

CS 1656 – Introduction to Data Science (Spring 2018)

Instructor: Dr. Alexandros Labrinidis

Teaching Assistant: Evangelos Karageorgos

Additional Credits: Phuong Pham, Anatoli Shein, Zuha Agha

Recitation 1: Python Tutorial

In this recitation, you will learn file I/O, JSON files, and CSV files in Python. Packages you will need are:

- os
- csv
- json

Part 1: File I/O

Reading and writing text files is typically very straightforward. Here is an overview of file I/O in Python.

a. Getting input filename from command line

```
file_in = raw_input('Enter name of input file: ')
```

b. Checking if file exists and manipulating pathnames

```
if os.path.isfile(os.path.join(os.getcwd(), file_in))
```

c. Open/Close and modifying file position

```
f = open(file_in)
print f.tell()           # current position
f.read(1)                # read one byte and move forward
f.next()                 # get bytes from the file until newline
f.seek(-3,2)              # move back 3 characters before the end
f.seek(0)                # move back to beginning of file
f.close()                # important if writing to file
```

Another way of opening the file is to use *with* statement. The use of *with* statement establishes a context in which the file will be used and when control leaves the *with* block, the file will be closed automatically. You don't need to use the *with* statement, but if you don't use it, make sure you remember to close the file.

```
with open(file_in, "r+") as f:
    print f.tell()
```

d. Read/Write to Files

There are several ways of reading and writing to files in Python. Here we show some of them.

One way of reading the file is:

```
read_data = f.read()
```

CS 1656 – Introduction to Data Science (Spring 2018)

Instructor: Dr. Alexandros Labrinidis

Teaching Assistant: Evangelos Karageorgos

Additional Credits: Phuong Pham, Anatoli Shein, Zuha Agha

An alternate way of reading line by line by iterating over the file is:

```
for line in f:
    print(line)
```

In order to write to a file,

```
f.writelines("Yay! Written to file on Friday, 9/9/16 :D \n")
```

Part 2: JSON

JSON file format is used to save generic data programmatic data structures as strings and is commonly used for serialization. JSON is good for storing simple data structures into text file and sending data to other people.

```
zipCodes = [60290, 60601, 60602, 60603, 60604, 60605, 60606]
# dump into a file
f = open('example2.json', 'wb')
json.dump(zipCodes, f)
f.close()
# load back into Python
f = open('example2.json', 'rb')
zipCodes2 = json.load(f)
f.close()
# compare
print "Checking zipcodes..."
print(zipCodes == zipCodes2)
```

Part 3: JSON / read

In this example, we will download a JSON file from a URL and display the file one line at a time.

```
print('Downloading JSON file and printing entire file:')
response = get('http://cs1656.org/data/hours.json')
print(response.content)
print('Loading as JSON and iterating one line at a time:')
hours = json.loads(response.content)
print(hours)
print('\nIterating over JSON:')
for line in hours:
    print line
```

CS 1656 – Introduction to Data Science (Spring 2018)

Instructor: Dr. Alexandros Labrinidis

Teaching Assistant: Evangelos Karageorgos

Additional Credits: Phuong Pham, Anatoli Shein, Zuha Agha

Part 4: Convert to CSV – On your own

For the next task you need to write a python program to convert the JSON file (downloaded from Part 4) into a CSV file that you need to save to disk. To do this you should look up the `csv` module and the `writer()` function in particular.

Part 5: Read CSV – On your own

For the next task you need to write a python program to read the CSV file from disk and display its raw contents (without recognizing rows or fields).

Part 6: Read from CSV / rows – On your own

For the next task you need to write a python program to read the CSV file from disk and display its contents one row at a time (without recognizing fields).

Part 7: Read from CSV / cells – On your own

For the next task you need to write a python program to read the CSV file from disk and display its contents one row at a time and then, within each row, one field at a time

References

[1] Martelli, Alex, Anna Ravenscroft, and David Ascher. *Python cookbook*. "O'Reilly Media, Inc.", 2005.

[2] <https://docs.python.org/3/>