Compsci 101, Software and Data

Owen Astrachan
Kristin Stephens-Martinez
March 29, 2018
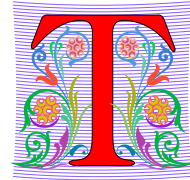

T is for …

- **Type**
  - From int to float to string to list to …
- **Text**
  - From .txt to editors to …
- **Turing Award**
  - From Alfred to Alan

Plan for the First Week of April

- Software: essential and accidental complexity
  - Fred Brooks: **No Silver Bullet**
  - Abstractions, Modules, and Design/Engineering

- Reading Files, Folders, and URLs
  - User-error's handled somehow

- Details of semester and midterm next week
  - What to study and how to study

Mid Survey

- Survey results are in and
  - The survey says:
    - +1 Final exam
    - What are we doing well? Improve?

The Power of Collaboration: WOTO

- Duke 2000: Music and Computer Science
  - https://www.stanforddaily.com/2016/03/09/qa-with-ge-wang-father-of-stanford-laptop-orchestra/
  - http://www.youtube.com/watch?v=ADEHmkL3HBg
- About Design in Compsci 308

Our investment into a huge and meticulous design process was a huge factor in making later progress. 35000+ lines of code / design / documentation gave us a project we were all very happy and proud to be a part of.

4/3/2018   Compsci 101, Spring 2018, Software and Data   19

Creating and Creativity

- You will create software in 101 and beyond
  - We are providing you tools and experience(s) to create, you may leverage these to be creative
- Do you need a license to create?

https://www.flickr.com/photos/llowrain/5523800294
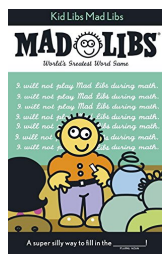
4/3/2018   Compsci 101, Spring 2018, Software and Data   20

Lab 9 Modules for Creating

- We should not call the lab/game "madlibs?
  - Does this really matter?
- Tag-a-Story

Today I've seen a <color> <noun>, but the <adjective> <food> isn't exactly how I think I want to <verb> with the <animal>s I've seen.

4/3/2018   Compsci 101, Spring 2018, Software and Data   21

Functions in Storyline.py

- Each function has a single purpose
  - Input and output: parameter(s) and return value

- Function **makeStory**:
  - Calls **getTemplateLines**
  - Calls **linesToStory**
  - Prints the story

- ***Use of other modules from same project***

4/3/2018   Compsci 101, Spring 2018, Software and Data   22

makeStory in Storyline.py

- Does this do one thing? Inputs and outputs?
  - This is the main program, as a function

```
41 def makeStory():
42     """
43     let user make a choice of
44     available templates and print
45     the story from the chosen template
46     """
47     lines = TemplateChooser.getTemplateLines("templates")
48     st = linesToStory(lines)
49     print(st)
```

We need to understand …

- Reading files and directories
  - Undersanding os, os.path, pathlib.Path, …
  - Do we really need to?
  - What about JSON?

- We're creating stories!!!

- https://www.youtube.com/watch?v=_gnypiKNaJE

Choosing from story templates

- Allow the user to select template/file at runtime
  - We don't use tkinter.filedialog
  - Navigation issues, change to web-interface

- How do we read a folder to get all filenames?
  - We can open a file, not a folder/directory?

```
>>> f = open("templates")
Traceback (most recent call last):  File "<stdin>", line
1, in <module>
IsADirectoryError: [Errno 21] Is a directory: 'templates'
```

The os library/module

- This is for "operating systems"
  - Files and directories are stored/named differently in Windows, Mac, Linux, other
  - https://docs.python.org/3/library/os.html

- See os documentation, can't always open a file given "hello.txt", may need full path
  - **/Users/ola/…/hello.txt**
  - **C:\Program Files\ola\…\hello.txt**

TemplateChooser.py Module

- What do we use from os library?
  - Functions from this module

- **os.listdir("templates")** get all files
  - "haiku.txt" and "labtemplate.txt"

- Need to create "templates/haiku.txt"
  - **os.path.join("templates", "haiku.txt")**
  - Isolate dependencies on "\" or "/" or …

Creating User "menu" of files

```
8  def dirToDictionary(dirname):
9      """
10     dirname is name of a folder containing
11     text files. Create and return a dictionary
12     in which keys are integers 0,1,2,...
13     and values are file names in folder
14     """
15     data = os.listdir(dirname)
16     d = {}
17     index = 0
18     for one in data:
19         f = os.path.join(dirname,one)
20         d[index] = f
21         index += 1
22         
23     return d
```

The pathlib and Path library

- The Path object in pathlib is more recent/Python3
  - Simpler, easier to use than functions from **os**

- Create Path object p
  - Use **p.iterdir()** to get all files/folders in p
  - Just the filenames:
    [x.parts[-1] for x in p.iterdir()]

PathTemplateChooser.py

- Using **iterdir** we can avoid **os.join**
  - Store Path objects as values in dictionary
  - Contrast storing string path names

- Can easily print just filename, but open full path
  - **p.parts[-1]** in contrast to "this/that/file.txt"
  - Path object has ***domain specific functions***
    - Contrast finding parts of "this\that\other" using **os.sep** to be portable

WOTO

http://bit.ly/101spr18-apr3-1

Python exceptions

- What should you do if you prompt user for a number and they enter "one"
  - Test to see if it has digits?

- Use exceptions with **try:** and **except:**
  - See code in function **chooseOne** from *TemplateChooser.py*

EXCEPTIONS

Handling Exceptions

- What happens: **x = int("123abc")**

```
46      st = input("choose one> ")
47      try:
48          val = int(st)
49          if 0 <= val and val < len(d):
50              return reader(d[val])
51      except ValueError:
52          print("please enter a number")
```

Christos Papadimitriou

- Gödel prize for concept of "price of anarchy"

I would say that, quite generally, computer scientists are going to find themselves interacting more with other fields. I encourage my students to go completely wild in their curriculum, to go out and learn not only that which they think they should learn in order to be good computer scientists—usually mathematics and programming and engineering and so on— but learn about everything else, about psychology, economics, about business, about biology, about the humanities. (2008)

Where do stories come from?

- Source is a template file + random choices
  - We get random choices from another module
  - Read labels and choices, choose at runtime

- The file verbs.txt stores choices of verbs
  - The file colors.txt stores choices for colors

- Create d["color"] = ["red", "orange", …, "violet"]

Replacements.py

- Choosing a color for <color> or verb for <verb>
  - Make a random choice
  - What if there is no <curling> tag?

- Creating the dictionary used in the code
  - **readDirectory()** calls **processFile()**
  - **processFile()** calls **readAndStore()**

Functions to process labels

- Why are there three functions:
  - What does **readDirectory** do?
  - What does **processFile** do?
  - What does **readAndStore** do?

- Try to make each function do one thing
  - Test independently of each other
  - Combine for desired effect/functionality

Why use files/folders?

- Could read/send <label>:[choices] via Internet
  - Store all on a web-page and parse strings

- Alternative: JSON
  - JavaScript Object Notation
  - Not restricted to JavaScript
  - Especially easy in Python: dictionary!!

JSON for story tags

- http://www.cs.duke.edu/csed/tag-a-story/tags.json

- What is format of this JSON string?
  - It's a dictionary!
  - See label : [list of choices]
  - Use commas to separate
  - Use white-space liberally

- Python json library and the function loads(…)

Peggy Li, Duke 2014

- Linked in, Clever, Nava, …

Undergraduate Teaching Assistant
Duke University
Aug 2011 – Dec 2012 • 1 yr 5 mos
Durham, NC

CS 6L / 101: Intro to Computer Science. Fall 2011, Spring 2012, and Fall 2012 (head TA).

This course inspired me to pursue Computer Science as a major, and I'm grateful I had the opportunity to pay it forward as a TA in later semesters. :)

Douglas Crockford

- Created/Popularized JSON, currently JavaScript architect at PayPal
- Releases JSMin
  - 2002, MIT license

"This software shall be used for Good, not Evil", grants IBM license "to use JSLint for evil"

JSON/Files in Replacements.py

- Code to read JSON label:[choices] dictionary
  - Compare to reading folder/files
  - We could do this locally and not over internet
  - Advantages? Disadvantages?

```
89 def readjson():
90     url = "http://www.cs.duke.edu/csed/tag-a-story/tags.json";
91     f = urllib.request.urlopen(url)
92     st = f.read().decode("utf-8")
93     d = json.loads(st)
94     return d
```

Why Use JSON?

- Particularly convenient in Python, dictionary
  - Keys and Values are typically strings or lists of strings, but int, float also possible

- A *lingua franca* for web-based communication
  - Ask for complex or hierarchical information
  - Get a JSON string

4/3/2018   Compsci 101, Spring 2018, Software and Data   43

lingua franca

Dictionary

lingua franca	🔍

lin·gua fran·ca
/ˌliNGgwə ˈfraNGkə/ 🔊

noun

a language that is adopted as a common language between speakers whose native languages are different.

- *historical*
  a mixture of Italian with French, Greek, Arabic, and Spanish, formerly used in the Levant.

⌄   Translations, word origin, and more definitions

4/3/2018   Compsci 101, Spring 2018, Software and Data   44

From Spotify to Spotipy

- Module Spotify101.py
  - No cost, no sign-up

4/3/2018   Compsci 101, Spring 2018, Software and Data   45

Questions

4/3/2018   Compsci 101, Spring 2018, Software and Data   46

8