LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

21 October 2016

Class Outline

- Practice with what to do before writing a program
- Reading a dictionary from a text file - My question to you

What to do before writing a program

1. Choosing the right data structure for a task
2. Writing loops appropriately

What to do before writing a program - Instructions

- The idea of next few slides is to make you think through and analyze what you need to write a program that suits the problem description.
- You should not write programs. You should only try to work through the problem scenarios and come with some pseudo-code
- Pseudo-code: where you describe what are your variables, what kind of objects are they (string, list, dictionary etc), what are the loops you should have etc.
- Working in groups is encouraged.
- Let us have 15-20 minutes for each question, and one person should come and explain their solution on whiteboard after that.

What to do before writing a program - Instructions

- The idea of next few slides is to make you think through and analyze what you need to write a program that suits the problem description.
- You should not write programs. You should only try to work through the problem scenarios and come with some pseudo-code
- Pseudo-code: where you describe what are your variables, what kind of objects are they (string, list, dictionary etc), what are the loops you should have etc.
- Working in groups is encouraged.
- Let us have 15-20 minutes for each question, and one person should come and explain their solution on whiteboard after that.

Note: This may look "useless" and "theoretical", but you can never "apply" anything without knowing some "theory".

When to use what data structure - 1

Morse Code

- Task: a program to convert a sentence it into morse code.
- Questions to address before starting to program:
  1. How will you take input from user?
  2. What information do you need to convert it to morse code?
  3. How will that information be stored internally?
  4. How will the output be displayed to the user?
  5. Once you have answers, write the pseudo code - describing step by step how the program should look.

When to use what data structure - 2

Ngram counting

- Let us say all books written by Author 1 are collected and stored in a single file author1.txt, and all books written by Author 2 in author2.txt.
- There are two problems: a) list the most frequent 100 bigrams in each of these authors' writings. b) compare how many bigrams overlap between these authors.
- What sort of information will you compute while writing this program?
- Once you have the answer, write the pseudo code - describing step by step how the program should look.

When to use what data structure - 3

POS Tagging

- Let us say you have access to a large newspaper corpus, and your goal is to answer the following questions:
  1. Problem 1: Find out which nouns are more commonly used in singular form and which are used in plural form.
  2. Problem 2: Which tags are nouns most commonly found after? What do these tags represent?
  3. Problem 3: How many words are ambiguous, in the sense that they appear with at least two tags?
- For all these problems, how will you design your solutions? What information do you need? What kind of data structures will you use to store data?
- Once you have answers, write the pseudo code

Note: based on Exercises 15–18 in Chapter 5 of NLTK book

Writing Loops Appropriately -1

Duplicate items in lists

- Problem: Write one program that takes a "list of numbers" as input from user, and prints out a list that shows numbers that repeat in this list.
- If I input [1,2,33,2,33,11], my program should output [2,33].
- How will you solve this problem? What are the issues you may face?
- Write your pseudo code.

Writing Loops Appropriately -1

Duplicate items in lists

- Problem: Write one program that takes a "list of numbers" as input from user, and prints out a list that shows numbers that repeat in this list.
- If I input [1,2,33,2,33,11], my program should output [2,33].
- How will you solve this problem? What are the issues you may face?
- Write your pseudo code.
- In the code sample I will show now, there are two functions. Predict what they will print. (DuplicateItems.py)

Writing Loops Appropriately -2

- Problem 1: Program should take a string as input, and print every third character in the string, if it is not a 'a'.

Writing Loops Appropriately -2

- Problem 1: Program should take a string as input, and print every third character in the string, if it is not a 'a'.
- Problem 2: Program should take a string as input, and replace each character with the character that follows it in English alphabet (i.e., if input is "Peter", it should become "Qfufs")
- Again, think through and prepare a pseudo code.

Writing Loops Appropriately -2

- Problem 1: Program should take a string as input, and print every third character in the string, if it is not a 'a'.
- Problem 2: Program should take a string as input, and replace each character with the character that follows it in English alphabet (i.e., if input is "Peter", it should become "Qfufs")
- Again, think through and prepare a pseudo code.

ForIf.py

calling a function from one program in another - skipping this as Stephanie showed it on Tuesday. We also discussed this in 516, if you remember.