



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Aim: Perform morphological analysis and word generation for any given text.

Objective:

To study morphological analysis.

Theory:

Morphological analysis is a field of linguistics that studies the structure of words. It identifies how a word is produced through the use of morphemes. A morpheme is a basic unit of the English language. The morpheme is the smallest element of a word that has grammatical function and meaning. In inflected languages, words are formed through morphological processes such as affixation. For example, by adding the suffix '-s' to the verb 'to dance', we form the third person singular 'dances'.

Parsing:

It is the process of determining the morphemes from which a given word is constructed. Morphemes are the smallest meaningful words which cannot be divided further. Morphemes can be stem or affix. Stem are the root word whereas affix can be prefix, suffix or infix. For example-

Unsuccessfull → un success ful

(prefix) (stem) (suffix)

Program:

```
import nltk

from nltk.stem import WordNetLemmatizer

nltk.download('wordnet')

lemmatizer = WordNetLemmatizer()

text = "The quick brown foxes are jumping over the lazy dogs"

words = nltk.word_tokenize(text)

lemmatized_words = [lemmatizer.lemmatize(word) for word in words]
```



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

```
generated_words = []  
for word in lemmatized_words:  
    plural_form = word + 's'  
    generated_words.append(plural_form)  
  
print("Original words:", words)  
print("Lemmatized words:", lemmatized_words)  
print("Generated words:", generated_words)
```

Output:

Original words: ['The', 'quick', 'brown', 'foxes', 'are', 'jumping', 'over', 'the', 'lazy', 'dogs']

Lemmatized words: ['The', 'quick', 'brown', 'fox', 'are', 'jumping', 'over', 'the', 'lazy', 'dog']

Generated words: ['Thes', 'quicks', 'browns', 'foxs', 'ares', 'jumpings', 'overs', 'thes', 'lazys', 'dogs']

Conclusion: Morphological analysis is of paramount importance in natural language processing (NLP) and computational linguistics. It serves as the foundational step for understanding the structure of words in a language, allowing NLP models to handle variations of words, such as tense, plurals, and derivational forms, effectively. Accurate morphological analysis aids in tasks like text normalization, information retrieval, machine translation, and sentiment analysis. Moreover, it contributes to better language understanding, enabling applications to comprehend and generate human language with greater precision, which is crucial in today's data-driven and language-centric technological landscape.