like BERT and GPT, elevating the capacity of machines to not merely parse language but to assimilate intricate contextual dependencies and generate human-like responses.
Despite these strides, NLP grapples with formidable challenges, including the disambiguation of polysemous words, understanding context nuances, and mitigating biases inherent in language corpora.
The application spectrum of NLP extends across multifaceted domains, from refining search engine algorithms and augmenting voice-driven virtual assistants to revolutionizing language translation services and bolstering the healthcare sector with advanced clinical text analysis.
As NLP continues its trajectory of innovation, it stands poised to redefine the very fabric of human-computer interaction, propelling communication with technology into a realm characterized by profound understanding and nuanced interpretation.
This is 'Sayak Ghorai'. email (gmail) me at "sayak.ghorai21@st.niituniversity.in". We are growing very fast! Are we?

Output:

Both the cases, ‘and’ frequency is maximum and it is a conjunction

```
/usr/local/bin/python3 /Users/sayakghorai/Desktop/Sem_Vi/NLP_TA/Text_Tokenizer.py
sayakghorai@Sayaks-MacBook-Air-M1 NLP_TA % /usr/local/bin/python3
/Users/sayakghorai/Desktop/Sem_Vi/NLP_TA/Text_Tokenizer.py
Without punctuation:
Word  ==»  Freq:
 and  ==»  11
 the  ==»  10
 of  ==»  8
 language  ==»  6
 to  ==»  6
 nlp  ==»  5
 is  ==»  3
 by  ==»  3
 with  ==»  3
 intricate  ==»  2
 advanced  ==»  2
 algorithms  ==»  2
 models  ==»  2
 nuances  ==»  2
 its  ==»  2
 formidable  ==»  2
 challenges  ==»  2
 context  ==»  2
 interaction  ==»  2
 machines  ==»  2
 into  ==»  2
 as  ==»  2
 analysis  ==»  2
 translation  ==»  2
 text  ==»  2
 trajectory  ==»  2
 understanding  ==»  2
 very  ==»  2
 we  ==»  2
 are  ==»  2
 natural  ==»  1
 processing  ==»  1
```

```
represents  ==»  1
intersection  ==»  1
artificial  ==»  1
intelligence  ==»  1
linguistics  ==»  1
deploying  ==»  1
statistical  ==»  1
unravel  ==»  1
subtle  ==»  1
embedded  ==»  1
within  ==»  1
human  ==»  1
overarching  ==»  1
goal  ==»  1
surmount  ==»  1
posed  ==»  1
linguistic  ==»  1
diversity  ==»  1
ambiguity  ==»  1
semantic  ==»  1
intricacies  ==»  1
thereby  ==»  1
fostering  ==»  1
seamless  ==»  1
between  ==»  1
humans  ==»  1
an  ==»  1
ever-evolving  ==»  1
domain  ==»  1
delving  ==»  1
tasks  ==»  1
immense  ==»  1
complexity  ==»  1
such  ==»  1
sentiment  ==»  1
named  ==»  1
entity  ==»  1
recognition  ==»  1
machine  ==»  1
abstractive  ==»  1
summarization  ==»  1
field's  ==»  1
has  ==»  1
been  ==»  1
substantially  ==»  1
influenced  ==»  1
advent  ==»  1
sophisticated  ==»  1
pre-trained  ==»  1
like  ==»  1
bert  ==»  1
gpt  ==»  1
elevating  ==»  1
capacity  ==»  1
not  ==»  1
merely  ==»  1
parse  ==»  1
but  ==»  1
assimilate  ==»  1
contextual  ==»  1
```

```
dependencies  ==»  1
generate  ==»  1
human-like  ==»  1
responses  ==»  1
despite  ==»  1
these  ==»  1
strides  ==»  1
grapples  ==»  1
including  ==»  1
disambiguation  ==»  1
polysemous  ==»  1
words  ==»  1
mitigating  ==»  1
biases  ==»  1
inherent  ==»  1
in  ==»  1
corpora  ==»  1
application  ==»  1
spectrum  ==»  1
extends  ==»  1
across  ==»  1
multifaceted  ==»  1
domains  ==»  1
from  ==»  1
refining  ==»  1
search  ==»  1
engine  ==»  1
augmenting  ==»  1
voice-driven  ==»  1
virtual  ==»  1
assistants  ==»  1
revolutionizing  ==»  1
services  ==»  1
bolstering  ==»  1
healthcare  ==»  1
sector  ==»  1
clinical  ==»  1
continues  ==»  1
innovation  ==»  1
it  ==»  1
stands  ==»  1
poised  ==»  1
redefine  ==»  1
fabric  ==»  1
human-computer  ==»  1
propelling  ==»  1
communication  ==»  1
technology  ==»  1
a  ==»  1
realm  ==»  1
characterized  ==»  1
profound  ==»  1
nuanced  ==»  1
interpretation  ==»  1
this  ==»  1
sayak  ==»  1
ghorai  ==»  1
email  ==»  1
gmail  ==»  1
me  ==»  1
```

```
 at  ==»  1
 sayak.ghorai21@st.niituniversity.in  ==»  1
 growing  ==»  1
 fast  ==»  1

Punctuation ==» Freq:
 ,  ==»   16
 .  ==»   12
 -  ==»   5
 (  ==»   3
 )  ==»   3
 '  ==»   3
 {  ==»   2
 }  ==»   2
 [  ==»   2
 ]  ==»   2
 "  ==»   2
 !  ==»   1
 ?  ==»   1




With punctuation:
Word  ==»  Freq:
 and  ==»  11
 the  ==»  10
 of  ==»  7
 to  ==»  6
 language  ==»  5
 nlp  ==»  4
 is  ==»  3
 by  ==»  3
 with  ==»  3
 advanced  ==»  2
 algorithms  ==»  2
 models  ==»  2
 its  ==»  2
 formidable  ==»  2
 context  ==»  2
 into  ==»  2
 as  ==»  2
 text  ==»  2
 trajectory  ==»  2
 understanding  ==»  2
 very  ==»  2
 are  ==»  2
 natural  ==»  1
 processing  ==»  1
 (nlp)  ==»  1
 represents  ==»  1
 {intricate}  ==»  1
 {intersection}  ==»  1
 [of]  ==»  1
 (artificial)  ==»  1
 [intelligence]  ==»  1
 linguistics,  ==»  1
 deploying  ==»  1
 statistical  ==»  1
 unravel  ==»  1
```

```
subtle  ==»  1
nuances  ==»  1
embedded  ==»  1
within  ==»  1
human  ==»  1
language.  ==»  1
overarching  ==»  1
goal  ==»  1
surmount  ==»  1
challenges  ==»  1
posed  ==»  1
linguistic  ==»  1
diversity,  ==»  1
ambiguity,  ==»  1
semantic  ==»  1
intricacies,  ==»  1
thereby  ==»  1
fostering  ==»  1
seamless  ==»  1
interaction  ==»  1
between  ==»  1
humans  ==»  1
machines.  ==»  1
an  ==»  1
ever-evolving  ==»  1
domain,  ==»  1
delving  ==»  1
tasks  ==»  1
immense  ==»  1
complexity  ==»  1
such  ==»  1
sentiment  ==»  1
analysis,  ==»  1
named  ==»  1
entity  ==»  1
recognition,  ==»  1
machine  ==»  1
translation,  ==»  1
abstractive  ==»  1
summarization.  ==»  1
field's  ==»  1
has  ==»  1
been  ==»  1
substantially  ==»  1
influenced  ==»  1
advent  ==»  1
sophisticated  ==»  1
pre-trained  ==»  1
like  ==»  1
bert  ==»  1
gpt,  ==»  1
elevating  ==»  1
capacity  ==»  1
machines  ==»  1
not  ==»  1
merely  ==»  1
parse  ==»  1
but  ==»  1
assimilate  ==»  1
intricate  ==»  1
```

```
contextual  ==»  1
dependencies  ==»  1
generate  ==»  1
human-like  ==»  1
responses.  ==»  1
despite  ==»  1
these  ==»  1
strides,  ==»  1
grapples  ==»  1
challenges,  ==»  1
including  ==»  1
disambiguation  ==»  1
polysemous  ==»  1
words,  ==»  1
nuances,  ==»  1
mitigating  ==»  1
biases  ==»  1
inherent  ==»  1
in  ==»  1
corpora.  ==»  1
application  ==»  1
spectrum  ==»  1
extends  ==»  1
across  ==»  1
multifaceted  ==»  1
domains,  ==»  1
from  ==»  1
refining  ==»  1
search  ==»  1
engine  ==»  1
augmenting  ==»  1
voice-driven  ==»  1
virtual  ==»  1
assistants  ==»  1
revolutionizing  ==»  1
translation  ==»  1
services  ==»  1
bolstering  ==»  1
healthcare  ==»  1
sector  ==»  1
clinical  ==»  1
analysis.  ==»  1
continues  ==»  1
innovation,  ==»  1
it  ==»  1
stands  ==»  1
poised  ==»  1
redefine  ==»  1
fabric  ==»  1
human-computer  ==»  1
interaction,  ==»  1
propelling  ==»  1
communication  ==»  1
technology  ==»  1
a  ==»  1
realm  ==»  1
characterized  ==»  1
profound  ==»  1
nuanced  ==»  1
interpretation.  ==»  1
```

```
 this  ==»  1
 'sayak  ==»  1
 ghorai'.  ==»  1
 email  ==»  1
 (gmail)  ==»  1
 me  ==»  1
 at  ==»  1
 "sayak.ghorai21@st.niituniversity.in".  ==»  1
 we  ==»  1
 growing  ==»  1
 fast!  ==»  1
 we?  ==»  1

Punctuation ==» Freq:
 ,  ==»   16
 .  ==»   12
 -  ==»   5
 (  ==»   3
 )  ==»   3
 '  ==»   3
 {  ==»   2
 }  ==»   2
 [  ==»   2
 ]  ==»   2
 "  ==»   2
 !  ==»   1
 ?  ==»   1
sayakghorai@Sayaks-MacBook-Air-M1 NLP_TA %
```