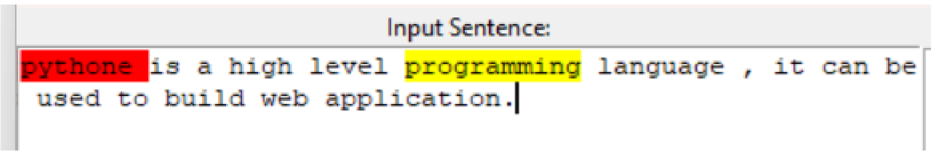
Results

The following test cases showed that the proposed spell-checking system is capable of detecting non-words and real-words errors. Non-words error which is due to typing error is highlighted as red while real-words error which happens due to wrong context is highlighted as yellow. For each error, the system will provide a list of suggested words based on edit distance (for non-words error) and real-words error is based on the bigram frequency.

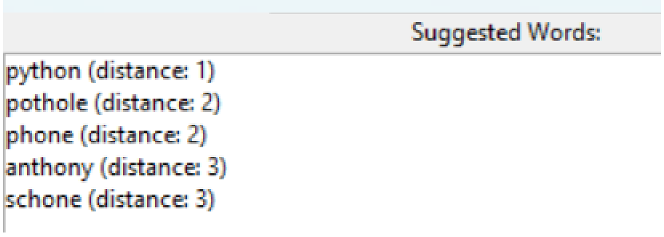
Case 1:



Input Sentence:

```
pythone is a high level programming language , it can be
 used to build web application.
```

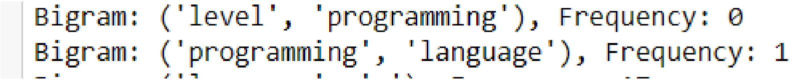
Non-words error:



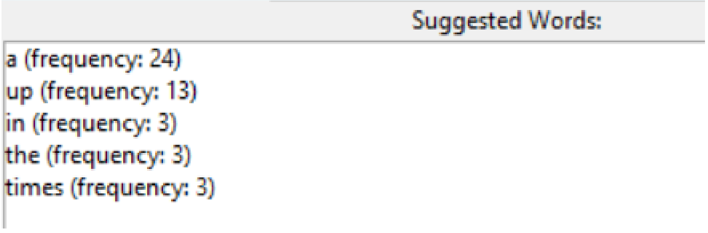
Suggested Words:

python (distance: 1)
pothole (distance: 2)
phone (distance: 2)
anthony (distance: 3)
schone (distance: 3)

Real-words error:



```
Bigram: ('level', 'programming'), Frequency: 0
Bigram: ('programming', 'language'), Frequency: 1
```

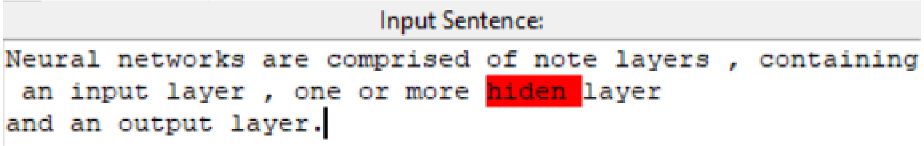


Suggested Words:

a (frequency: 24)
up (frequency: 13)
in (frequency: 3)
the (frequency: 3)
times (frequency: 3)

In this example, ‘pythone’ was correctly detected by the system as non-word error and provided the correct words suggestion. Meanwhile, the system also detected ‘programming’ as the real-word error due to the low bigram frequency of ‘level programming’ and ‘programming language’ in the corpus.

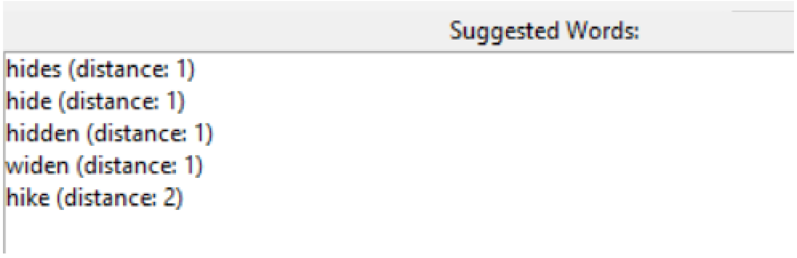
Case 2:



Input Sentence:

```
Neural networks are comprised of note layers , containing
 an input layer , one or more hiden layer
and an output layer.
```

Non-words error:

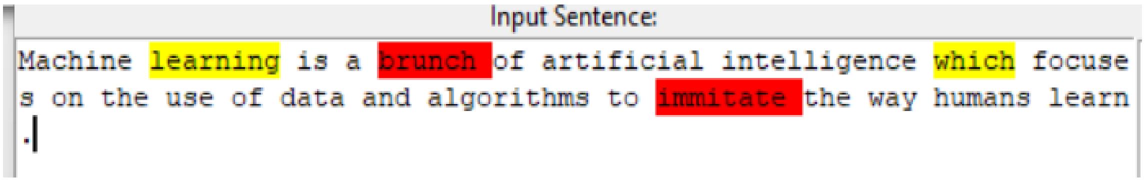


Suggested Words:

hides (distance: 1)
hide (distance: 1)
hidden (distance: 1)
widen (distance: 1)
hike (distance: 2)

In this example, the system is able to capture the only non-words spelling error and the expected word ‘hidden’ was also suggested by the system for correction. However, the phrase ‘node layers’ was missed.

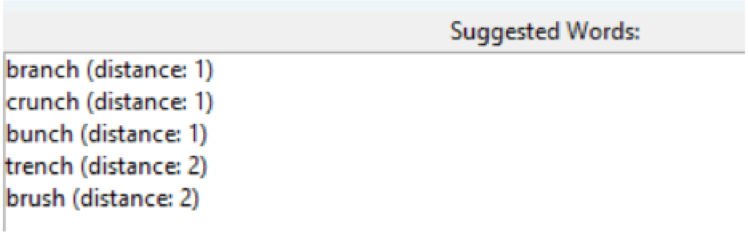
Case 3:



Input Sentence:

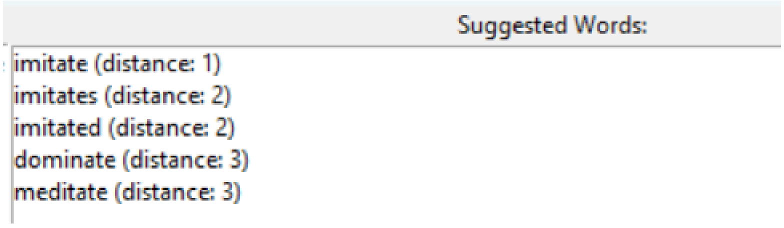
```
Machine learning is a brunch of artificial intelligence which focuse
s on the use of data and algorithms to immitate the way humans learn
.
```

Non-words error:



Suggested Words:

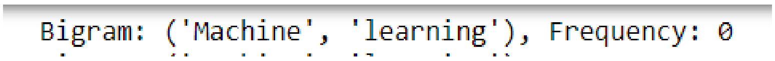
branch (distance: 1)
crunch (distance: 1)
bunch (distance: 1)
trench (distance: 2)
brush (distance: 2)



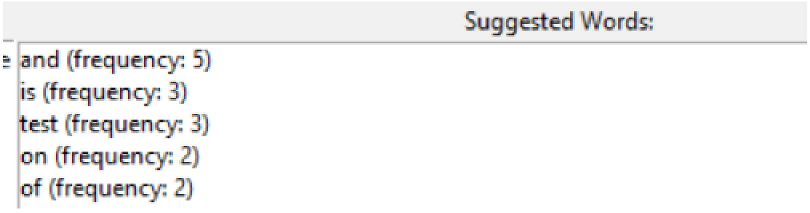
Suggested Words:

imitate (distance: 1)
imitates (distance: 2)
imitated (distance: 2)
dominate (distance: 3)
meditate (distance: 3)

Real-words error:

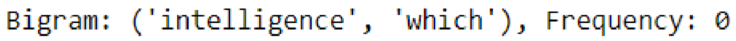


```
 Bigram: ('Machine', 'learning'), Frequency: 0
```

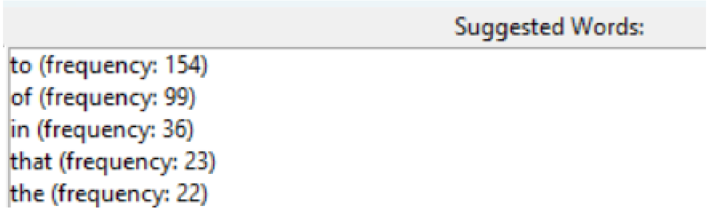


Suggested Words:

```
and (frequency: 5)
is (frequency: 3)
test (frequency: 3)
on (frequency: 2)
of (frequency: 2)
```



```
Bigram: ('intelligence', 'which'), Frequency: 0
```

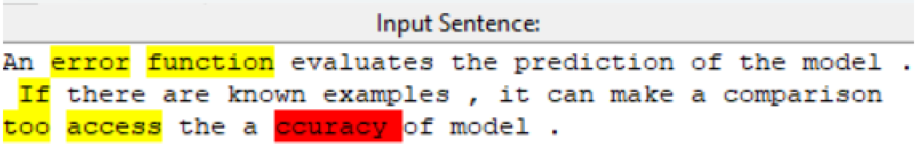


Suggested Words:

```
to (frequency: 154)
of (frequency: 99)
in (frequency: 36)
that (frequency: 23)
the (frequency: 22)
```

Two non-words error were correctly detected by the system with correct words suggestion. However, due to the low bigram frequency of ‘machine learning’ and ‘intelligence which’, the system detected ‘learning’ and ‘which’ as real-words error.

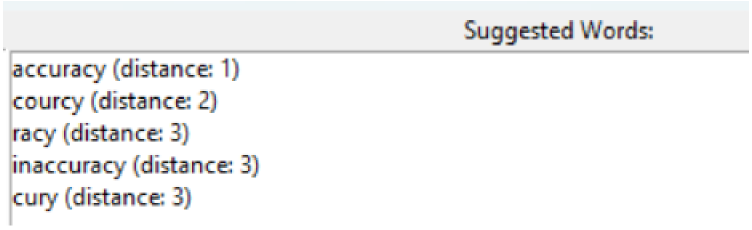
Case 4:



Input Sentence:

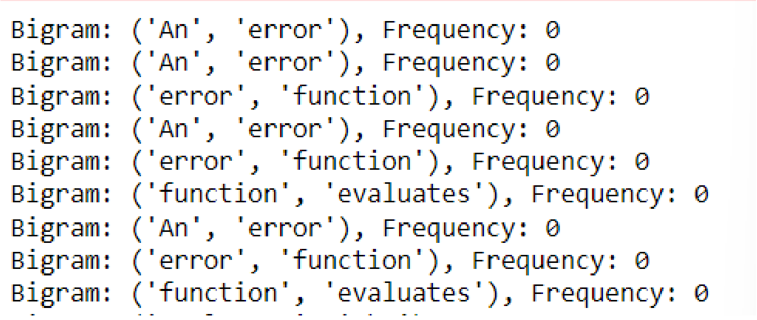
```
An error function evaluates the prediction of the model .
 If there are known examples , it can make a comparison
too access the a ccuracy of model .
```

Non-words error:

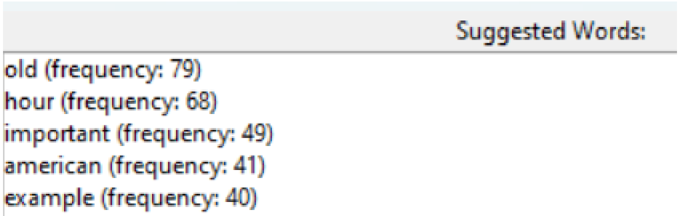


Suggested Words:

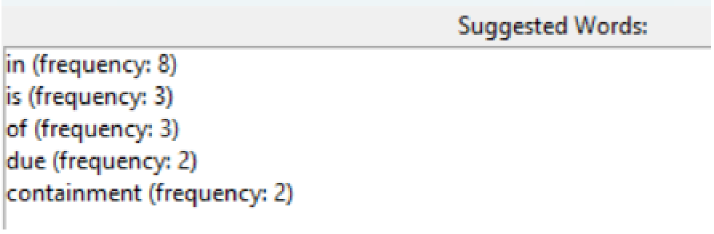
```
accuracy (distance: 1)
courcy (distance: 2)
racy (distance: 3)
inaccuracy (distance: 3)
cury (distance: 3)
```



```
Bigram: ('An', 'error'), Frequency: 0
Bigram: ('An', 'error'), Frequency: 0
Bigram: ('error', 'function'), Frequency: 0
Bigram: ('An', 'error'), Frequency: 0
Bigram: ('error', 'function'), Frequency: 0
Bigram: ('function', 'evaluates'), Frequency: 0
Bigram: ('An', 'error'), Frequency: 0
Bigram: ('error', 'function'), Frequency: 0
Bigram: ('function', 'evaluates'), Frequency: 0
```

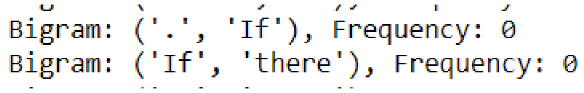

Suggested Words:

```
old (frequency: 79)
hour (frequency: 68)
important (frequency: 49)
american (frequency: 41)
example (frequency: 40)
```

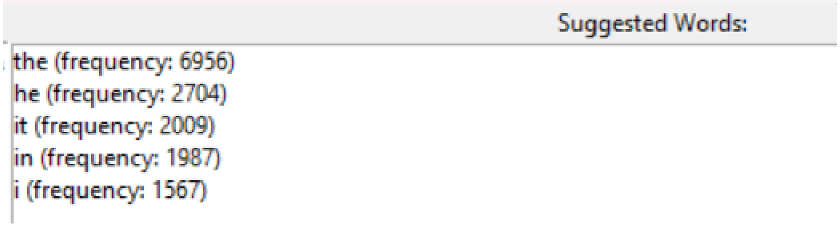

Suggested Words:

```
in (frequency: 8)
is (frequency: 3)
of (frequency: 3)
due (frequency: 2)
containment (frequency: 2)
```

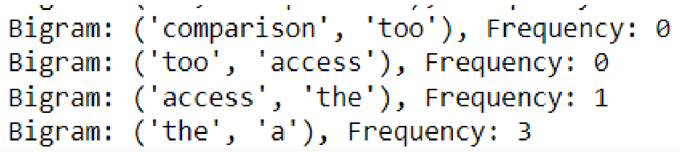
Real-words error:



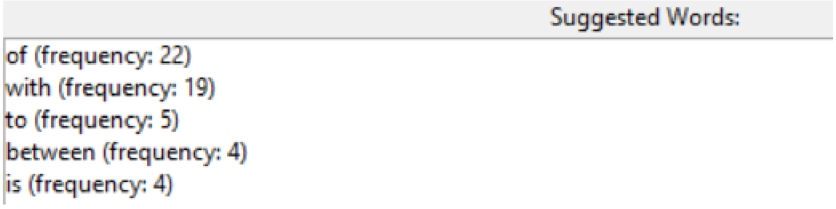
```
Bigram: ('.', 'If'), Frequency: 0
Bigram: ('If', 'there'), Frequency: 0
```


Suggested Words:

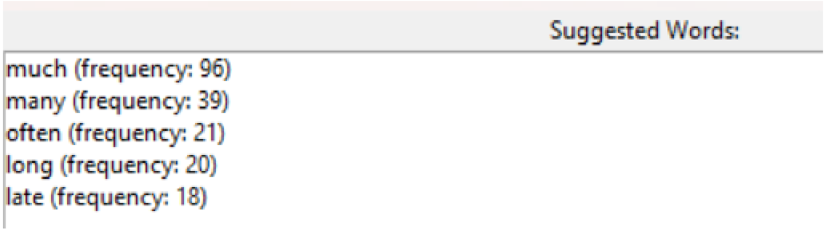
```
the (frequency: 6956)
he (frequency: 2704)
it (frequency: 2009)
in (frequency: 1987)
i (frequency: 1567)
```



```
Bigram: ('comparison', 'too'), Frequency: 0
Bigram: ('too', 'access'), Frequency: 0
Bigram: ('access', 'the'), Frequency: 1
Bigram: ('the', 'a'), Frequency: 3
```


Suggested Words:

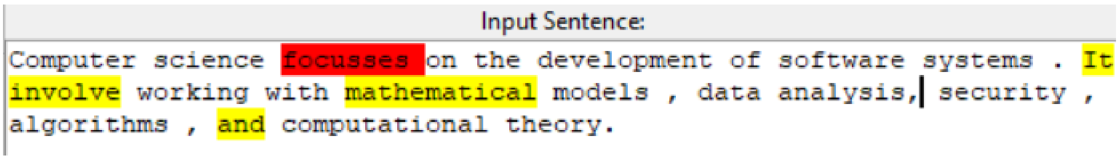
```
of (frequency: 22)
with (frequency: 19)
to (frequency: 5)
between (frequency: 4)
is (frequency: 4)
```


Suggested Words:

```
much (frequency: 96)
many (frequency: 39)
often (frequency: 21)
long (frequency: 20)
late (frequency: 18)
```

In this sentence, one non-word spelling error was detected by the system and the expected word ‘accuracy’ was provided as the first suggested word. ‘Error function, if’ was detected as a real-word spelling error as these phrases are not found in the corpus. Meanwhile, the system also correctly detected ‘too access’ as real-word error and the correct word was among the suggested words shown.

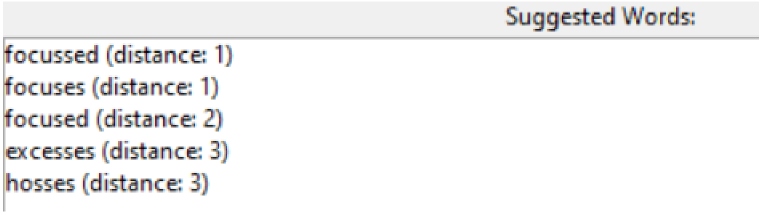
Case 5:



Input Sentence:

```
Computer science focusses on the development of software systems . It
involve working with mathematical models , data analysis, security ,
algorithms , and computational theory.
```

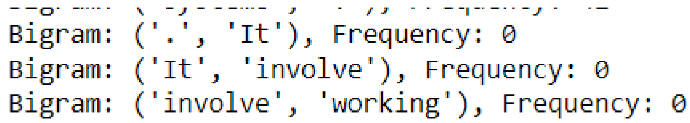
Non-words error:



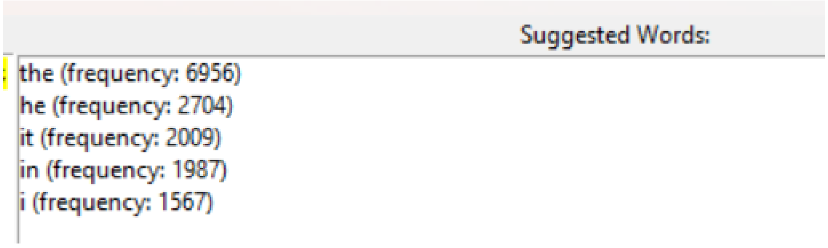
Suggested Words:

focussed (distance: 1)
focuses (distance: 1)
focused (distance: 2)
excesses (distance: 3)
hosses (distance: 3)

Real-words error:

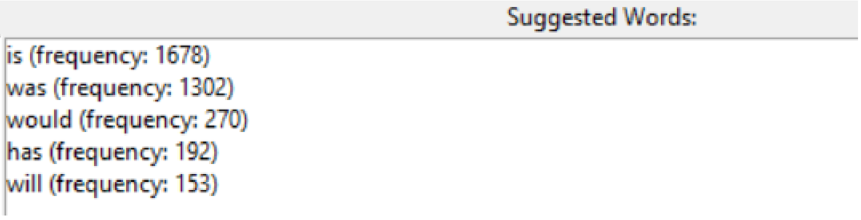


```
Bigram: ('.', 'It'), Frequency: 0
Bigram: ('It', 'involve'), Frequency: 0
Bigram: ('involve', 'working'), Frequency: 0
```



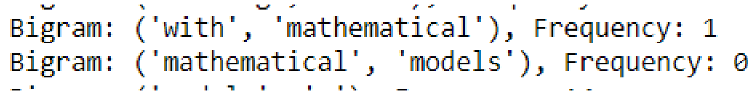
Suggested Words:

the (frequency: 6956)
he (frequency: 2704)
it (frequency: 2009)
in (frequency: 1987)
i (frequency: 1567)

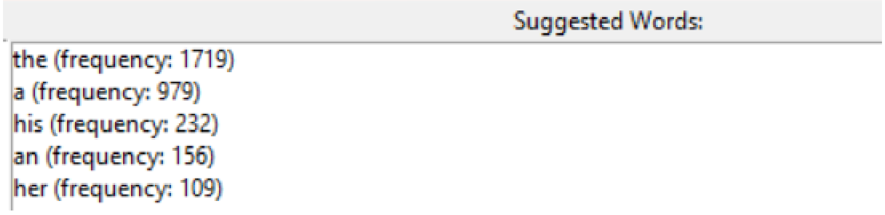


Suggested Words:

is (frequency: 1678)
was (frequency: 1302)
would (frequency: 270)
has (frequency: 192)
will (frequency: 153)



```
Bigram: ('with', 'mathematical'), Frequency: 1
Bigram: ('mathematical', 'models'), Frequency: 0
```

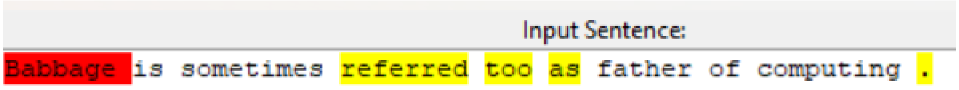


Suggested Words:

the (frequency: 1719)
a (frequency: 979)
his (frequency: 232)
an (frequency: 156)
her (frequency: 109)

In this sentence, one non-word spelling error was detected by the system and the expected word ‘focuses’ was among the suggested words. Meanwhile, the system also correctly detected ‘it involve’ as real-word error in addition to other two phrases which was mistakenly detected due to the low appearance of these phrases in the corpus.

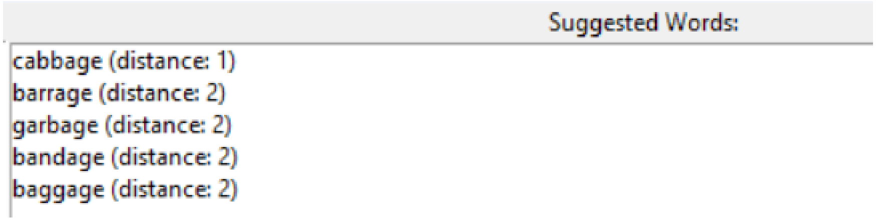
Case 6:



Input Sentence:

```
Babbage is sometimes referred too as father of computing .
```

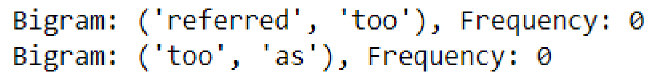
Non-words error:



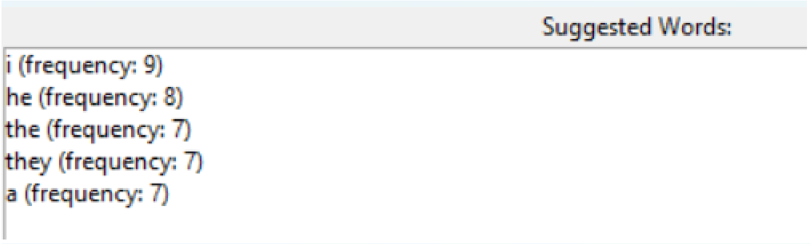
Suggested Words:

cabbage (distance: 1)
barrage (distance: 2)
garbage (distance: 2)
bandage (distance: 2)
baggage (distance: 2)

Real-words error:

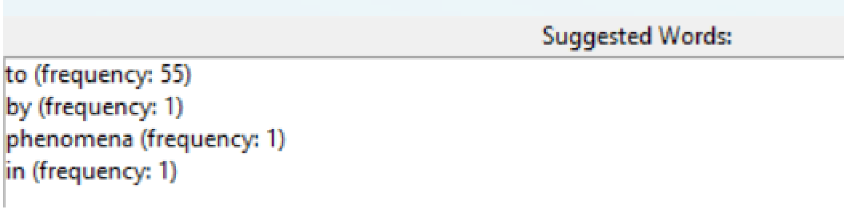


```
Bigram: ('referred', 'too'), Frequency: 0
Bigram: ('too', 'as'), Frequency: 0
```



Suggested Words:

i (frequency: 9)
he (frequency: 8)
the (frequency: 7)
they (frequency: 7)
a (frequency: 7)

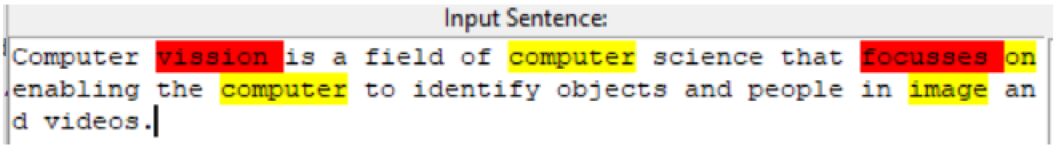


Suggested Words:

to (frequency: 55)
by (frequency: 1)
phenomena (frequency: 1)
in (frequency: 1)

In this sentence, an individual's name was detected as non-word spelling error. Besides that, the system has correctly detected real-word spelling error 'referred too as' and suggested the expected correct word 'to'.

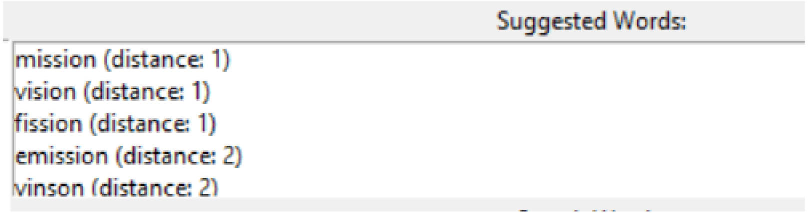
Case 7:



Input Sentence:

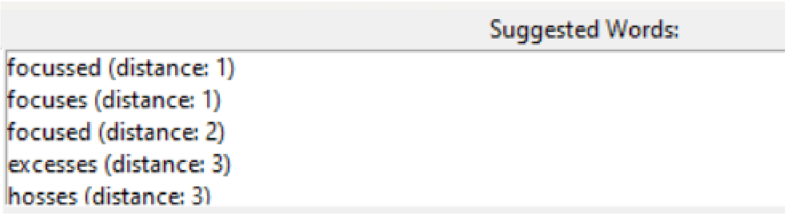
Computer vission is a field of computer science that focusses on
enabling the computer to identify objects and people in image an
d videos.

Non-words error:



Suggested Words:

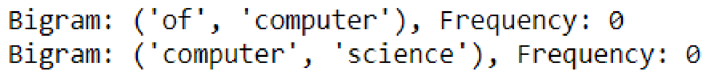
mission (distance: 1)
vision (distance: 1)
fission (distance: 1)
emission (distance: 2)
vinson (distance: 2)



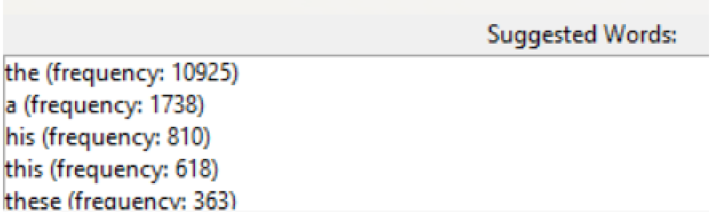
Suggested Words:

focussed (distance: 1)
focuses (distance: 1)
focused (distance: 2)
excesses (distance: 3)
hosses (distance: 3)

Real-words error:

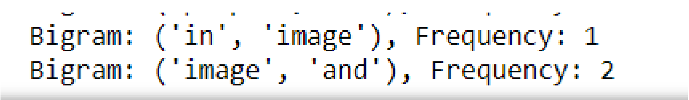


```
Bigram: ('of', 'computer'), Frequency: 0
Bigram: ('computer', 'science'), Frequency: 0
```

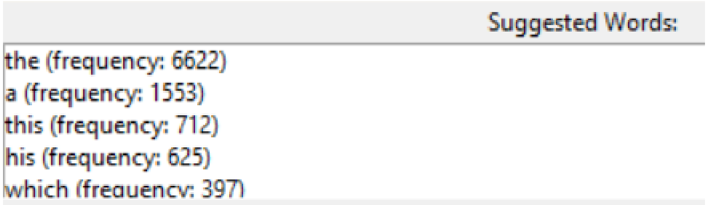


Suggested Words:

the (frequency: 10925)
a (frequency: 1738)
his (frequency: 810)
this (frequency: 618)
these (frequency: 363)



```
Bigram: ('in', 'image'), Frequency: 1
Bigram: ('image', 'and'), Frequency: 2
```

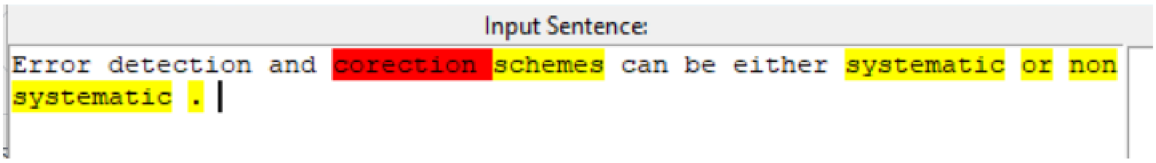


Suggested Words:

the (frequency: 6622)
a (frequency: 1553)
this (frequency: 712)
his (frequency: 625)
which (frequency: 397)

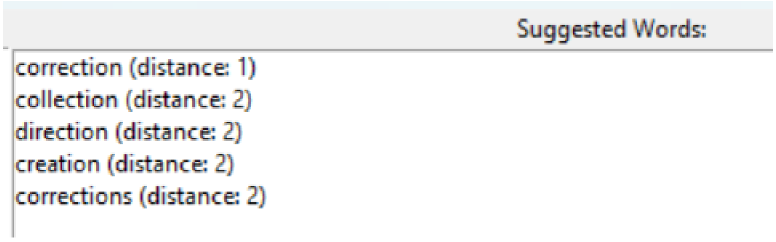
Two non-words error were correctly detected by the system with correct words suggestion. Meanwhile, the system detected mistakenly three real-words error due to the low bigram frequency.

Case 8:



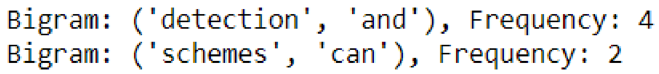
Error detection and corection schemes can be either systematic or non systematic .

Non-word error:

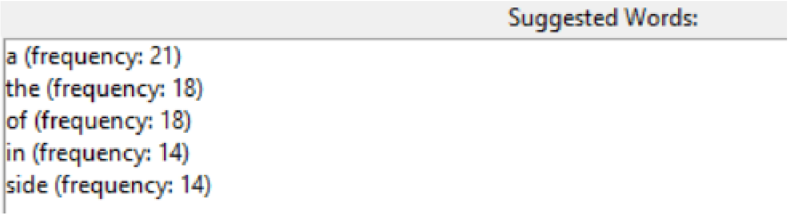


correction (distance: 1)
collection (distance: 2)
direction (distance: 2)
creation (distance: 2)
corrections (distance: 2)

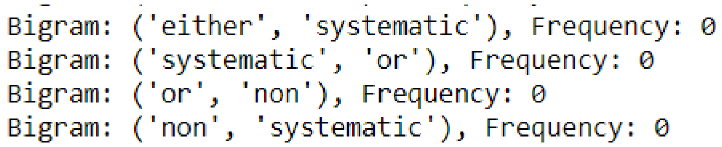
Real-word error:



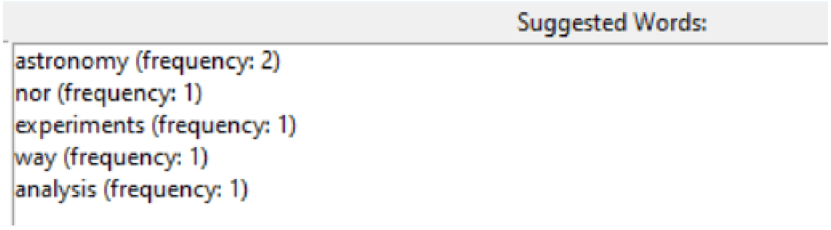
```
Bigram: ('detection', 'and'), Frequency: 4
Bigram: ('schemes', 'can'), Frequency: 2
```



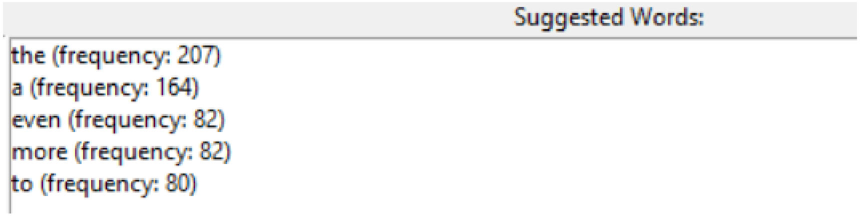
a (frequency: 21)
the (frequency: 18)
of (frequency: 18)
in (frequency: 14)
side (frequency: 14)



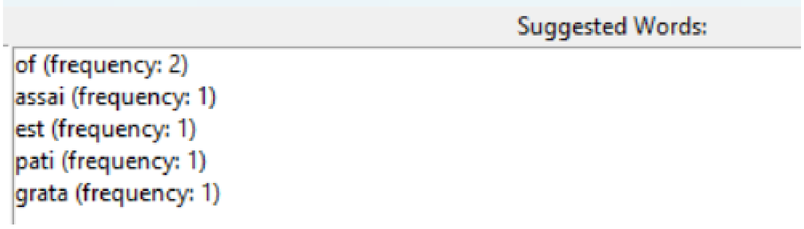
```
Bigram: ('either', 'systematic'), Frequency: 0
Bigram: ('systematic', 'or'), Frequency: 0
Bigram: ('or', 'non'), Frequency: 0
Bigram: ('non', 'systematic'), Frequency: 0
```



astronomy (frequency: 2)
nor (frequency: 1)
experiments (frequency: 1)
way (frequency: 1)
analysis (frequency: 1)



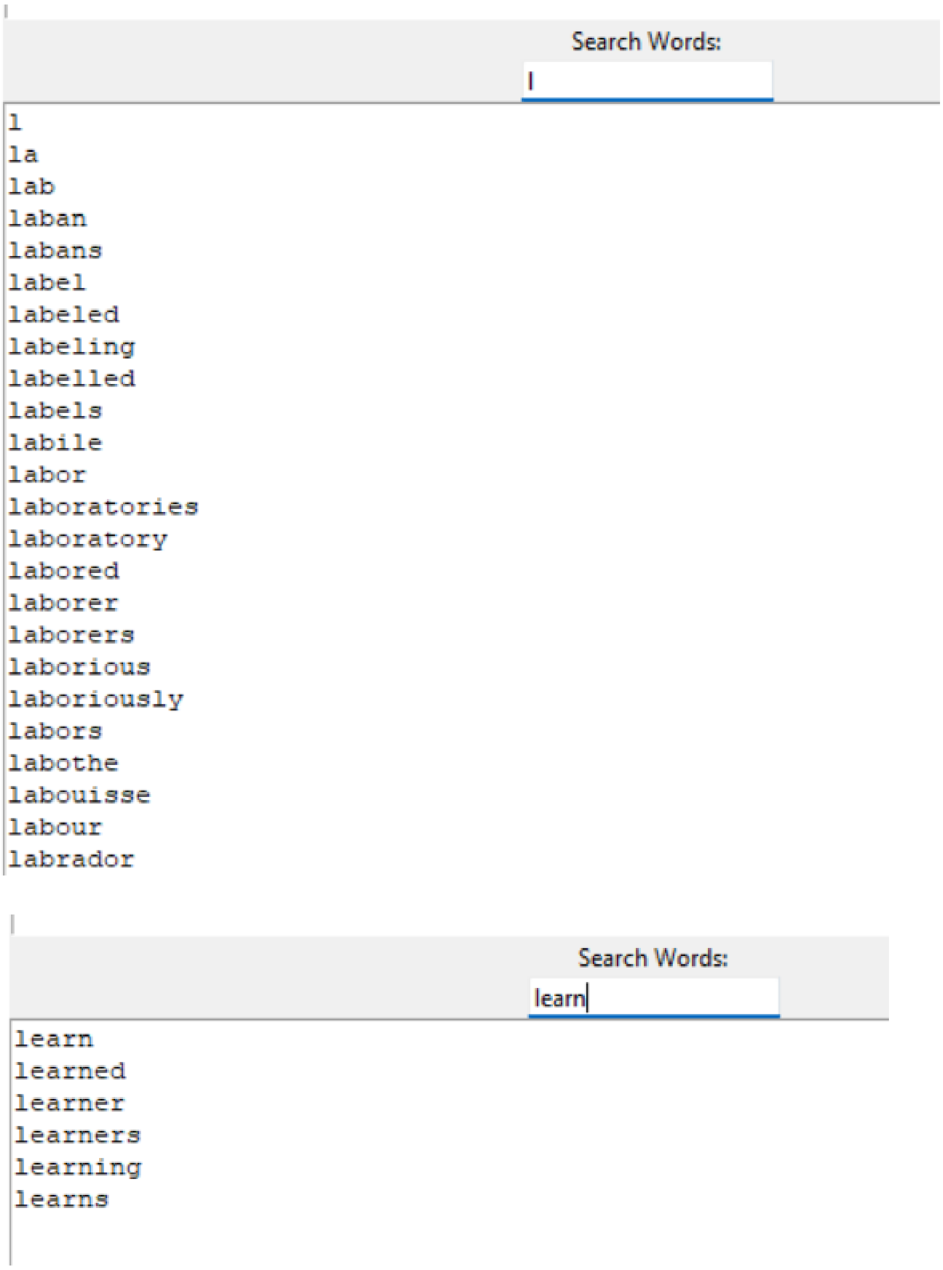
the (frequency: 207)
a (frequency: 164)
even (frequency: 82)
more (frequency: 82)
to (frequency: 80)



of (frequency: 2)
assai (frequency: 1)
est (frequency: 1)
pati (frequency: 1)
grata (frequency: 1)

In this example, one non-word error was correctly detected with expected correct word provided. However, due to the low bigram frequency of ‘correction schemes, systemic or non-systemic’, these phrases were also detected as real-words error.

Search Words



Search Words: l

```
l
la
lab
laban
labans
label
labeled
labeling
labelled
labels
labile
labor
laboratories
laboratory
labored
laborer
laborers
laborious
laboriously
labors
labothe
labouisse
labour
labrador
```

Search Words: learn

```
learn
learned
learner
learners
learning
learns
```

Figure 1 Search Words Function

Besides error detection, this system is embedded with a search words function. The user can pass in one letter and a list of possible words will be displayed. As more letters are entered, the displayed words become more specific. In this example, when the word ‘learn’ was entered in the search bar, all possible words derived from the root word ‘learn’ are displayed.

Overall, the proposed system is capable in detecting all non-words error. However, due to the corpus size used which consists of 41733 unique words, 1095516 bi-gram, the system is sensitive in capturing real-words error as some of the phrases do not exist in the corpus.