



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Batch:-B-2 Roll No:-16010122151

Experiment No. 10

Title : Virtual lab on text processing (word analysis)

Aim: To implement text processing (Word analysis)

Expected Outcome of Experiment:

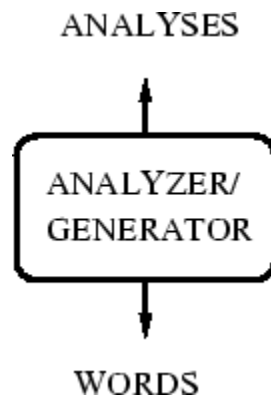
CO4 : Understand the basic concept and techniques of Machine Learning clustering.

Books/ Journals/ Websites referred:

1. <https://nlp-iiith.vlabs.ac.in/exp/word-analysis/index.html>

Introduction

A word can be simple or complex. For example, the word 'cat' is simple because one cannot further decompose the word into smaller part. On the other hand, the word 'cats' is complex, because the word is made up of two parts: root 'cat' and plural suffix '-s'



Procedure

Simulation:-

- 1) Play:-



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs

Word Analysis

Select a Language which you know better

English

Select a word from the below dropbox and do a morphological analysis on that word

play

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	play	
ROOT	play	✓
CATEGORY	verb	✓
GENDER	na	✓
NUMBER	singular	✓
PERSON	na	✓
CASE	na	✓
TENSE	na	✓
Check	Right answer!!!	

2) Plays

Virtual Labs

Word Analysis

Select a Language which you know better

English

Select a word from the below dropbox and do a morphological analysis on that word

plays

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	plays	
ROOT	play	✓
CATEGORY	verb	✓
GENDER	na	✓
NUMBER	plural	✓
PERSON	na	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

3) Playing



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual
LELUS
An ISO 9001:2015 Certified

Word Analysis

Select a Language which you know better

English

Select a word from the below dropdown and do a morphological analysis on that word

playing

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	playing	
ROOT	play	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	second	✓
CASE	na	✓
TENSE	present-continuous	✓
Check	Right answer!!!	

4) Played

Virtual
LELUS
An ISO 9001:2015 Certified

Word Analysis

Select a Language which you know better

English

Select a word from the below dropdown and do a morphological analysis on that word

played


Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	played	
ROOT	play	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	plural	✓
PERSON	third	✓
CASE	na	✓
TENSE	past-perfect	✓
Check	Right answer!!!	



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

1) Train



Word Analysis

Select a Language which you know better

English


Select a word from the below dropbox and do a morphological analysis on that word

train

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	train	
ROOT	train	✓
CATEGORY	noun	✓
GENDER	na	✓
NUMBER	singular	✓
PERSON	second	✓
CASE	na	✓
TENSE	na	✓
Check	Right answer!!!	

2) Trains



Word Analysis

Select a Language which you know better

English

Select a word from the below dropbox and do a morphological analysis on that word

trains


Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	trains	
ROOT	train	✓
CATEGORY	noun	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	third	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

3) Training



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

**Word Analysis**

Select a Language which you know better

English


Select a word from the below dropbox and do a morphological analysis on that word

training

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	training	
ROOT	train	✓
CATEGORY	noun	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	first	✓
CASE	na	✓
TENSE	future-perfect-continuous	✓
<div>Check</div>	Right answer!!!	

4) Trained

**Word Analysis**

Select a Language which you know better

English

Select a word from the below dropbox and do a morphological analysis on that word

trained

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)


WORD	trained	
ROOT	train	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	second	✓
CASE	na	✓
TENSE	past-perfect	✓
<div>Check</div>	Right answer!!!	

HINDI



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

1) रेलगाड़ी

Virtual Labs
The World, Open to Your Imagination

Word Analysis

Select a Language which you know better

Hindi


Select a word from the below dropbox and do a morphological analysis on that word

रेलगाड़ी

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	रेलगाड़ी	
ROOT	रेलगाड़ी	✓
CATEGORY	noun	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	na	✓
CASE	direct	✓
TENSE	na	✓
Check	Right answer!!!	

2) रेलगाडडयाँ

Virtual Labs
The World, Open to Your Imagination

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropbox and do a morphological analysis on that word

रेलगाड़ियाँ

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	रेलगाड़ियाँ	
ROOT	रेलगाड़ी	✓
CATEGORY	noun	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	na	✓
CASE	direct	✓
TENSE	na	✓
Check	Right answer!!!	

3) रेलगाडडय के



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

रेलगाड़िया

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	रेलगाड़िया	
ROOT	रेलगाड़ी	✓
CATEGORY	noun	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	na	✓
CASE	oblique	✓
TENSE	na	✓
Check	Right answer!!!	

Hindi Verb

1) पढ़

Virtual Labs

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	plural	✓
PERSON	second	✓
CASE	na	✓
TENSE	present-continuous	✓
Check	Right answer!!!	

2) पढ़ना



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs
An IIT Bombay Initiative

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ना

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ना	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

3) पढ़ता

Virtual Labs
An IIT Bombay Initiative

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ता

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ता	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	first	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

4) पढ़ते



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering



Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropbox and do a morphological analysis on that word

पढ़ते

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़ते	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	plural	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

5) पढ़ती



Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropbox and do a morphological analysis on that word

पढ़ती

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़ती	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	third	✓
CASE	na	✓
TENSE	simple-present	✓
Check	Right answer!!!	

6) पढ़ा



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs
An IEEE Society of India Initiative

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropbox and do a morphological analysis on that word

पढ़ा

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़ा	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	third	✓
CASE	na	✓
TENSE	simple-past	✓
Check	Right answer!!!	

7) पढ़े

Virtual Labs
An IEEE Society of India Initiative

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropbox and do a morphological analysis on that word

पढ़े

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़े	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	third	✓
CASE	na	✓
TENSE	simple-past	✓
Check	Right answer!!!	

8) पढ़ी



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ेगा

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ेगा	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	first	✓
CASE	na	✓
TENSE	simple-past	✓
Check	Right answer!!!	

9) पढ़ेगा

Virtual Labs

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ेगा

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ेगा	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-future	✓
Check	Right answer!!!	

10) पढ़ेगा



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering



Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ेंगा

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ेंगा	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	singular	✓
PERSON	first	✓
CASE	na	✓
TENSE	simple-future	✓
Check		Right answer!!!

11) पढ़ेंगे



Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ेंगे


Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ेंगे	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	first	✓
CASE	na	✓
TENSE	simple-future	✓
Check		Right answer!!!

12) पढ़े



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

**Virtual Labs**
An IIT Bombay initiative

Word Analysis

Select a Language which you know better

Hindi


Select a word from the below dropdown and do a morphological analysis on that word

पढ़ाई

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़ाई	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	male	✓
NUMBER	plural	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-future	✓
Check	Right answer!!!	

13) पढ़ेगी

**Virtual Labs**
An IIT Bombay initiative

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ेगी

Select the Correct morphological analysis for the above word using dropboxes (NOTE : na = not applicable)

WORD	पढ़ेगी	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-future	✓
Check	Right answer!!!	

14) पढ़ाई



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Virtual Labs
An IIT Bombay Center of Excellence

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ाई

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ाई	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	singular	✓
PERSON	first	✓
CASE	na	✓
TENSE	simple-future	✓
<div>Check</div>		Right answer!!!

15) पढ़ गी

Virtual Labs
An IIT Bombay Center of Excellence

Word Analysis

Select a Language which you know better

Hindi

Select a word from the below dropdown and do a morphological analysis on that word

पढ़ाई

Select the Correct morphological analysis for the above word using dropdowns (NOTE : na = not applicable)

WORD	पढ़ाई	
ROOT	पढ़	✓
CATEGORY	verb	✓
GENDER	female	✓
NUMBER	plural	✓
PERSON	second	✓
CASE	na	✓
TENSE	simple-future	✓
<div>Check</div>		Right answer!!!

Post Lab questions

Q.1 List any two real-time applications of NLP?

A)

- **Chatbots/Virtual Assistants:-** NLP is extensively used in chatbots and virtual assistants like Siri, Google Assistant, and chat support systems. These systems interpret user queries or commands, understand the context, and respond accordingly, providing real-time assistance.
- **Sentiment Analysis:-** NLP is used for sentiment analysis in social media monitoring tools and customer feedback analysis systems. By analyzing the sentiment of text data in real-time, businesses can gauge public opinion, customer satisfaction, and market trends, allowing them to make informed decisions promptly.



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Q.2 What are stop words?

A) Stop words are common words that are often filtered out during text preprocessing in Natural Language Processing (NLP) tasks. These words typically occur frequently in a language and don't carry significant meaning on their own. Examples of stop words include "the," "is," "and," "in," "of," "for," etc.

The purpose of removing stop words is to reduce the dimensionality of the data and focus on the more meaningful words that carry the essence of the text. By eliminating stop words, NLP models can focus more on the relevant content and improve accuracy in tasks like text classification, sentiment analysis, and information retrieval. However, the list of stop words may vary depending on the specific task or language being analyzed.

Q.3 List the components of Natural Language Processing.

Here are the key components:

- **Tokenization:-** Breaking down text into smaller units such as words, phrases, or sentences, known as tokens.
- **Text Cleaning/Preprocessing:-** This involves removing noise from the text, which may include tasks like removing punctuation, converting text to lowercase, handling contractions, and removing stop words.
- **Part-of-Speech (POS) Tagging:-** Assigning grammatical tags to words in a sentence, such as noun, verb, adjective, etc.
- **Named Entity Recognition (NER):-** Identifying and categorizing entities within text into predefined categories such as names of people, organizations, locations, dates, etc.
- **Parsing:-** Analyzing the grammatical structure of sentences to understand their syntactic relationships.
- **Semantic Analysis:-** Extracting the meaning of text by understanding the relationships between words and phrases in a given context.
- **Sentiment Analysis:-** Determining the sentiment or opinion expressed in text, whether it's positive, negative, or neutral.
- **Language Modeling:-** Building statistical models that predict the probability of a sequence of words occurring in a given context.
- **Machine Translation:** Translating text from one language to another automatically.
- **Text Generation:-** Creating new text based on patterns learned from existing text data.
- **Information Retrieval:-** Finding relevant documents or information based on user queries.
- **Question Answering:** Automatically generating answers to questions posed in natural language.



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering

Conclusion:-

To concluded, Natural Language Processing (NLP) encompasses a wide range of techniques and components aimed at processing and understanding human language. From tokenization to text generation, NLP involves breaking down text, analyzing its structure, extracting meaning, and performing various language-related tasks. These tasks include sentiment analysis, named entity recognition, machine translation, and more. By leveraging NLP, we can automate tasks involving text, extract valuable insights from textual data, and enable interactions between humans and machines in natural language. As NLP continues to advance, it holds immense potential to revolutionize how we interact with technology, understand language, and extract knowledge from vast amounts of textual data.



K. J. Somaiya College of Engineering, Mumbai-77
(A Constituent College of Somaiya Vidyavihar University)
Department of Electronics and Computer Engineering