**Web Mining (CSE3024)**

**Lab Assignment 10**

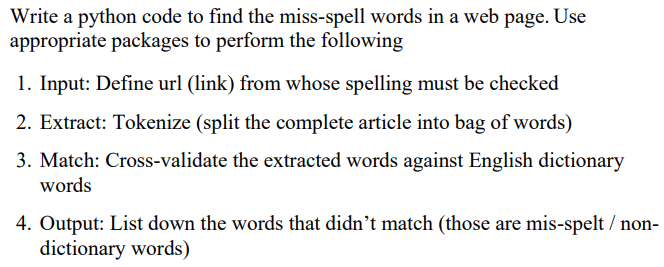
Name: **Kritika Mishra**

Registration Number: **16BCI0041**

Slot: L15+L16

Date: 24th October 2018

Question:



Code:

|  |
| --- |
| **import newspaper**  **from nltk import word\_tokenize** |
| **import enchant** |
| **import re** |
| **url = 'https://hackernoon.com/dilemmas-of-a-digital-lifestyle-27c044940157' #input URL** |
| **my\_article = newspaper.Article(url,language='en')** |
| **my\_article.download()** |
| **my\_article.parse()** |
| **d = enchant.Dict("en\_US")** |
| **print(list(set([word.encode('ascii', 'ignore') for word in word\_tokenize(my\_article.text) if d.check(word) is False and re.match('^[a-zA-Z ]\*$',word)] )))** |

Output:

