Inspire…Educate…Transform.

Text Mining

Lab 01 - 2017-02-18

Which of the following languages do you know

- C or C++
- Java
- Perl
- Scheme
- Fortran
- Etc..

Today's Agenda

- How to Learn Python

Outline

- Installation
- Basic datatypes
- Conditional Statements and Loops
- Functions

Installation

- Introduction to Python
  - Anaconda installation- platform python 2.7
  - Please follow the steps given in piazza

Different Notebooks to be used

- Jupyter
- Spyder
- Idle

you can choose whatever you want,

Just type in start.

But we choose jupyter(for this module).

Simple Intro for Jupyter

- Jupyter Environment
  - Opening a new notebook
  - Kernel and cell
  - Executing cell (using Shift+Enter)
  - Creating a new cell
  - Creating a heading/markdown
  - Using help (? and tab)

Printing simple messages

- Usage of print
  - Eg: print (‘hai python’)
  - Eg: print(2)
- Elements separated by commas print with a space between them
  - Eg: print(123),
  - print(‘hai’)
- A comma at the end of the statement (print ‘hello’,) will not print a newline character

Basic Data Types in Python

- Integers
- Lists
- Strings
- Tuples
- Dictionary

Integers

- Are not declared, just assigned
- The variable is created the first time you assign it a value
- Everything in Python is an object

Eg: a=10

a, b=10,15

it assigns a=10 b=15

Lists

- List is a collection of elements of same or different data type.

  l1=[‘a’,1,2,’4.5’]

Group of elements given in a pair of ’[‘, ’]’(square brackets) is called list.

Strings

- Strings are amongst the most popular types in Python.
- We can create them simply by enclosing characters in quotes. Python treats single quotes the same as double quotes.
- Creating strings is as simple as assigning a value to a variable

  Ex: s1='python' s2='hai'

Tuples

- A tuple is a sequence of immutable Python objects.
- Tuples are sequences, just like lists. The differences between tuples and lists are, the tuples cannot be changed unlike lists
- Tuples use parentheses, whereas lists use square brackets.
  - X=(1,2,3)
  - Y=()
  - Z=(15,)

Sets

- A set is an unordered collection with no duplicate elements.
- Basic uses include membership testing and eliminating duplicate entries.
- Set objects also support mathematical operations like union, intersection, difference.
  - basket = ['apple', 'orange', 'apple', 'pear', 'orange', 'banana']
  - fruit = set(basket)

Dictionary

- A dictionary is a key value pair called item.
- Each key is separated from its value by a colon (:), the items are separated by commas, and the whole thing is enclosed in curly braces.
- An empty dictionary without any items is written with just two curly braces, like this: {}.
  - Eg: dict={‘Name’:’ABC’, ’Marks’:12}

Accessing values:

- Simple variables can be accessed through their names.

  Eg: a=10; print(a)

- Strings,Lists and Tuples are accessed through the position of the elements called index.

- Dictionary values can be accessed through keys

Accessing values in Lists:

- Lists index starts from zero.
  - Eg: L1=[10,15,20,25]
    - L1[0] gives 10
    - L1[1] gives 15
    - L1[2] gives 20
    - L1[3] gives 25

    These can also be accessed by simply writing

    L1[0:4] called slicing.

    You can try and see L1[0:],L1[2:3],L1[:3]

Accessing values in Lists:

- Lists can also be accessed from the end.
  - Eg: L1=[3,4,5]
  - L1[-1] prints 5
  - L1[-2] prints 4
  - L1[-3] prints 3
- Eg: L1=[5,10,'are',5.5,6.7]
  - Try L1[2:-1]

Built in functions and methods of lists

len()—gives length of list

max()---- maximum element of list

min()----- minmum element of list

append()

extend()

count()

index()

sort()

reverse()

Etc...

Built in methods of lists

Append()

l1=[15,18,20]

l1.append(21)

Now l1 updated to [15,18,20,21]

We can only add only single element

Built in methods of lists

- extend
  - List1=[‘ML’,’TEXT’,’23’,23]
  - List1.extend([4,5,6])
  - Now list1 is updated to
    - [ML’,’TEXT’,’23’,23,4,5,6]
    - You try List1+[1,2,3] and print List1
    - List2=List1+[1,2,3] print List1 and List2 see the difference

Built in methods of lists

- Remaining methods please explore..
  - https://www.tutorialspoint.com/python/python_lists.htm
  - https://www.learnpython.org/en/Lists

  Etc..

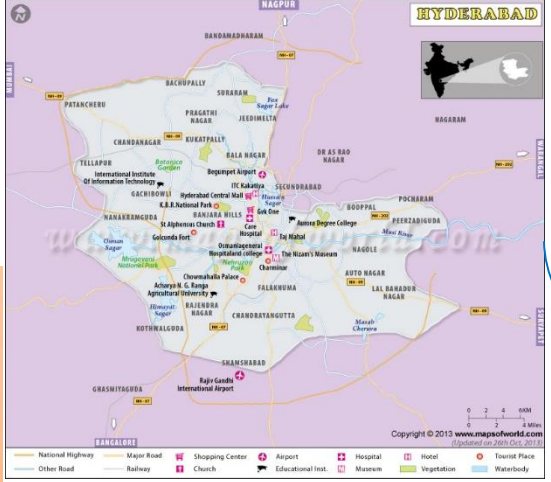
- Remaining data types and their uses

  Please check in python_Basics.ipynb file.

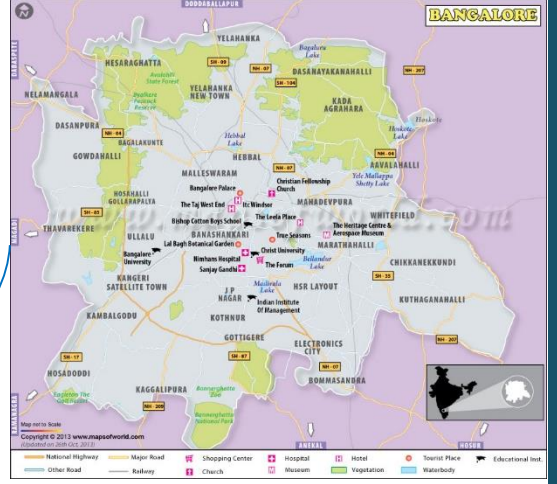
  Thanks for your co-operation.

Text Preprocessing

- Splitting the words
- Removal of Stopwords
- Removal of whitespaces, considering only words
- Removal of numbers/ certain characters-** (Regular expressions)
- Stemming and Lemmatization



INSOFE
Inspire...Educate...Transform.



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar,
Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road,
Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza,
Jakkasandra Main Road, Teacher's Colony, 14th Main
Road, Sector – 5, HSR Layout, Bengaluru - 560102

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.