

Python/NLTK Tutorial

CSC 485/2501

Jackie CK Cheung

jcheung@cs.toronto.edu

22 Sept., 2011

Outline

- Introduction to Python and NLTK
- Interactive Python sessions
- Basic syntax and data types
- Modules and classes

Python

- High-level, general-purpose language
- Readable code, clear syntax
- Dynamic typing (duck typing)
- Automatic garbage collection and memory management
- Large standard library

Natural Language Toolkit (NLTK)

- Python package that implements many standard NLP data structures, algorithms
 - led by Steven Bird, Edward Loper, Ewan Klein
- Open-source
- <http://www.nltk.org>

Accessing Python and NLTK

- Option 1: Log in to your CDF account

```
% python
```

```
>>> import nltk
```

- Option 2: Install on your own machine
 - At your own risk! Make sure your code for assignments runs on CDF.
 - First install Python 2.6, then PyYAML, and NLTK.
 - <http://www.nltk.org/download>

Interactive Shell

- Enter using `python` command

Most Important Commands!

- `help`
- `dir`
- `print`

Immutable Data Types

- Cannot be changed!
 - Methods produce new objects
- Booleans: True, False
 - and, or, not
- Numeric: int, float
 - +, -, *, /, //, **, %
- Strings: str
 - A character is simply a one-character long str.

Mutable Data Types

- Can be changed
 - Methods change the object itself
- Lists: list
 - zero-indexed
- Maps: dict
 - associative container
 - hash table or map in other languages

Statements

- `if, for, while`
 - Mind the colons!
- Indentation takes the place of curly braces in other languages!
 - Recommendation: use an IDE that has Python syntax highlighting, converts tabs to space

Functions

- `def`, `return`
- Lambda notation
- Functions are objects

Classes and Objects

- Initialize using `class_name(arguments)`
- Call using `object.method_name(arguments)`
- `self` in the definition of a class refers to the object instance itself
 - Similar to `this` in C++ or Java.

Script files

- You can put code that is to be reused in a script with the extension `.py`.
- `% python your_file.py`

Modules and Import

- **Option 1:** `import module_name`
 - Access by `module_name.field`
- **Option 2:**
`from module_name import field`
 - Access by `field`
- **Option 3 (not recommended):**
`from module_name import *`
 - Access by `field`
 - Potential for name clashes, unexpected behaviour

```
if __name__ == '__main__':
```

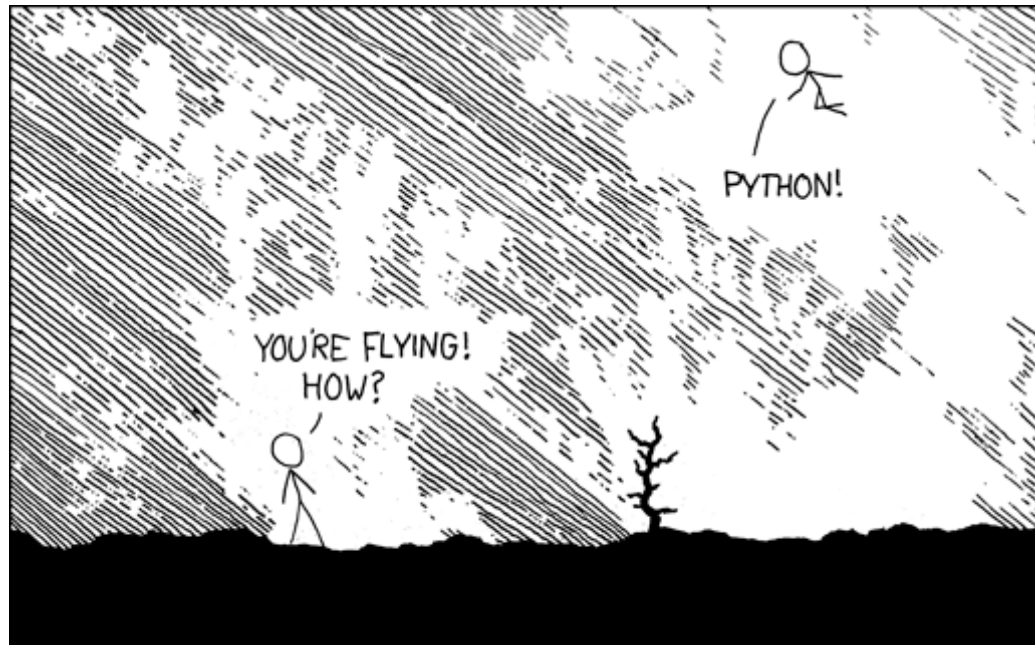
- All code in a module is executed when the module is imported.
- Put all of the code that is intended to be run with a script in the body of the above if statement.

Python Editors and IDEs

- Installed on CDF:
 - emacs (powerful, but steep learning curve)
 - wing-101 (needs X forwarding for remote access)
 - IDLE (X forwarding, comes with Python)
- Others:
 - eclipse with Python plug-in (slow, but good)
 - Notepad++ (basic editor with highlighting)

List of Tutorials

- General Python
 - <http://docs.python.org/tutorial>
- NLTK-specific
 - <http://www.nltk.org/book>



I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
|
HELLO WORLD IS JUST
print "Hello, world!"

I DUNNO...
DYNAMIC TYPING?
WHITESPACE?

COME JOIN US!
PROGRAMMING IS FUN AGAIN!
IT'S A WHOLE NEW WORLD UP HERE!




BUT HOW ARE YOU FLYING?

I JUST TYPED
import antigravity

THAT'S IT?

... I ALSO SAMPLED EVERYTHING IN THE MEDICINE CABINET FOR COMPARISON.



BUT I THINK THIS IS THE PYTHON.

<http://xkcd.com/353/>