Web Mining Lab Assignment

Name: Om Ashish Mishra
Registration Number: 16BCE0789
Slot: F2

Q.

Write a python code to find the miss-spell words in a web page.
Use appropriate packages to perform the following

1. Input: Define url (link) from whose spelling must be checked

2. Extract: Tokenize (split the complete article into bag of words)

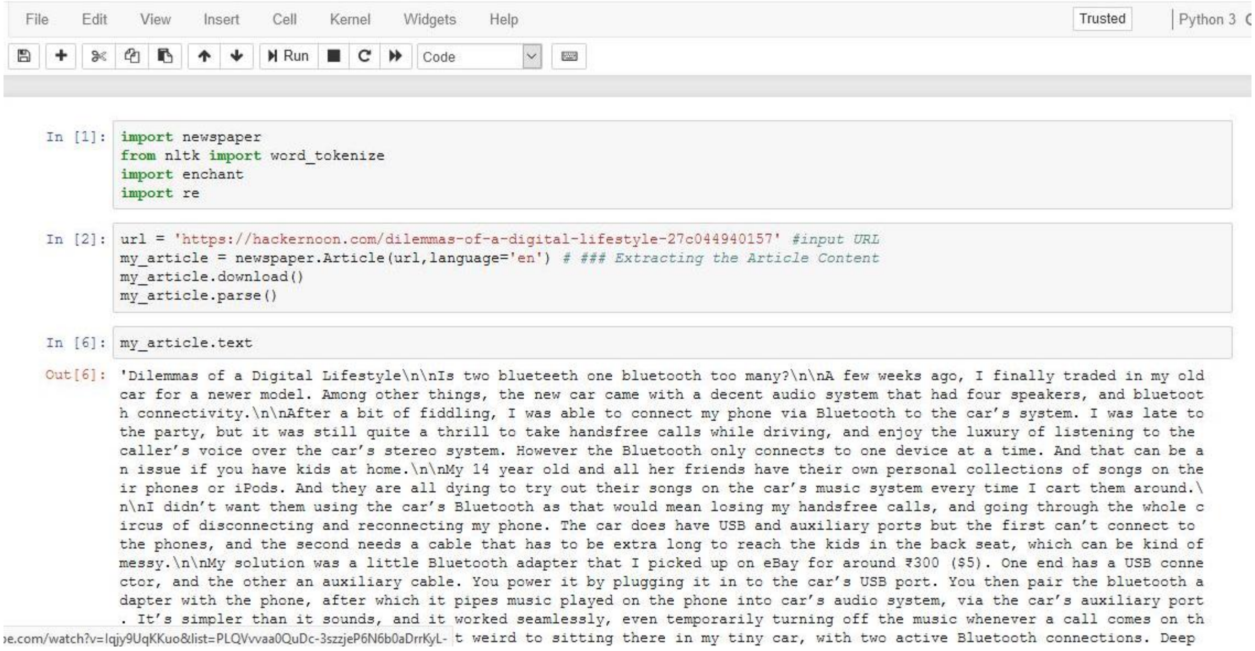
3. Match: Cross-validate the extracted words against English dictionary words

4. Output: List down the words that didn’t match (those are mis-spelt / non-dictionary words)

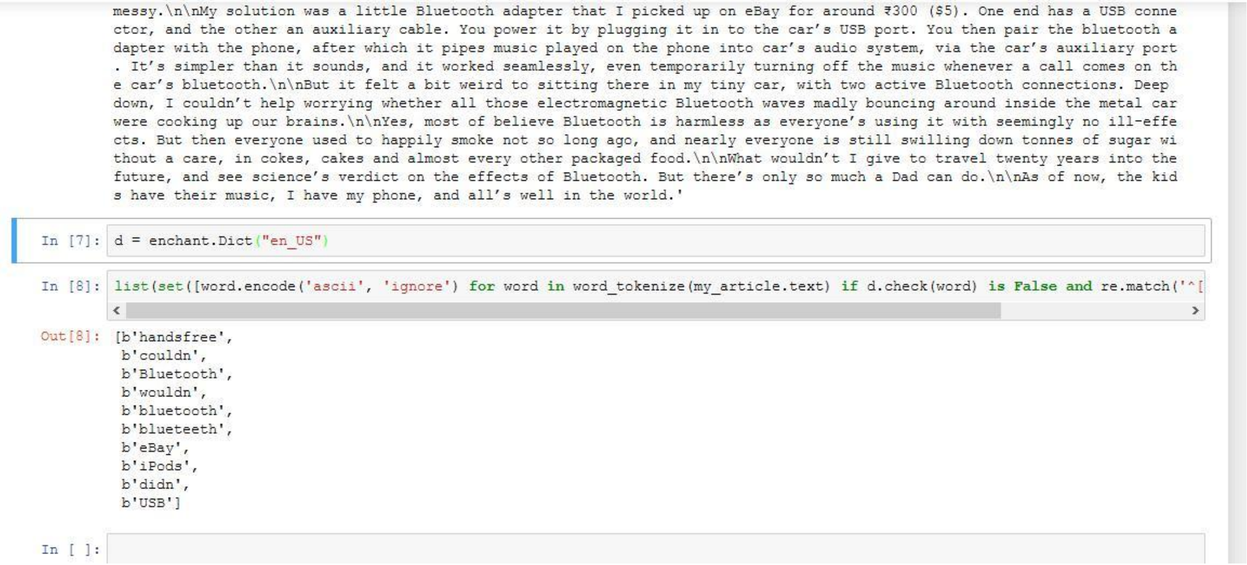
CODE:

```
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') # ### Extracting the Article Content

my_article.download()

my_article.parse()

my_article.text

d = enchant.Dict("en_US")

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:


File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3
Run
Code
be.com/watch?v=lqjy9UqKKuo&list=PLQVvvaa0QuDc-3szzjeP6N6b0aDrrKyL-

```
In [1]: import newspaper
        from nltk import word_tokenize
        import enchant
        import re

In [2]: url = 'https://hackernoon.com/dilemmas-of-a-digital-lifestyle-27c044940157' #input URL
        my_article = newspaper.Article(url,language='en') # ### Extracting the Article Content
        my_article.download()
        my_article.parse()

In [6]: my_article.text

Out[6]: 'Dilemmas of a Digital Lifestyle\n\nIs two blueteeth one bluetooth too many?\n\nA few weeks ago, I finally traded in my old
        car for a newer model. Among other things, the new car came with a decent audio system that had four speakers, and bluetoot
        h connectivity.\n\nAfter a bit of fiddling, I was able to connect my phone via Bluetooth to the car’s system. I was late to
        the party, but it was still quite a thrill to take handsfree calls while driving, and enjoy the luxury of listening to the
        caller’s voice over the car’s stereo system. However the Bluetooth only connects to one device at a time. And that can be a
        n issue if you have kids at home.\n\nMy 14 year old and all her friends have their own personal collections of songs on the
        ir phones or iPods. And they are all dying to try out their songs on the car’s music system every time I cart them around.\
        n\nI didn’t want them using the car’s Bluetooth as that would mean losing my handsfree calls, and going through the whole c
        ircus of disconnecting and reconnecting my phone. The car does have USB and auxiliary ports but the first can’t connect to
        the phones, and the second needs a cable that has to be extra long to reach the kids in the back seat, which can be kind of
        messy.\n\nMy solution was a little Bluetooth adapter that I picked up on eBay for around ₹300 ($5). One end has a USB conne
        ctor, and the other an auxiliary cable. You power it by plugging it in to the car’s USB port. You then pair the bluetooth a
        dapter with the phone, after which it pipes music played on the phone into car’s audio system, via the car’s auxiliary port
        . It’s simpler than it sounds, and it worked seamlessly, even temporarily turning off the music whenever a call comes on th
                                                                t weird to sitting there in my tiny car, with two active Bluetooth connections. Deep
```



```
        messy.\n\nMy solution was a little Bluetooth adapter that I picked up on eBay for around ₹300 ($5). One end has a USB conne
        ctor, and the other an auxiliary cable. You power it by plugging it in to the car’s USB port. You then pair the bluetooth a
        dapter with the phone, after which it pipes music played on the phone into car’s audio system, via the car’s auxiliary port
        . It’s simpler than it sounds, and it worked seamlessly, even temporarily turning off the music whenever a call comes on th
```

```
        e car’s bluetooth.\n\nBut it felt a bit weird to sitting there in my tiny car, with two active Bluetooth connections. Deep
        down, I couldn’t help worrying whether all those electromagnetic Bluetooth waves madly bouncing around inside the metal car
        were cooking up our brains.\n\nYes, most of believe Bluetooth is harmless as everyone’s using it with seemingly no ill-effe
        cts. But then everyone used to happily smoke not so long ago, and nearly everyone is still swilling down tonnes of sugar wi
        thout a care, in cokes, cakes and almost every other packaged food.\n\nWhat wouldn’t I give to travel twenty years into the
        future, and see science’s verdict on the effects of Bluetooth. But there’s only so much a Dad can do.\n\nAs of now, the kid
        s have their music, I have my phone, and all’s well in the world.'

In [7]: d = enchant.Dict("en_US")

In [8]: list(set([word.encode('ascii', 'ignore') for word in word_tokenize(my_article.text) if d.check(word) is False and re.match('^[

Out[8]: [b'handsfree',
         b'couldn',
         b'Bluetooth',
         b'wouldn',
         b'bluetooth',
         b'blueteeth',
         b'eBay',
         b'iPods',
         b'didn',
         b'USB']

In [ ]:
```