Social Media & Text Analysis

part 1 - Intro to Python

Ohio State University
Instructor: Wei Xu
Website: socialmedia-class.org
Why Python?

• Python is an object-oriented and high level programming language (first released in 1991).

• Very beginner-friendly!
  - shorter code needed for the same task.

• Very powerful!
  - many well-maintained libraries (e.g. numpy, Matplotlib, Scikit-learn, PyTorch, TensorFlow,)
  - a popular programming language in AI and machine learning research
Simplicity of Python

• Create a list of integers:
  • Python

```
nums = [45, 23, 51, 32, 5]
```

• Java, in contrast:

```
List<Integer> nums =
    Arrays.asList(new Integer[] {45, 23, 51, 32, 5});
```
Simplicity of Python

• Create a list of integers, and print them out:

  • Python

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print(idx, num)
```

  • Java, in contrast:

```java
List<Integer> nums = Arrays.asList(new Integer[]{45, 23, 51, 32, 5});
for (int i = 0; i < nums.size(); i++) {
    String number = nums.get(i);
    System.out.println(i + " " + number);
}
```

Simplicity of Python

• Create a list of integers, and print them out:

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print (idx, num)
```

output

```
(0, 45)
(1, 23)
(2, 51)
(3, 32)
(4, 5)
```
Simplicity of Python

- Create a list of integers, and print them out:

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print (idx, num)
```

Lists are declared using `[item1, item2, item3, ...]`
Simplicity of Python

- Create a list of integers, and print them out:

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print (idx, num)
```

```
0 45
1 23
2 51
3 32
4 5
```

Iterates over each item in the list, yielding (index, value) tuples.
Simplicity of Python

• Create a list of integers, and print them out:

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print (idx, num)
```

```
0 45
1 23
2 51
3 32
4 5
```

simple and very useful print() function
Simplicity of Python

• Create a list of integers, and print them out:

```python
nums = [45, 23, 51, 32, 5]
for idx, num in enumerate(nums):
    print (idx, num)
```

Indentation is very important in Python!
Try it Out!

• We will use Google's Colab programming environment:

![Google's Colab](image-url)
Basic String Operations

• A Code Sample:

```c
sent1 = "Hello world!"  # A comment.
print (sent1)           # Another one.

print (sent1[4])        # The 5th char (index starts from 0)

l = len(sent1)          # The length (in number of characters)
print ("There are " + str(l) + " characters.")

tokens = sent1.split()  # Split a string by space
print (tokens)

print (len(tokens))     # The length (in number of tokens/words)
```

Hello world!

- There are 12 characters.
- ['Hello', 'world!']
- 2
**Basic String Operations**

- **A Code Sample:**

  ```python
  sent1 = "Hello world!"  # A comment.
  print (sent1)          # Another one.
  print (sent1[4])       # The 5th char (index starts from 0)
  l = len(sent1)         # The length (in number of characters)
  print ("There are " + str(l) + " charasters.")
  tokens = sent1.split() # Split a string by space
  print (tokens)
  print (len(tokens))   # The length (in number of tokens/words)
  ```

  ![Code output]

  **output**

  - Hello world!
  - There are 12 charasters.
  - ['Hello', 'world!']
  - 2

Characters in a string can be assessed using the [] syntax.
Basic String Operations

• A Code Sample:

```python
sent1 = "Hello world!"  # A comment.
print (sent1)           # Another one.

print (sent1[4])        # The 5th char (index starts from 0)
l = len(sent1)          # The length (in number of characters)
print ("There are " + str(l) + " charasters.")
tokens = sent1.split()  # Split a string by space
print (tokens)
print (len(tokens))     # The length (in number of tokens/words)
```

Hello world!

There are 12 charasters.
["Hello", "world!"]
2

The `len(string)` function returns the length of a string.
Basic String Operations

- A Code Sample:

```python
sent1 = "Hello world!"  # A comment.
print (sent1)           # Another one.
print (sent1[4])        # The 5th char (index starts from 0)
l = len(sent1)          # The length (in number of characters)
print ("There are " + str(l) + " characters.")
tokens = sent1.split()  # Split a string by space
print (tokens)          # The length (in number of tokens/words)
print (len(tokens))
```

output

Hello world!
There are 12 characters.
['Hello', 'world!']
2

The `str()` function converts values to a string data type.
Basic String Operations

• A Code Sample:

```python
sent1 = "Hello world!"  # A comment.
print (sent1)           # Another one.
print (sent1[4])        # The 5th char (index starts from 0)
l = len(sent1)          # The length (in number of characters)
print ("There are " + str(l) + " charasters.")
tokens = sent1.split()  # Split a string by space
print (tokens)
print (len(tokens))     # The length (in number of tokens/words)
```

Hello world!

There are 12 charasters.
['Hello', 'world!']

The + operator can concatenate two strings.
Basic String Operations

• A Code Sample:

```python
sent1 = "Hello world!"  # A comment.
print (sent1)  # Another one.
print (sent1[4])  # The 5th char (index starts from 0)
l = len(sent1)  # The length (in number of characters)
print ("There are " + str(l) + " charasters.")
tokens = sent1.split()  # Split a string by space
print (tokens)
print (len(tokens))  # The length (in number of tokens/words)
```

```
Hello world!
There are 12 charasters.
["Hello", "world!"]
2
```

The `split()` function returns a list of substrings.
More Resources

• Python :
  - Google’s class: https://developers.google.com/edu/python/
  - Christophe Morisset’s notebook: https://github.com/Morisset/Python-lectures-Notebooks/blob/master/Notebooks/intro_Python.pdf
  - and many others …

• Got a Error Messages or questions?
  - Search on Google
  - StackOverflow