

PYTHON: LOADING & PARSING TEXT DATA

9.2.2020

ANACONDA INSTALL?



FOR LOOP

```
* words = ["this", "is", "a", "list"]  
  
* for w in words:  
    print(w)
```

FOR LOOP

- * `words = ["this", "is", "a", "list"]`
- * `for w in words:`
 `print(w)`
- * `for index in range(len(words)):`
 `print(words[index])`

OBJECT METHODS

- * *Everything in python is an **object***
- * *Objects have “methods”—functions that are attached to that object*
- * *You can access methods (and other object “attributes”) using the . syntax:*
 - * *e.g. my_variable.some_method()*

READING A TEXT FILE

- * `f = open("data_file.txt")`
- * *the file object acts like a pointer into the file*
- * `f.readline()` - *returns the next line*
- * `f.read()` - *returns the rest of the file*
- * Details: <https://docs.python.org/3/library/functions.html#open>

STRING SPLITTING

- * `s = "this is a string"`
- * `s.split() -> ["this", "is", "a", "string"]`
- * *splits on whitespace by default*
- * Details: <https://docs.python.org/3/library/stdtypes.html#str>

LIST APPEND

- * `my_list = ["wow"]`
- * `my_list.append("such")` - *add "such" to list*
`my_list.append("list")` - *add "list" to list*
- * `print(my_list)`
`=> ["wow", "such", "list"]`

THE PROBLEM

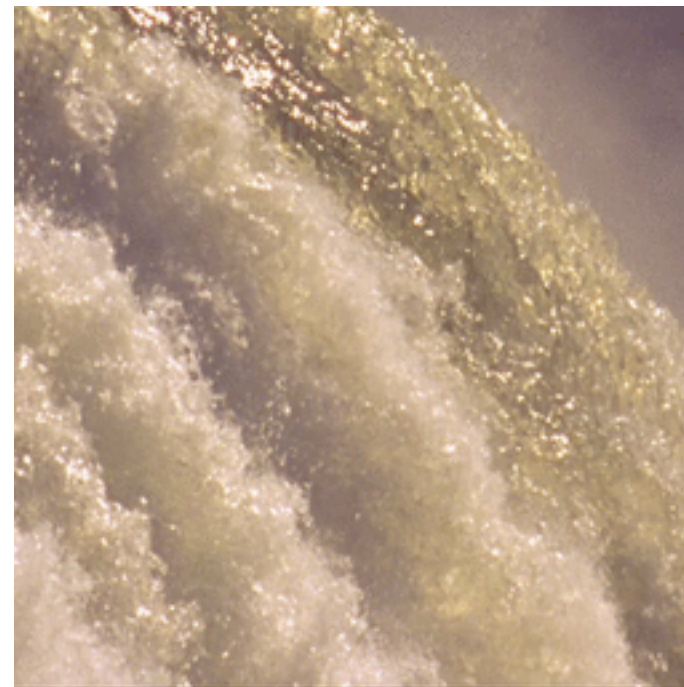
- * We need to parse a text file (labels.txt) that contains labels for 7200 stimuli used in our experiment
- * Each line looks like this:

15 ani16.gif text.n.01 shelf.n.01 sweater.n.01 hand.n.01

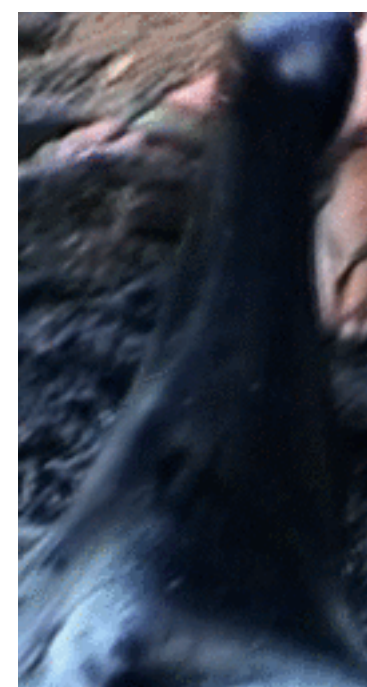
index |
video clip
filename

object/action labels

VISUAL SEMANTICS



horseman
horse
walk
road
spectator
mountain



THE PROBLEM

- * What kind of data structure do we want the result to be in?
- * It should be easy for us to use later
- * It should be ~obvious what's going on

THE SOLUTION

- * Let's read the file line-by-line, turn each line into a dictionary, and put all the dictionaries in a list
- * *See `parse_labels.py` in this lecture directory for a demo of how this works*

THAT'S ALL, FOLKS!