Computing in 571
Programming

- For standalone code, you can use anything you like
  - That runs on the department cluster

- For some exercises, we will use a Python-based toolikt
Department Cluster

- Resources on CLMA wiki
  - http://depts.washington.edu/uwcl/twiki/bin/view.cgi/Main

- Installed corpora, software, etc.
  - patas.ling.washington.edu
  - dryas.ling.washington.edu
Condor

- Distributes software processes to cluster nodes
- All homework will be tested with condor_submit
  - See documentation on CLMA wiki
    - Construction of condor scripts
Natural Language Toolkit (NLTK)

- Large, integrated, fairly comprehensive
  - Stemmers
  - Taggers
  - Parsers
  - Semantic analysis
  - Corpus samples, etc
- Extensively documented
- Pedagogically oriented
  - Implementations strive for clarity
    - Sometimes at the expense of speed/efficiency
NLTK Information

- [http://www.nltk.org](http://www.nltk.org)
  - Online book
- Demos of software
- HOWTOs for specific components
- API information, etc
Python & NLTK

- NLTK is installed on cluster
  - Use python2.6 with NLTK

- NLTK data is also installed
  - /corpora/nltk/nltk-data

- NLTK is written in Python
  - http://www.python.org; http://docs.python.org
    - Many good online intros, fairly simple
Python & NLTK

- Interactive mode allows experimentation, introspection
  - patas$ python2.6
  - >>> import nltk
  - >>> dir(nltk)
  - ..... AbstractLazySequence', 'AffixTagger', 'AnnotationTask',
        'Assignment', 'BigramAssocMeasures', 'BigramCollocationFinder',
        'BigramTagger', 'BinaryMaxentFeatureEncoding',
  - >>>> help(AffixTagger)
  - ..... 
        - Prints properties, methods, comments,...
Turning in Homework

- Class CollectIt
  - Linked from course webpage

- Homeworks due Tuesday night
  - CollectIt time = Wednesday (00:01)

- Should submit as hw#.tar
  - Where # = homework number
  - Tar file contains top-level condor scripts to run
Programming with NLTK