

Homework 2: Lists

L435/L555

Due Thursday, September 15

1. Write a basic python program which asks the user for a number and, using the `math` package, calculates the natural log of that number, printing it to the screen. Do not worry about ill-formed input (i.e., non-numbers) or pretty output. Also, remember that python will give you a string as input, but you need to use it as a number (i.e., a `float`).

2. (a) Write a program that takes the following list, asks the user from which position elements should be deleted, reads in the number, and deletes 5 elements from that position on. (If there are fewer than 5 elements, then it deletes all the items from that position on) As a final step, the program should print the modified list. Make sure that the program does not crash if there are fewer than 5 elements left.

```
months = [ 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep',
           'Oct', 'Nov', 'Dec' ]
```

- (b) Extend the program so that the list has the number of days inserted after each month (i.e., the number for each month is hard-coded). This time, the number a user indicates is the number of the month, and instead of deleting 5 months, the program retrieves the appropriate month and its number of days from the list and prints sentences like “March has 31 days.” For example, if the user types in 2, he/she will get February as the month and thus see “February has 28 days.”

3. **Mystery:** I know you haven't really learned conditionals yet (`if`, `elif`, `else`), but I want you to figure out what the following program does.

```
numbers = [1,3,5,7,11, 13, 17, 19]
numbers.sort()
nn = int(input('new_number:_'))
if nn in numbers:
    print('tough_luck')
elif nn > 20:
    numbers.append(nn)
    print('attached')
else:
    numbers.insert(0, nn)
    print('first!')
```

4. L555 only (extra credit for L435): Go to the American National Corpus (ANC) site and download the Mini-MASC (Manually Annotated SubCorpus): <http://www.anc.org/data/masc/downloads/data-download/>. (Feel free to use some other portion of MASC; I just want to keep your download small.)

- (a) Explore the data a bit, using your unix skills and your web skills. What do the different xml files represent?
- (b) In the `-s` and `-seg` files, what are `anchors`?
- (c) Give me a `grep` command to obtain lines only with part of speech (POS) information.
- (d) At the top of the `.ptb` files, it lists tags with their `occurs` counts. Using your unix skills (and ignoring tags with asterisks), what pipeline of commands can you use to verify that these counts are accurate?