**2017 Fall : COMP-SCI 5590 – Special Topics**

**Python Programming**

**Lesson3: Sets and Dictionaries**

In this lesson, we learn about sets and dictionaries in python. We also exercise some useful application of them.

1. Sets: Usage, Features, Operations on sets as union, Intersection and Difference
2. Dictionaries: Usage, Features, Operations on dictionaries as remove, update and Insert, using dictionaries in loops
3. Some simple functions and features of extracting data of web pages like wikipedia

**Use case Description:**

Extract data of Wikipedia then save that data as a file.

There is different kind of library that helps to extract data of Wikipedia. Some works on the features of the Wikipedia, structure getting revisions and so on. Here we learn how to extract data of Wikipedia with a keyword then save those data in the file.

**In class exercise:**

Write a simple program that parse a wikipage then extract the headers of the page.

Steps to do this simple program:

1. Import these libraries

import requests

from bs4 import BeautifulSoup

import os

2. define a variable and put the link you are willing to extract data of that

3. use the Request library to download the url in another variable

ex: sourcecode=Request.get(url)

4. parse the sourceCode using the BeautifulSoap library and save the parsed code in a variable

5.use findAll(‘div’) to find all the div in the parsed sourcecode

6. use a loop by the result that you have got of the step 5 to find heading

ex: for div in result:

R1=div.find(‘h1’)

7. print the content of the R1

8. do the same for printing the body

**Dictionaries and sets**

1. Write a Python program to get a comma separated sequence of words then the output should be sorted in comma separated manner.

Ex:

Input: music, film, python

Output: film, music, python

1. Use a list which contains three dictionaries. For example [{},{},{}]. Then sum the values of the dictionaries if they are integer.

For more explanations:

The input should be something like this:

[{**'course '**: **python**, **'LastGPA'** : 90, **'CurrentGPA'**: 80},  
{ **'course'**: **python**, **'LastGPA'** : 95, **'CurrentGPA'**: 85},  
{ **'course'**: **python**, **'LastGPA'** : 100,**'CurrentGPA'**: 100}]

And the output like this: notice that one element has been deleted, and the new element is the join of them, the number in the new list is the average of the old numbers.

[{**'course'**: **python**, **' LastGPA+CurrentGPA'** : 85},  
{ **'course'**: **python**, **' LastGPA+CurrentGPA'** : 90},  
{ **'course'**: **python**, **' LastGPA+CurrentGPA'** : 100}]

You can use a for loop to iterate through the list, then in each iteration, delete the item you don’t want using pop,keeping its value in a variable like n1, again deleting the item you do not want, and keeping its value in a variable like n2. Finally get the average of the vaues in n1 and n2, assign that to new item in the list like: d[‘lastGPAandCurrentGPA’]=85