

NLP Lab - Installing the data for NLTK Toolkit

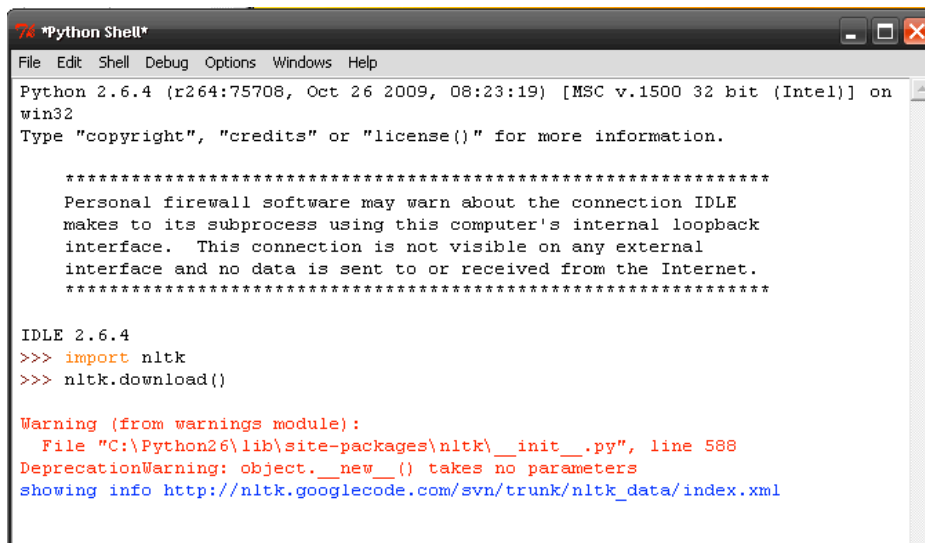
1. Get the data

Open Python 2.6 IDLE (Python GUI) by going to All Programs->Python 2.6 -> IDLE (Python GUI)

Now in the Python interpreter window, we'll run the NLTK download function. First, type the first command shown after the >>> prompt. After it is done, you'll get another >>> prompt and type the second command.

```
>>> import nltk
>>> nltk.download()
```

Ignore the warning and look for the NLTK Downloader window, which may be behind others.



```
Python Shell
File Edit Shell Debug Options Windows Help
Python 2.6.4 (r264:75708, Oct 26 2009, 08:23:19) [MSC v.1500 32 bit (Intel)] on
win32
Type "copyright", "credits" or "license()" for more information.

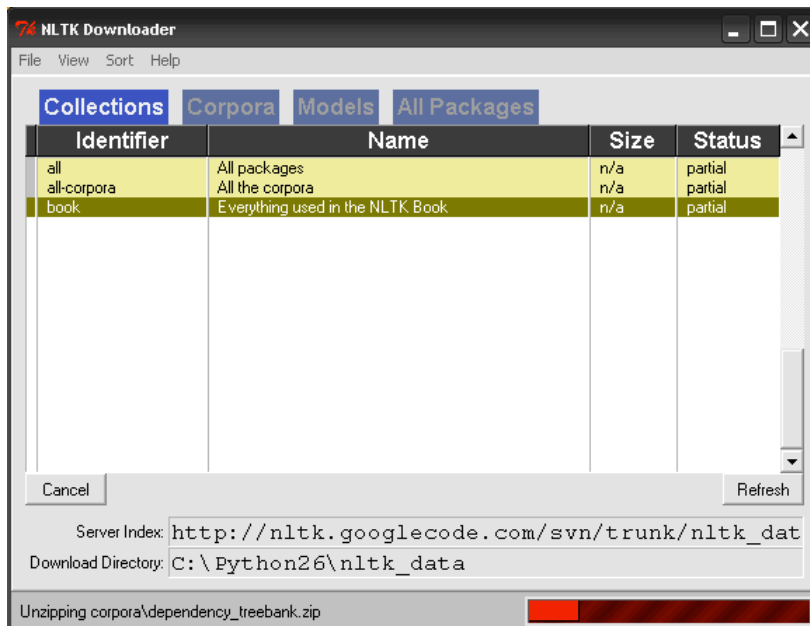
*****
Personal firewall software may warn about the connection IDLE
makes to its subprocess using this computer's internal loopback
interface. This connection is not visible on any external
interface and no data is sent to or received from the Internet.
*****

IDLE 2.6.4
>>> import nltk
>>> nltk.download()

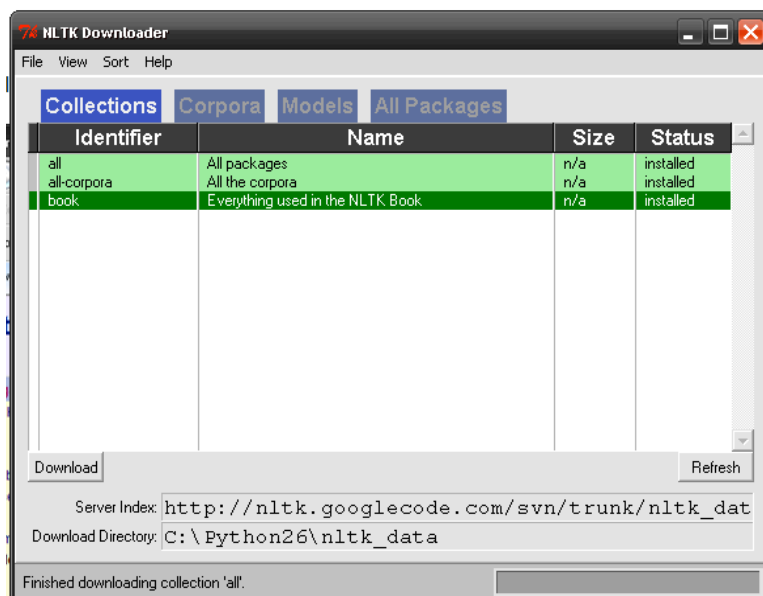
Warning (from warnings module):
  File "C:\Python26\lib\site-packages\nltk\__init__.py", line 588
  DeprecationWarning: object.__new__() takes no parameters
  showing info http://nltk.googlecode.com/svn/trunk/nltk_data/index.xml
```

In the NLTK Downloader window, look for the Download Directory textbox near the bottom and **change the Download Directory to be H:\nltk_data**

Select “book” is to get the packages needed for the NLTK book examples and click the download button under the packages pane on the left. This will take a long time.



When it's finished downloading, you can close that window.



Now go back to the Python IDLE window and test that the data has been installed as follows. (This uses the downloaded Brown Corpus): Type the command shown after the >>> prompts.

```
>>> from nltk.corpus import brown
>>> brown.words()
['The', 'Fulton', 'County', 'Grand', 'Jury', 'said', ...]
```