

NLTK – Natural Language Toolkit (<http://www.nltk.org/Home>)

1. **If you want to use NLTK in the Unix environment from your CS account, you need to set up the environment-**

Add this line in the \$HOME/.profile file in your CS account –

```
export PYTHONPATH=/home/cs4705/nltk-2.0b5
```

Then, either restart the shell or do

```
knl2102@helsinki /home/cs4705 $ . $HOME/.profile
```

2. You can also use NLTK by downloading it to Windows from the following web site:

3. WHAT IS NLTK -

**Natural Language Toolkit (NLTK)**

NLTK was originally created in 2001 as part of a computational linguistics course in the Department of Computer and Information Science at the University of Pennsylvania. Since then it has been developed and expanded with the help of dozens of contributors. It has now been adopted in courses in dozens of universities, and serves as the basis of many research projects.

4. What we'll be using NLTK for –  
We'll be using nltk to write custom CFGs for the set of given sentences and produce parse trees for those sentences to check the accuracy of the grammar.
5. A simple example of defining a CFG and using it to parse a sentence using Recursive Descent Parser-

```
knl2102@helsinki /home/cs4705 $ python
Python 2.5.2 (r252:60911, Jul 22 2009, 15:35:03)
[GCC 4.2.4 (Ubuntu 4.2.4-1ubuntu3)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk;
>>> grammar1 = nltk.parse_cfg("""
... S -> NP VP
... VP -> V NP | V NP PP
... PP -> P NP
... V -> "saw" | "ate" | "walked"
... NP -> "John" | "Mary" | "Bob" | Det N | Det N PP
... Det -> "a" | "an" | "the" | "my"
... N -> "man" | "dog" | "cat" | "telescope" | "park"
```

```
... P -> "in" | "on" | "by" | "with"
... """")

>>> sent = ("Bob saw Mary").split()
>>> rd_parser = nltk.RecursiveDescentParser(grammar1)
>>> for tree in rd_parser.nbest_parse(sent):
...     print tree
...
(S (NP Bob) (VP (V saw) (NP Mary)))
>>>
```

(More examples and complete documentation is available at  
<http://nltk.googlecode.com/svn/trunk/doc/book/ch08.html>)

6. You should use the interactive chart parser for developing the grammar. Use the following commands from Python to invoke it:

```
import nltk
nltk.app.chartparser()
```

You can then define your grammar in a file and load it via the GUI. You would then enter a sentence and test it in the GUI. This is the easiest way to develop your grammar for the assignment.