

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 toolkit

Department Cluster

- Resources on CLMS wiki
 - <http://depts.washington.edu/uwcl>
 - 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 CLMS wiki
 - Construction of condor scripts

NLTK

- 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>
 - 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 Thursday night
 - CollectIt time = Friday (00:01)
- Should submit as hw#.tar
 - Where # = homework number
 - Tar file contains top-level condor scripts to run

Programming with NLTK

NLTK is a suite of programs and libraries for处理自然语言。NLTK 是一个处理自然语言的套件，包含了许多用于处理文本的程序和库。

NLTK 提供了多种语言模型、语义分析、句法分析、词性标注、命名实体识别等工具，可以帮助开发者快速地构建自然语言处理系统。

NLTK 的主要特点包括：

- 易于使用：提供了直观的 API 和丰富的示例代码，方便开发者上手。

- 功能强大：支持多种自然语言处理任务，如文本分类、情感分析、机器翻译等。

- 开源免费：完全开源，任何人都可以自由使用、修改和分发。

- 社区活跃：拥有一个活跃的开发者社区，定期发布更新和维护。