Computing in Ling571

Scott Farrar
CLMA, University of Washington
farrar@uw.edu

January 3, 2010

Today's lecture

1. Languages
   - Resources
   - Python cmds, etc
2. The NLTK
   - Intro
   - Using the NLTK
3. Class Discussion and Assignments
   - GoPost
   - Patas
   - Condor
   - Assignments
   - CollectIt

Choice of Programming Language

You can use any language you like for assignments, well sort of.

Some require the NLTK and certain Java packages.

Choice of Programming Language

You can use any language you like for assignments, well sort of.

Some require the NLTK and certain Java packages.

Focus on Python

For the lecture, I'll focus on Python because the first assignment requires it.

You should get used to using the NLTK as well. More later on the NLTK.

Main resources

- Python hub: http://www.python.org. start here
- Python 2.6 Docs: http://docs.python.org/
- A general Python code repository, including NLP code: http://www.vex.net/parnassus/

Main resources

- Python hub: http://www.python.org. start here
- Python 2.6 Docs: http://docs.python.org/
- A general Python code repository, including NLP code: http://www.vex.net/parnassus/

Main resources

- Python hub: http://www.python.org. start here
- Python 2.6 Docs: http://docs.python.org/
- A general Python code repository, including NLP code: http://www.vex.net/parnassus/

Main resources

- Python hub: http://www.python.org. start here
- Python 2.6 Docs: http://docs.python.org/
- A general Python code repository, including NLP code: http://www.vex.net/parnassus/

Books

- Lutz & Ascher *Learning Python* O'Reilly. **(beginner)**
- Chun *Core Python Programming* Prentice Hall. **(beginner/intermediate)**
- Martelli *Python in a Nutshell* O'Reilly. **(beginner/intermediate)**
- Beazley *Python Essential Reference* Developer's Library, 4th edition. **(intermediate)**
- Goldwasser & Letscher *Object-Oriented Programming in Python* . **(advanced)**

See this list for other (and multilingual) recommendations: http://wiki.python.org/moin/PythonBooks

Python tutorials

- For the simplest kind of tutorial for **non-programmers**, see http://programming-crash-course.com/
- A nice thorough intro is the on-line book, *How to Think Like a Computer Scientist: Learning with Python*, http://www.ibiblio.org/obp/thinkCSpy/
- For **linguists**, you might start first with the Natural Language Took Kit: http://www.nltk.org.
- **For programmers**, see the tutorials that already assume prior CS knowledge: http://docs.python.org/tutorial/index.html http://www.diveintopython.org/.

Scenario for Ling571

What's the best way to tinker with the code, and to do an assignment?

- Have the API documentation open.
- Do a Web search to solve specific issues.
- Work through a tutorial or book.
- Test short code snippets using 'interactive' mode.
- For longer programs, use your favorite editor to create **.py** files and execute the interpreter over that file.

Scenario for Ling571

What's the best way to tinker with the code, and to do an assignment?

- Have the API documentation open.
- Do a Web search to solve specific issues.
- Work through a tutorial or book.
- Test short code snippets using 'interactive' mode.
- For longer programs, use your favorite editor to create **.py** files and execute the interpreter over that file.

Scenario for Ling571

What's the best way to tinker with the code, and to do an assignment?

- Have the API documentation open.
- Do a Web search to solve specific issues.
- Work through a tutorial or book.
- Test short code snippets using 'interactive' mode.
- For longer programs, use your favorite editor to create **.py** files and execute the interpreter over that file.

Scenario for Ling571

What's the best way to tinker with the code, and to do an assignment?

- Have the API documentation open.
- Do a Web search to solve specific issues.
- Work through a tutorial or book.
- Test short code snippets using 'interactive' mode.
- For longer programs, use your favorite editor to create **.py** files and execute the interpreter over that file.

Scenario for Ling571

What's the best way to tinker with the code, and to do an assignment?

- Have the API documentation open.
- Do a Web search to solve specific issues.
- Work through a tutorial or book.
- Test short code snippets using 'interactive' mode.
- For longer programs, use your favorite editor to create **.py** files and execute the interpreter over that file.

Handy template

sample.py

```
#write your test code here
def myfunc(x):
    print  x*2

if __name__=='__main__':
    #call test code from here
    myfunc('moin')
```

Python in Vim and Emacs

Program Vim or Emacs to execute the code. For example for Vim, add this to **.vimrc** on Patas:

```
map <f2> :w\|!python2.6 %<cr>
```

For Emacs, see http://www.emacswiki.org/emacs/PythonMode

Essential Python cmds

dir(...) Returns a list of names comprising the attributes of a given object:

```
>>> dir(nltk.chat)
['Chat', '__builtins__', '__doc__', '__file__',
 '__name__', '__path__',
'demo', 'eliza', 'eliza_chat', 'iesha',
'iesha_chat', 'random', 're', 'reflections',
'rude', 'rude_chat', 'string', 'suntsu',
'suntsu_chat', 'util', 'zen', 'zen_chat']
```

Essential Python cmds

help(...): a built function to show documentation for some object:

```
>>help(nltk.chat.rude)

Help on module nltk.chat.rude in nltk.chat:

NAME
    nltk.chat.rude

FILE
    /usr/lib/python2.5/site-packages/nltk/chat/rude.py

DESCRIPTION
    # Natural Language Toolkit: Zen Chatbot
    #
    # Copyright (C) 2001-2008 NLTK Project
    # Author: Peter Spiller <pspiller@csse.unimelb.edu.au>
    # URL: <http://www.nltk.org/>
    # For license information, see LICENSE.TXT
```

Python on Patas

- Many versions of Python on Patas
- Be sure to use **$ python2.6** to execute your code
- By default, **$ python** runs Python2.5
- Python versions are installed in **/opt**

Python on Patas

- Many versions of Python on Patas
- Be sure to use **$ python2.6** to execute your code
- By default, **$ python** runs Python2.5
- Python versions are installed in **/opt**

Python on Patas

- Many versions of Python on Patas
- Be sure to use **$ python2.6** to execute your code
- By default, **$ python** runs Python2.5
- Python versions are installed in **/opt**

Python on Patas

- Many versions of Python on Patas
- Be sure to use **$ python2.6** to execute your code
- By default, **$ python** runs Python2.5
- Python versions are installed in **/opt**

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

Natural Language Tool Kit (www.nltk.org)

- The NLTK is a bundle of NLP modules, mostly designed for learning computational linguistics:
  - parsers
  - taggers
  - corpus readers
  - evaluation modules
  - semantic processors
  - chatbots
- The NLTK comes with copious documentation, demos, tutorials and data.
- It's *all* integrated (demos, code and data).
- The most complete NLP package ever constructed.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK Documentation maze

- Where to find NLTK documentation can be a little confusing.
- Only use links from www.nltk.org (some old stuff lying around in other sites)
- Go to: **http://www.nltk.org/documentation**
- The NLTK Book: use this for specific reading assignments (linked from the on-line course schedule).
- API Doc: to see source code and the API
- HOWTOs: go here when the Book and API aren't enough (can be out of date)
- All this material is linked from the course website.

NLTK on Patas

- The NLTK is usable with your Patas acct
- test this with:

```
farrar@patas:~$ python2.5
Python 2.5.2 (r252:60911, Dec  4 2008, 14:21:22)
Type "help", "copyright", "credits" or "license"
for more information.
>>> import nltk
>>>
```

- NLTK data lives at **/corpora/nltk/nltk-data**.
- **cp** specific data sets to your home to inspect, or import with the NLTK and use API.

NLTK on Patas

- The NLTK is usable with your Patas acct
- test this with:

```
farrar@patas:~$ python2.5
Python 2.5.2 (r252:60911, Dec  4 2008, 14:21:22)
Type "help", "copyright", "credits" or "license"
for more information.
>>> import nltk
>>>
```

- NLTK data lives at **/corpora/nltk/nltk-data**.
- **cp** specific data sets to your home to inspect, or import with the NLTK and use API.

GoPost discussion

- Please use GoPost for after-hours class discussions. Steve or I will try our best to get back to within a day. (Don't expect us to spend our evenings stalking you on GoPost.)
- Send specific questions about your grade to Scott.
- Use GoPost for: coding issues, specific NLP questions, homework questions.
- Post useful code to GoPost (not your whole assignment, but parts are fine). A good example is code that gets around a bug in the NLTK.

GoPost discussion

- Please use GoPost for after-hours class discussions. Steve or I will try our best to get back to within a day. (Don't expect us to spend our evenings stalking you on GoPost.)
- Send specific questions about your grade to Scott.
- Use GoPost for: coding issues, specific NLP questions, homework questions.
- Post useful code to GoPost (not your whole assignment, but parts are fine). A good example is code that gets around a bug in the NLTK.

GoPost discussion

- Please use GoPost for after-hours class discussions. Steve or I will try our best to get back to within a day. (Don't expect us to spend our evenings stalking you on GoPost.)
- Send specific questions about your grade to Scott.
- Use GoPost for: coding issues, specific NLP questions, homework questions.
- Post useful code to GoPost (not your whole assignment, but parts are fine). A good example is code that gets around a bug in the NLTK.

GoPost discussion

- Please use GoPost for after-hours class discussions. Steve or I will try our best to get back to within a day. (Don't expect us to spend our evenings stalking you on GoPost.)
- Send specific questions about your grade to Scott.
- Use GoPost for: coding issues, specific NLP questions, homework questions.
- Post useful code to GoPost (not your whole assignment, but parts are fine). A good example is code that gets around a bug in the NLTK.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Using Patas

- Patas is the CLMA computing cluster.
- patas.ling.washington.edu
- See the CLMA Wiki for more info.
- All hw code must run on Patas.
- For assignments, all input and sample files are found on the website.
- Sample code from lectures will be posted here as well.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

Homeworks and Condor

- Condor is a system that optimizes cluster resources.
- You are asked to submit a Condor script with each assignment.
- See the CLMA wiki pages for help on this.
- We will run your assignments using *condor_submit*.
- Everyones code will be run the same way.
- The condor script allows individual flexibility as to the language, arguments and structure of your code.

How to organize homeworks

- Create a top-level folder called **hw1** to contain your work. At the top level in that folder, create a Condor script called **hw1.cmd**. Include all other files within that directory, or in a subdirectory of your choosing.
- For more involved assignments, this will give you flexibility in design while allowing us to run everyones code in the same way, that is, by issuing a single command, **condor_submit hw1.cmd**.

How to organize homeworks

- Create a top-level folder called **hw1** to contain your work. At the top level in that folder, create a Condor script called **hw1.cmd**. Include all other files within that directory, or in a subdirectory of your choosing.
- For more involved assignments, this will give you flexibility in design while allowing us to run everyones code in the same way, that is, by issuing a single command, **condor_submit hw1.cmd**.

How to turn in homeworks

- We'll be using the CollectIt system for assignments.
- Most homeworks will be due on Tuesdays at 11:59 PM
- Please compress (`tar`) your homework directory to a single file called:
  hw1.tar.

How to turn in homeworks

- We'll be using the CollectIt system for assignments.
- Most homeworks will be due on Tuesdays at 11:59 PM
- Please compress (`tar`) your homework directory to a single file called:
  hw1.tar.

How to turn in homeworks

- We'll be using the CollectIt system for assignments.
- Most homeworks will be due on Tuesdays at 11:59 PM
- Please compress (`tar`) your homework directory to a single file called:
  hw1.tar.