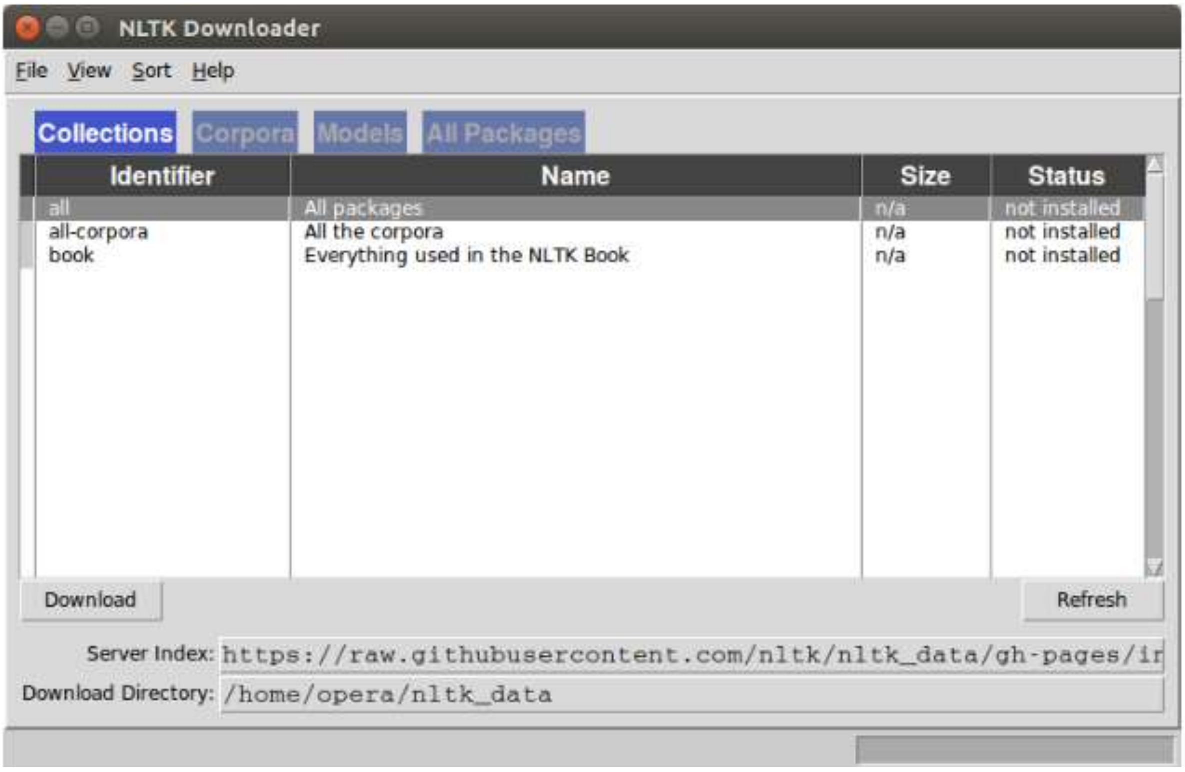
*****The following code will help you to tokenize sentences*****

```
import nltk
nltk.download()


from nltk.tokenize import sent_tokenize, word_tokenize

data = "All work and no play makes jack a dull boy, all work and no play"
print(word_tokenize(data))
```



Identifier	Name	Size	Status
all	All packages	n/a	not installed
all-corpora	All the corpora	n/a	not installed
book	Everything used in the NLTK Book	n/a	not installed

Server Index: https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/ir

Download Directory: /home/opera/nltk_data

Click all and then click download. It will download all the required packages which may take a while, the bar on the bottom shows the progress.

A sentence or data can be split into words using the method **word_tokenize():**

```
from nltk.tokenize import sent_tokenize, word_tokenize

data = "All work and no play makes jack a dull boy, all work and no play"
print(word_tokenize(data))
```

Tokenize words

A sentence or data can be split into words using the method word_tokenize():

```
from nltk.tokenize import sent_tokenize, word_tokenize

data = "All work and no play makes jack a dull boy, all work and no play"
print(word_tokenize(data))
```

This will output:

```
['All', 'work', 'and', 'no', 'play', 'makes', 'jack', 'dull', 'boy', ',', 'all', 'work', 'and', 'no', 'play']
```

All of them are words except the comma. Special characters are treated as separate tokens.

Tokenizing sentences

The same principle can be applied to sentences. Simply change the to sent_tokenize()
We have added two sentences to the variable data:

```
from nltk.tokenize import sent_tokenize, word_tokenize

data = "All work and no play makes jack dull boy. All work and no play makes jack a dull boy."
print(sent_tokenize(data))
```

Outputs:

```
['All work and no play makes jack dull boy.', 'All work and no play makes jack a dull boy.']
```

NLTK and arrays

If you wish to you can store the words and sentences in arrays:

```
from nltk.tokenize import sent_tokenize, word_tokenize

data = "All work and no play makes jack dull boy. All work and no play makes jack a dull boy."

phrases = sent_tokenize(data)
words = word_tokenize(data)

print(phrases)
print(words)
```