```
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')
nltk.download('omw-1.4')
     [nltk data] Downloading package stopwords to /
     [nltk data]
                   Package stopwords is already up-
     [nltk data] Downloading package wordnet to /ro
     [nltk data]
                   Package wordnet is already up-to
     [nltk data] Downloading package punkt to /root
                   Package punkt is already up-to-d
     [nltk data]
     [nltk data] Downloading package omw-1.4 to /ro
     True
from nltk.book import *
     *** Introductory Examples for the NLTK Book **
     Loading text1, ..., text9 and sent1, ..., sent
     Type the name of the text or sentence to view
     Type: 'texts()' or 'sents()' to list the mater
     text1: Moby Dick by Herman Melville 1851
     text2: Sense and Sensibility by Jane Austen 18
     text3: The Book of Genesis
     text4: Inaugural Address Corpus
     text5: Chat Corpus
     text6: Monty Python and the Holy Grail
     text7: Wall Street Journal
     text8: Personals Corpus
     text9: The Man Who Was Thursday by G . K . Che
text1.tokens[:20]
     ['[',
      'Moby',
      'Dick',
      'by',
      'Herman',
      'Melville',
      '1851',
      ']',
      'ETYMOLOGY',
      ٠٠,
      '(',
      'Supplied',
      'by',
      'a',
      'Late',
      'Consumptive',
      'Usher',
```

```
'to',
'a',
'Grammar']
```

The two things I would describe about the tokens method is that it's easy to use for a couple of words in a text and it seperates all the words with commas which makes it very easy to read.

```
text1.concordance('sea', 5, 5)

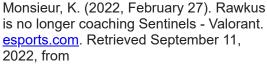
Displaying 5 of 455 matches:
the sea
Indian Sea
the sea
the sea
the sea
```

The count methods act the same as it outputs the number of times the word appears in the text or list.

```
text1.tokens.count('sea')
     433
text1.count('sea')
     433
raw_text = 'To make matters even more interesting, Ra
from nltk.tokenize import word tokenize
tokens = word tokenize(raw text)
tokens[:10]
     ['To',
      'make',
      'matters',
      'even',
      'more',
      'interesting',
      ٠,٠,
      'Rawkus',
```

'says', 'he']



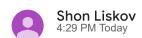


https://www.esports.com/en/rawkus-is-no-longer-coaching-sentinels-329630

```
from nltk.tokenize import sent tokenize
tokens = sent_tokenize(raw_text)
tokens
     ['To make matters even more interesting,
     Rawkus says he hasn't been coaching the team
     for two weeks already.',
      'Unfortunately for the FaZe player turned
     Sentinels coach, the only tournament that
     will go on his resume is VCT Champions, where
     Sentinels was eliminated in the group stage
     after losses to FURIA and KRÜ Esports.',
      'The ex-coach also detailed the deal FaZe
     and Sentinels had in place for him.',
      'Both organizations worked out a deal after
     FaZe was eliminated from the VCT Last Chance
     Qualifier.',
      'Sentinels essentially borrowed Rawkus'
     services while he was still contracted to
     FaZe, up until January 1st.']
from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
tokens = word_tokenize(raw_text)
for·w·in·tokens:
..print(stemmer.stem(w))
     rawku
     say
     he
     hasn
     t
     been
     coach
     the
     team
     for
     two
     week
     alreadi
     unfortun
     for
     the
     faze
     player
     turn
     sentinel
     coach
     the
```

onli

```
tournament
that
will
go
on
hi
resum
is
vct
champion
where
sentinel
wa
elimin
in
the
group
stage
after
loss
to
furia
and
krü
esport
the
ex-coach
also
detail
the
```



The NLTK library is very useful and has many packages to choose from for many NLP projects. The code quality of NLTK is very slick and simple to read and write. Only takes about three to five lines to output the result you wish for. Chat bots are perfect examples for the use of NLTK and utilizing the works of how NLTK can generate

different outputs for user input.

interest - interesting rawku - rawkus elimin - eliminated wa - wa week - week

```
from nltk.stem import WordNetLemmatizer
lemma = WordNetLemmatizer()
words = nltk.word_tokenize(raw_text)
for w in words:
  print(lemma.lemmatize(w))
     Sentineis
     coach
     the
     only
     tournament
     that
     will
     go
     on
     his
     resume
```

VCT Champions where Sentinels eliminated in the group stage after loss to **FURIA** and KRÜ Esports The ex-coach also detailed the deal FaZe and Sentinels had in place for him Both organization worked out а deal after FaZe wa eliminated from +ha

Colab paid products - Cancel contracts here

✓ 0s completed at 4:35 PM

×