

Algorithms and Data Structures for Data Science

lab_parsing

CS 277

February 3, 2023

Brad Solomon



UNIVERSITY OF
ILLINOIS
URBANA - CHAMPAIGN

Department of Computer Science

Learning Objectives

Review fundamentals of Python I/O

Review built-in string methods

Create useful code that can read common text file formats

File I/O in Python

```
1 readableFile = open('inputFile.txt', 'r')
2
3
4
5
6 writableFile = open('outputFile.txt', 'w')
7
8
9
10
11 carefulWriteFile = open('outputFile.txt', 'x')
12
13
14
15
16 appendableFile = open('outputFile.txt', 'a')
17
18
19
20
21
22
23
```

File I/O in Python

Which approach do you prefer?

```
1 # Approach 1
2
3 readableFile = open('inputFile.txt', 'r')
4
5
6 fileData = readableFile.read()
7
8
9 readableFile.close()
10
11
12 # Approach 2
13
14 with open('inputFile.txt', 'r') as myFile:
15     fileData = myFile.read()
16
17
18
```

File I/O in Python

```
1 with open('data/temp1.txt', 'x') as myFile:
2     for i in range(10):
3         myFile.write(str(i))
4     myFile.write("\n")
5     myFile.write("Line 2")
6
7
8
9 myFile = open('data/temp2.txt', 'w')
10 for i in range(5):
11     myFile.write(str(i) + "\n")
12 myFile.close()
13
14
15
16 with open('data/temp2.txt', 'a') as myFile:
17     myFile.write("Hello World!\n")
18
19 with open('data/temp2.txt', 'a') as myFile:
20     myFile.write("Hello World!\n")
21
22
23
```

File I/O in Python

```
1 with open('data/temp1.txt', 'x') as myFile:
2     for i in range(10):
3         myFile.write(str(i))
4     myFile.write("\n")
5     myFile.write("Line 2")
6
7
8
9 myFile = open('data/temp2.txt', 'w')
10 for i in range(5):
11     myFile.write(str(i) + "\n")
12 myFile.close()
13
14
15
16 with open('data/temp2.txt', 'a') as myFile:
17     myFile.write("Hello World!\n")
18
19 with open('data/temp2.txt', 'a') as myFile:
20     myFile.write("Hello World!\n")
21
22
23
```

1	0123456789
2	Line 2

temp1.txt

1	0
2	1
3	2
4	3
5	4
6	

temp2.txt

File I/O in Python

```
1 with open('data/temp1.txt', 'x') as myFile:
2     for i in range(10):
3         myFile.write(str(i))
4     myFile.write("\n")
5     myFile.write("Line 2")
6
7
8
9 myFile = open('data/temp2.txt', 'w')
10 for i in range(5):
11     myFile.write(str(i) + "\n")
12 myFile.close()
13
14
15
16 with open('data/temp2.txt', 'a') as myFile:
17     myFile.write("Hello World!\n")
18
19 with open('data/temp2.txt', 'a') as myFile:
20     myFile.write("Hello World!\n")
21
22
23
```

1	0123456789
2	Line 2

temp1.txt

1	0
2	1
3	2
4	3
5	4
6	Hello World!
7	

temp2.txt

File I/O in Python

```
1 with open('data/temp1.txt', 'x') as myFile:
2     for i in range(10):
3         myFile.write(str(i))
4     myFile.write("\n")
5     myFile.write("Line 2")
6
7
8
9 myFile = open('data/temp2.txt', 'w')
10 for i in range(5):
11     myFile.write(str(i) + "\n")
12 myFile.close()
13
14
15
16 with open('data/temp2.txt', 'a') as myFile:
17     myFile.write("Hello World!\n")
18
19 with open('data/temp2.txt', 'a') as myFile:
20     myFile.write("Hello World!\n")
21
22
23
```

1	0123456789
2	Line 2

temp1.txt

1	0
2	1
3	2
4	3
5	4
6	Hello World!
7	Hello World!
8	

temp2.txt

File I/O in Python

```
1 with open('data/temp1.txt') as myFile:
2     inList = myFile.readlines()
3 print(inList)
4
5
6
7
8
9 myFile = open('data/temp2.txt')
10 for i in range(10):
11     print("Line Content: {}".format(myFile.readline()))
12 myFile.close()
13
14
15
16
17
18 with open('data/temp1.txt') as myFile:
19     print(myFile.read())
20
21
22
23
```

1	0123456789
2	Line 2
3	
4	
5	
6	

temp1.txt

1	0
2	1
3	2
4	3
5	4
6	

temp2.txt

File I/O in Python

```
1 with open('data/temp1.txt') as myFile:
2     inList = myFile.readlines()
3 print(inList)
4
5
6
7
8
9 myFile = open('data/temp2.txt')
10 for i in range(10):
11     print("Line Content: {}".format(myFile.readline()))
12 myFile.close()
13
14
15
16
17
18 with open('data/temp1.txt') as myFile:
19     print(myFile.read())
20
21
22
23
```

1	0123456789
2	Line 2
3	
4	
5	
6	

temp1.txt

1	0
2	1
3	2
4	3
5	4
6	

temp2.txt

```
['0123456789\n', 'Line 2']
Line Content: 0

Line Content: 1

Line Content: 2

Line Content: 3

Line Content: 4

Line Content:
Line Content:
Line Content:
Line Content:
0123456789
Line 2
```

File I/O in Python

`str.strip()` will remove whitespace from the string.

1	0
2	1
3	2
4	3
5	4
6	

temp2.txt

```
1 myFile = open('data/temp2.txt')
2 for i in range(6):
3     print("Line Content: {}".format(myFile.readline().strip()))
4 myFile.close()
5
6 with open('data/temp2.txt') as myFile:
7     for line in myFile:
8         print(line.strip())
```

```
Line Content: 0
Line Content: 1
Line Content: 2
Line Content: 3
Line Content: 4
Line Content:
0
1
2
3
4
```

File I/O in Python

`str.split()` will 'split' a string into a list of substrings

The split is determined by a break character, which defaults to whitespace

```
1 with open('data/temp1.txt') as myFile:
2     text = myFile.read()
3     print("{}'".format(text))
4
5     print(text.split())
6
7     print(text.split("5"))
8
9
10
11
12
13
```

File I/O in Python

`str.split()` will 'split' a string into a list of substrings

1	0123456789
2	Line 2
3	
4	
5	
6	

temp1.txt

The split is determined by a break character, which defaults to whitespace

```
1 with open('data/temp1.txt') as myFile:
2     text = myFile.read()
3     print("{}'".format(text))
4
5     print(text.split())
6
7     print(text.split("5"))
8
9
10
11
12
13
```

'0123456789
Line 2'

['0123456789', 'Line', '2']

['01234', '6789\nLine 2']

Python File I/O

Whats wrong here:

```
1  
2 with open('dataFile2.txt','w') as myFile:  
3     data = myFile.readlines()  
4  
5 for line in data:  
6     print(line)  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18
```

Coding the lab

1) Complete the functions in order — use previous functions!

2) Consider what problem you are trying to solve

What resolution do I need to answer the question?

What is the easiest or fastest way to access the parts I need?