

```
!pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.6)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.3.2)
Requirement already satisfied: regex<=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2023.6.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.66.1)
```

```
import nltk
```

```
nltk.download()
```



NLTK Downloader

```
-----
d) Download  l) List    u) Update  c) Config  h) Help   q) Quit
-----
```

Downloader> d

Download which package (l=list; x=cancel)?

Identifier> punkt

Downloading package punkt to /root/nltk_data...

Unzipping tokenizers/punkt.zip.

```
-----
d) Download  l) List    u) Update  c) Config  h) Help   q) Quit
-----
```

Downloader> q

True

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

```
text='''Stephenson 2-18 is now known as being one of the largest, if not the current largest star ever discovered, surpassing other stars like VY Canis
        Stephenson 2-18 has a radius of 2,150 solar radii, being larger than almost the entire orbit of Saturn (1,940 - 2,169 solar radii).'''
```

```
text
```

```
'Stephenson 2-18 is now known as being one of the largest, if not the current largest star ever discovered, surpassing other stars like VY Canis
ing larger than almost the entire orbit of Saturn (1,940 - 2,169 solar radii).'
```

```
sentences = sent_tokenize(text)
```

```
sentences
```

```
['Stephenson 2-18 is now known as being one of the largest, if not the current largest star ever discovered, surpassing other stars like VY Canis
Majoris and UY Scuti.',
 'Stephenson 2-18 has a radius of 2,150 solar radii, being larger than almost the entire orbit of Saturn (1,940 - 2,169 solar radii).']
```

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

```
words = word_tokenize(text)
```

```
words
```

```
['Stephenson',
 '2-18',
 'is',
 'now',
 'known',
 'as',
 'being',
 'one',
 'of',
 'the',
 'largest',
 ',',
 ',',
 'if',
 'not',
 'the',
 'current',
 'largest',
 'star',
 'ever',
 'discovered',
 ',',
 ',',
 'surpassing',
 'other',
 'stars',
 'like',
 'VY',
```

```
'Canis',
'Majoris',
'and',
'UY',
'Scuti',
'.',
'Stephenson',
'2-18',
'has',
'a',
'radius',
'of',
'2,150',
'solar',
'radii',
',',
'being',
'larger',
'than',
'almost',
'the',
'entire',
'orbit',
'of',
'Saturn',
'(',
'1,940',
'-',
'2,169',
'solar',
'radii',
')',
```

```
for w in words:
    print (w)
```

```
Stephenson
2-18
is
now
known
as
being
one
of
the
largest
,
if
not
the
current
largest
star
ever
discovered
,
surpassing
other
stars
like
VY
Canis
Majoris
and
UY
Scuti
.
Stephenson
2-18
has
a
radius
of
2,150
solar
radii
,
being
larger
than
almost
the
entire
orbit
of
Saturn
(
1,940
-
2,169
solar
radii
)
```

```
sent_tokenize(text)
```

```
['Stephenson 2-18 is now known as being one of the largest, if not the current largest star ever discovered, surpassing other stars like VY Canis Majoris and UY Scuti.',  
'Stephenson 2-18 has a radius of 2,150 solar radii, being larger than almost the entire orbit of Saturn (1,940 - 2,169 solar radii).']
```

```
[word_tokenize(text) for t in sent_tokenize(text)]
```

```
[['Stephenson',  
'2-18',  
'is',  
'now',  
'known',  
'as',  
'being',  
'one',  
'of',  
'the',  
'largest',  
,,  
'if',  
'not',  
'the',  
'current',  
'largest',  
'star',  
'ever',  
'discovered',  
,,  
'surpassing',  
'other',  
'stars',  
'like',  
'VY',  
'Canis',  
'Majoris',  
'and',  
'UY',  
'Scuti',  
,,  
'Stephenson',  
'2-18',  
'has',  
'a',  
'radius',  
'of',  
'2,150',  
'solar',  
'radii',  
,,  
'being',  
'larger',  
'than',  
'almost',  
'the',  
'entire',  
'orbit',  
'of',  
'Saturn',  
'(',  
'1,940',  
'-',  
'2,169',  
'solar',  
'radii',  
)'],
```

```
from nltk.tokenize import wordpunct_tokenize
```

```
wordpunct_tokenize(text)
```

```
['Stephenson',  
'2',  
'-',  
'18',  
'is',  
'now',  
'known',  
'as',  
'being',  
'one',  
'of',  
'the',  
'largest',  
,,  
'if',  
'not',  
'the',  
'current',  
'largest',  
'star',  
'ever',  
'discovered',
```

```
,',  
'surpassing',  
'other',  
'stars',  
'like',  
'VY',  
'Canis',  
'Majoris',  
'and',  
'UY',  
'Scuti',  
,',  
'Stephenson',  
'2',  
'-',  
'18',  
'has',  
'a',  
'radius',  
'of',  
'2',  
,',  
,',  
'150',  
'solar',  
'radii',  
,',  
'being',  
'larger',  
'than',  
'almost',  
'the',  
'entire',  
'orbit',  
'of',  
'Saturn',  
'(',
```

```
text.lower()
```

```
"stephenson 2-18 is now known as being one of the largest, if not the current largest star ever discovered, surpassing other stars like vy canis majoris. it is 150 times larger than almost the entire orbit of saturn (1,940 - 2,169 solar radii)."
```

```
text.upper()
```

```
"STEPHENSON 2-18 IS NOW KNOWN AS BEING ONE OF THE LARGEST, IF NOT THE CURRENT LARGEST STAR EVER DISCOVERED, SURPASSING OTHER STARS LIKE VY CANIS MAJORIS. IT IS 150 TIMES LARGER THAN ALMOST THE ENTIRE ORBIT OF SATURN (1,940 - 2,169 SOLAR RADII)."
```