AIM: NLP processing of on English Language & Hindi Language

WORD ANALYSIS:

Word analysis is a process of learning more about word meanings by studying their origins and parts.
A “morpheme” is the smallest meaningful part of a word.
Other terms for word analysis: Morphemic analysis & Word study

Morphological analysis is the process of examining possible resolutions to unquantifiable, complex problems involving many factors. The root of the word morphology comes from the Greek word, morphe, for form.
Morphological analysis takes a problem with many known solutions and breaks them down into their most basic elements, or forms, in order to more completely understand them.
Morphological analysis is used in general problem solving, linguistics and biology. In many fields of study morphology facilitates clearer instruction for teachers to help students understand problems and their solutions.
For general problem solving, morphological analysis provides a formalized structure to help examine the problem and possible solutions. The elements of a problem and its solutions are arranged in a matrix to help eliminate illogical solutions.
In biology, the study of forms helps understand mutations, adaptation and evolution. The study of the features and structure of organisms helps us understand organisms and their place in the greater environment.
In linguistics, words are broken down into the smallest units of meaning: morphemes.
Morphemes can sometimes be words themselves as in the case of free morphemes, which can stand on their own. Other morphemes can add meaning but not stand as words on their own; bound morphemes need to be used along with another morpheme to make a word. Cats, for example, is a two-morpheme word. Its base, cat, is a free morpheme and its suffix an s, to denote pluralization, a bound morpheme.

SOURCE CODE:

```
wordListE=[{"word":"boy","rt":"boy","cat":"n","gen":"m","num":"sg","case":"","per":"","tense":"","aspect":""},

{"word":"boys","rt":"boy","cat":"n","gen":"m","num":"pl","case":"","per":"","tense":"","aspect":""},

{"word":"toys","rt":"toy","cat":"n","gen":"","num":"pl","case":"","per":"3","tense":"","aspect":""}]
```

```
wordListH=[{"word":"लड़के","rt":"लड़का","cat":"n","gen":"m","num":"sg","case":"obl","per":"",
"tense":"","aspect":"",},

{"word":"लड़के","rt":"लड़का","cat":"n","gen":"m","num":"pl","case":"dir","per":"","tense":"","as
pect":""},

{"word":"लड़को","rt":"लड़का","cat":"n","gen":"m","num":"pl","case":"obl","per":"","tense":"","a
spect":""},

{"word":"हंसी","rt":"हंस","cat":"v","gen":"fem","num":"sg/pl","case":"obl","per":"1/2/3","tense
":"past","aspect":"pft"},

{"word":"हंसी","rt":"हंसी","cat":"n","gen":"","num":"pl","case":"obl","per":"","tense":"","aspect
":""},
          ]

def processEng(word):
  for x in wordListE:
    if word==x["word"]:
      if x["cat"]=="v":
        print("Root of word is", x['rt'],"category is verb ","gender is ",x['gen'],"number is ",
x['num'], "tense is ", x['tense'],"person is", x['per'], "aspect is", x['aspect'])
      else:
        print("Root of word is", x['rt'],"category is noun ","gender is ",x['gen'],"number is ",
x['num'])

def processHindi(word):
  for x in wordListH:
    if word==x["word"]:
      if x["cat"]=="v":
        print("Root of word is", x['rt'],"category is verb ","gender is ",x['gen'],"number is ",
x['num'], "tense is ", x['tense'],"person is", x['per'], "aspect is", x['aspect'])
      else:
        print("Root of word is", x['rt'],"category is noun ","gender is ",x['gen'],"number is ",
x['num'], "case is ", x['case'])

wordLang = input("Enter E for English or H for Hindi ")
```

```
print("You Selected ", wordLang)
inputWord = input("Enter the word ")
wordlang = wordLang.lower()
if wordLang == "e":
  processEng(inputWord)
elif wordLang == "h":
  processHindi(inputWord)
else:
  print("Wrong Choice Selected")
```

OUTPUT:

```
Enter the word toys
Root of word is toy category is noun  gender is   number is  pl

Enter the word boys
Root of word is boy category is noun  gender is  m number is  pl

Enter the word boy
Root of word is boy category is noun  gender is  m number is  sg

Enter E for English or H for Hindi h
You Selected  h
Enter the word लड़को
Root of word is लड़का category is noun  gender is  m number is  pl case is  obl
```

TEXT ANALYSIS:

Text analysis (TA) is a machine learning technique used to automatically extract valuable insights from unstructured text data. Companies use text analysis tools to quickly digest online data and documents, and transform them into actionable insights.

Text analysis (also known as Text Mining or Content Analysis) is a technique that computers use to extract worthwhile information from the human language in a smart and efficient manner. Researchers and developers can use this method to assemble diverse and unorganized data in a structured form. In this process, documents are disintegrated for hassle-free management of data pieces, simply put: unstructured text gets converted into structured data.
Once the sentences are put to paper, the integral sections such as the proper nouns are branched off using the list of words in the dictionary. The analysis of text, those in either

documents or graphics, qualitative details can be transformed into quantitative details. It's safe to establish text analysis is a research approach adapted to put together reasonable conclusions by decoding the content.
Developers and researchers use text analysis for establishing correlation between two entities, automated summary generation, translation, speech recognition and other similar content management tasks that come their way. If we talk about the business aspect, text analysis offers many wider subjects such as semantic search and content management to gather information.

SOURCE CODE:

```
text = "This is my test text. We're keeping this text short to keep things manageable."
text = text.lower()

from collections import Counter

# counts word frequency
def count_words(text):
  skips = [".", ",", "", ":", ";", "'", '"']
  for ch in skips:
    text = text.replace(ch, "")
  word_counts = {}
  for word in text.split(" "):
    if word in word_counts:
      word_counts[word]+= 1
    else:
      word_counts[word]= 1
  return word_counts

def count_words_fast(text):
  text = text.lower()
  skips = [".", ",", "", ":", ";", "'", '"']
  for ch in skips:
    text = text.replace(ch, "")
  word_counts = Counter(text.split(" "))
  return word_counts


count_words(text)
count_words_fast(text)
```

OUTPUT:

```
Counter({'is': 1,
         'keep': 1,
         'keeping': 1,
         'manageable': 1,
         'my': 1,
         'short': 1,
         'test': 1,
         'text': 2,
         'things': 1,
         'this': 2,
         'to': 1,
         'were': 1})
```

CONCLUSION:

From this tutorial, I have learned about word analysis and text analysis.