

Updated for Anaconda 5.0 on November 13th, 2017

Installation of NLTK for NLP Day

Expected Installation Time : 30 minutes, depending on internet connection.

Contents

Method 1: Command Prompt / Terminal	2
Step 1: Open Anaconda Prompt / Terminal.....	2
Step 2: Import NLTK	2
DO NOT CLOSE THE TERMINAL / COMMAND PROMPT	3
Step 3: Download all-corpora	3
Step 4: Download the Punkt from All Packages.....	4
Step 5: Test the Package	4
Method 2: Anaconda Navigator	5
Step 1: Navigate to Environments	5
Step 2: Open a New Notebook	5
Step 3: Download all-corpora	6
Step 4: Download the Punkt from All Packages.....	7
Step 5: Test the Package	7

Method 1: Command Prompt / Terminal

Step 1: Open Anaconda Prompt / Terminal

conda install -c anaconda nltk

```
Anaconda Prompt - python

(C:\Users\ktbernoulli\Anaconda3) C:\Users\ktbernoulli>conda install -c anaconda nltk
Fetching package metadata .....
Solving package specifications: .

Package plan for installation in environment C:\Users\ktbernoulli\Anaconda3:
```

```
administrator — python — 104x29

Last login: Mon Nov 13 12:32:28 on ttys000
[The-Most-Extreme-Thing-Ive-Done-Under-300-Bucks:~ ktbernoulli$ conda install -c anaconda nltk
Fetching package metadata .....
Solving package specifications: .

Package plan for installation in environment /Applications/anaconda:
```

Proceed ([y]/n)? y

Proceed ([y]/n)? y

Step 2: Import NLTK

Then nltk.download()

```
(C:\Users\ktbernoulli\Anaconda3) C:\Users\ktbernoulli>python
Python 3.6.3 |Anaconda custom (64-bit)| (default, Oct 15 2017, 03:27:45) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

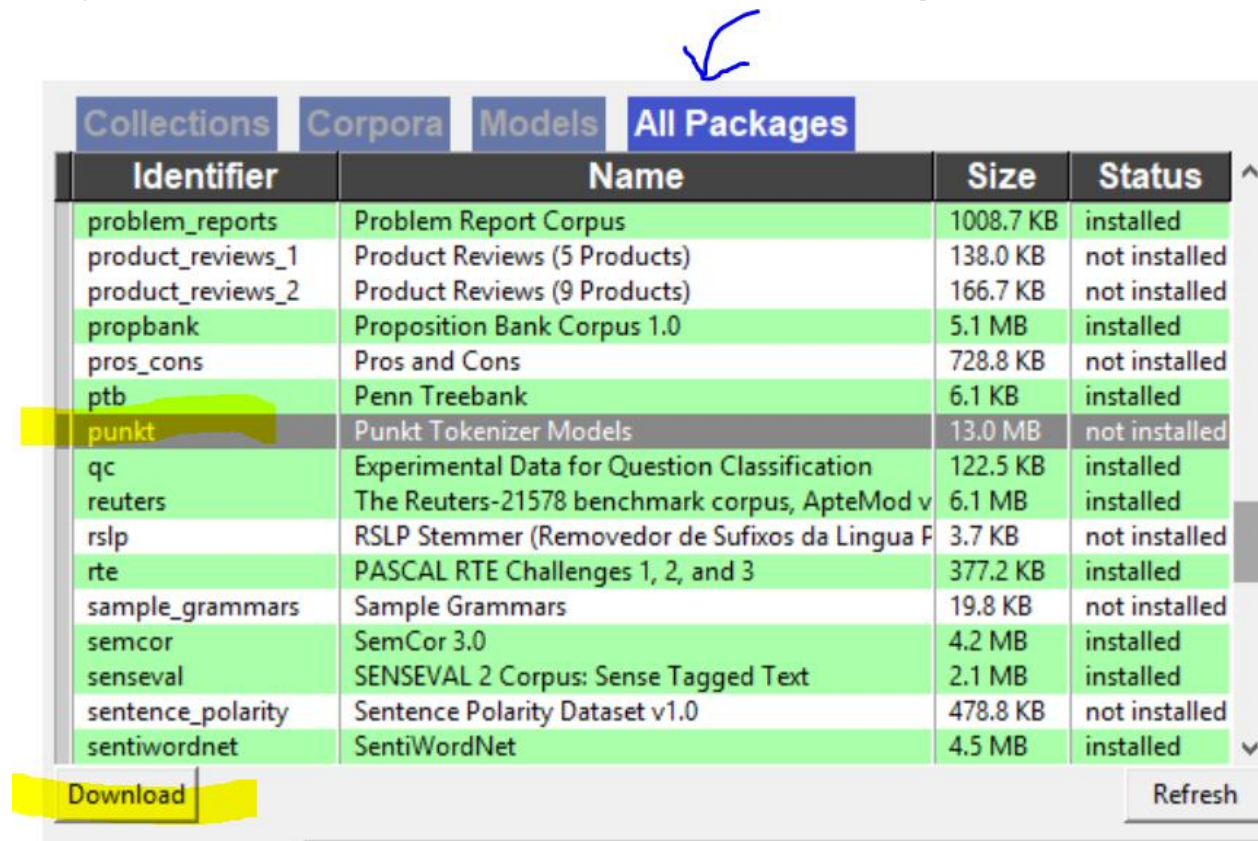
```
The-Most-Extreme-Thing-Ive-Done-Under-300-Bucks:~ ktbernoulli$ python
Python 3.6.1 |Anaconda custom (x86_64)| (default, May 11 2017, 13:04:09)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

DO NOT CLOSE THE TERMINAL / COMMAND PROMPT.

Step 3: Download all-corpora

Collections				
Corpora				
Models				
All Packages				
Identifier	Name		Size	Status
all	All packages		n/a	partial
all-corpora	All the corpora		n/a	partial
all-nltk	All packages available on nltk_data gh-pages bran		n/a	partial
book	Everything used in the NLTK Book		n/a	partial
popular	Popular packages		n/a	partial
third-party	Third-party data packages		n/a	installed

Step 4: Download the Punkt from All Packages



Collections	Corpora	Models	All Packages
Identifier	Name	Size	Status
problem_reports	Problem Report Corpus	1008.7 KB	installed
product_reviews_1	Product Reviews (5 Products)	138.0 KB	not installed
product_reviews_2	Product Reviews (9 Products)	166.7 KB	not installed
probank	Proposition Bank Corpus 1.0	5.1 MB	installed
pros_cons	Pros and Cons	728.8 KB	not installed
ptb	Penn Treebank	6.1 KB	installed
punkt	Punkt Tokenizer Models	13.0 MB	not installed
qc	Experimental Data for Question Classification	122.5 KB	installed
reuters	The Reuters-21578 benchmark corpus, ApteMod v	6.1 MB	installed
rsdp	RSLP Stemmer (Removedor de Sufixos da Lingua P	3.7 KB	not installed
rte	PASCAL RTE Challenges 1, 2, and 3	377.2 KB	installed
sample_grammars	Sample Grammars	19.8 KB	not installed
semcor	SemCor 3.0	4.2 MB	installed
senseval	SENSEVAL 2 Corpus: Sense Tagged Text	2.1 MB	installed
sentence_polarity	Sentence Polarity Dataset v1.0	478.8 KB	not installed
sentiwordnet	SentiWordNet	4.5 MB	installed

Download Refresh

Step 5: Test the Package

```
(C:\Users\ktbernoulli\Anaconda3) C:\Users\ktbernoulli>python
Python 3.6.3 |Anaconda custom (64-bit)| (default, Oct 15 2017, 03:27:45) [MSC v.1900
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
True
>>> from nltk.tokenize import word_tokenize
>>> word_tokenize("Hello World")
['Hello', 'World']
>>>
```

Method 2: Anaconda Navigator

Step 1: Navigate to Environments

The screenshot shows the Anaconda Navigator interface. At the top, there is a search bar with 'nltk' entered. Below it, a table lists the installed packages. The 'nltk' package is listed with a status of 'Not installed'. A dialog box titled 'Install Packages' is open, showing a table with one row for 'nltk' and buttons for 'Cancel' and 'Apply'.

Name	Description
nltk	Build python programs to work with human language data

1 package will be installed

	Name	Unlink	Link	Channel
1	nltk	-	3.2.4	defaults

Cancel **Apply**


Step 2: Open a New Notebook

```
In [*]: import nltk
        nltk.download()
```

Step 3: Download all-corpora

<div>CollectionsCorporaModelsAll Packages</div>			
Identifier	Name	Size	Status
all	All packages	n/a	partial
all-corpora	All the corpora	n/a	partial
all-nltk	All packages available on nltk_data gh-pages bran	n/a	partial
book	Everything used in the NLTK Book	n/a	partial
popular	Popular packages	n/a	partial
third-party	Third-party data packages	n/a	installed
<div>CancelRefresh</div>			

Step 4: Download the Punkt from All Packages



Collections	Corpora	Models	All Packages
Identifier	Name	Size	Status
problem_reports	Problem Report Corpus	1008.7 KB	installed
product_reviews_1	Product Reviews (5 Products)	138.0 KB	not installed
product_reviews_2	Product Reviews (9 Products)	166.7 KB	not installed
proppbank	Proposition Bank Corpus 1.0	5.1 MB	installed
pros_cons	Pros and Cons	728.8 KB	not installed
ptb	Penn Treebank	6.1 KB	installed
punkt	Punkt Tokenizer Models	13.0 MB	not installed
qc	Experimental Data for Question Classification	122.5 KB	installed
reuters	The Reuters-21578 benchmark corpus, ApteMod v	6.1 MB	installed
rsdp	RSLP Stemmer (Removedor de Sufixos da Lingua P	3.7 KB	not installed
rte	PASCAL RTE Challenges 1, 2, and 3	377.2 KB	installed
sample_grammars	Sample Grammars	19.8 KB	not installed
semcor	SemCor 3.0	4.2 MB	installed
senseval	SENSEVAL 2 Corpus: Sense Tagged Text	2.1 MB	installed
sentence_polarity	Sentence Polarity Dataset v1.0	478.8 KB	not installed
sentiwordnet	SentiWordNet	4.5 MB	installed

Download Refresh

Step 5: Test the Package

```
In [1]: import nltk

In [3]: from nltk.tokenize import word_tokenize

In [4]: word_tokenize("Hello World")

['Hello', 'World']
```