**Batch: A - 3 Roll No.: 16014022050**

**Experiment No.: 10**

|  |
| --- |
| **Title: Virtual lab on text processing (word analysis)** |

**Aim:** To implement text processing (Word analysis)

**Expected Outcome of Experiment:**

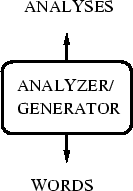
**CO4:** Understand the basic concept and techniques of Machine Learning clustering.

**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:**

**STEP 1**: Select the language.

**OUTPUT**: Drop down for selecting words will appear.

**STEP 2**: Select the word.

**OUTPUT**: Drop down for selecting features will appear.

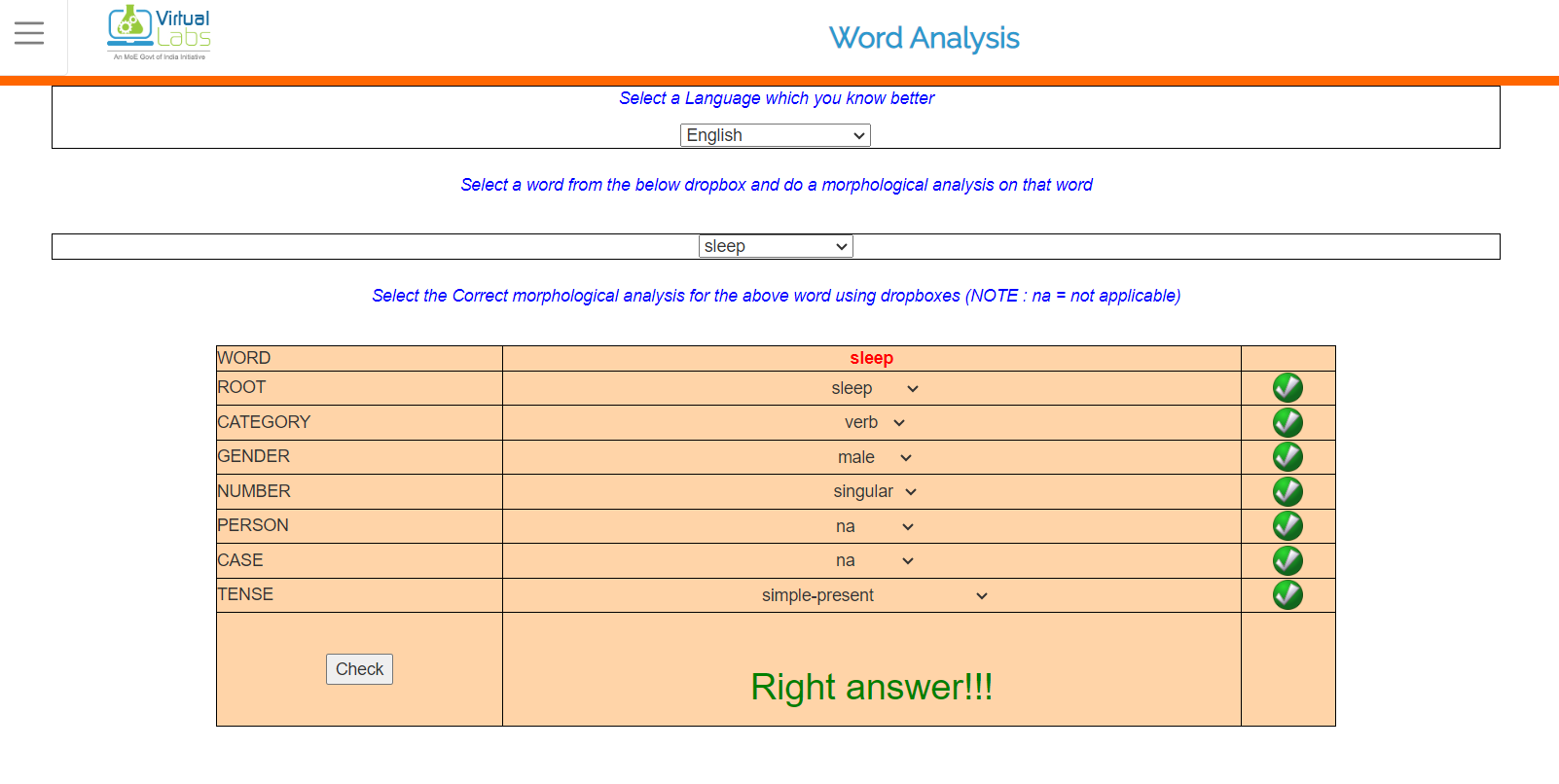
**STEP 3**: Select the features.

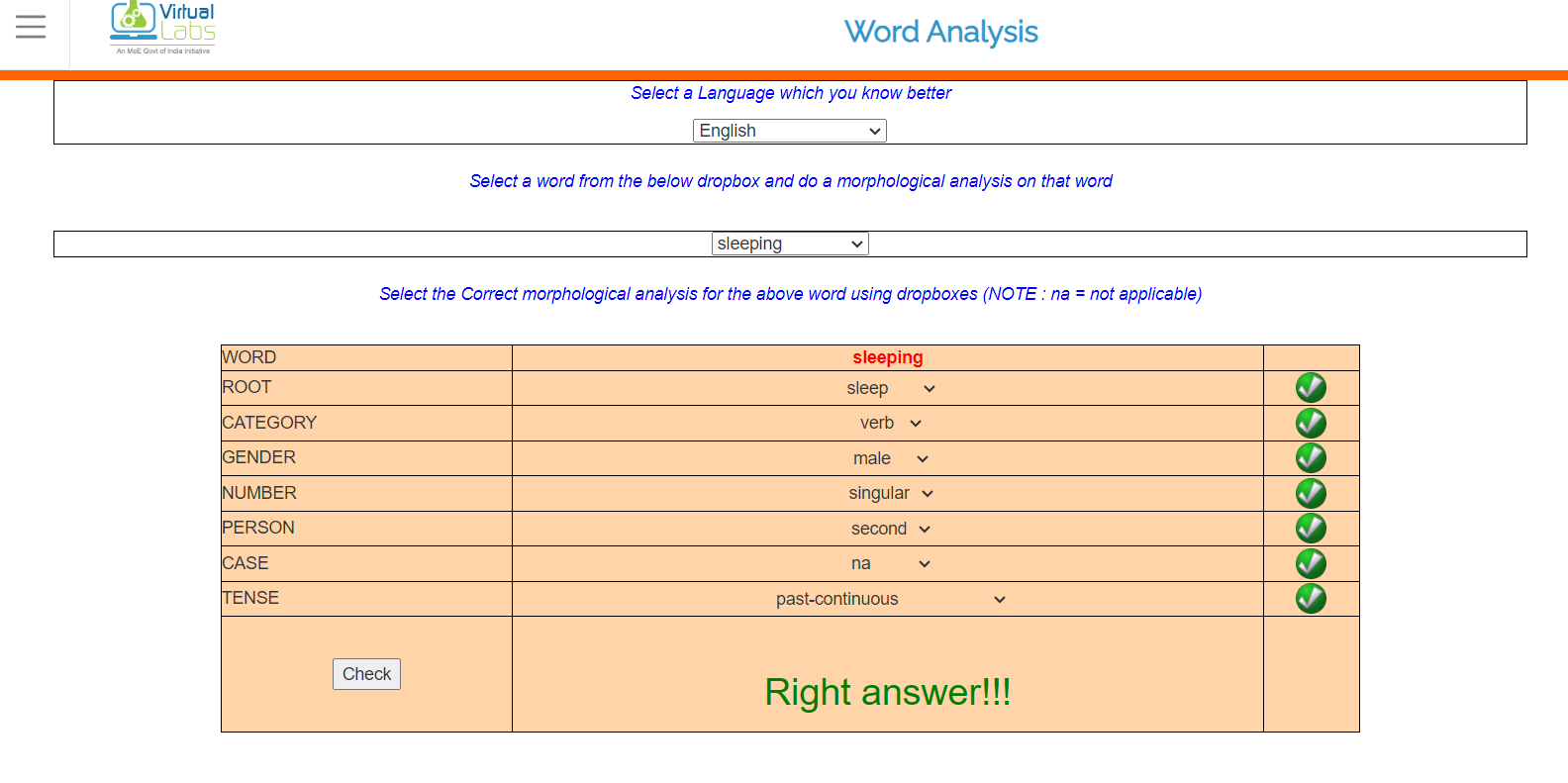
**STEP 4**: Click "Check" button to check your answer.

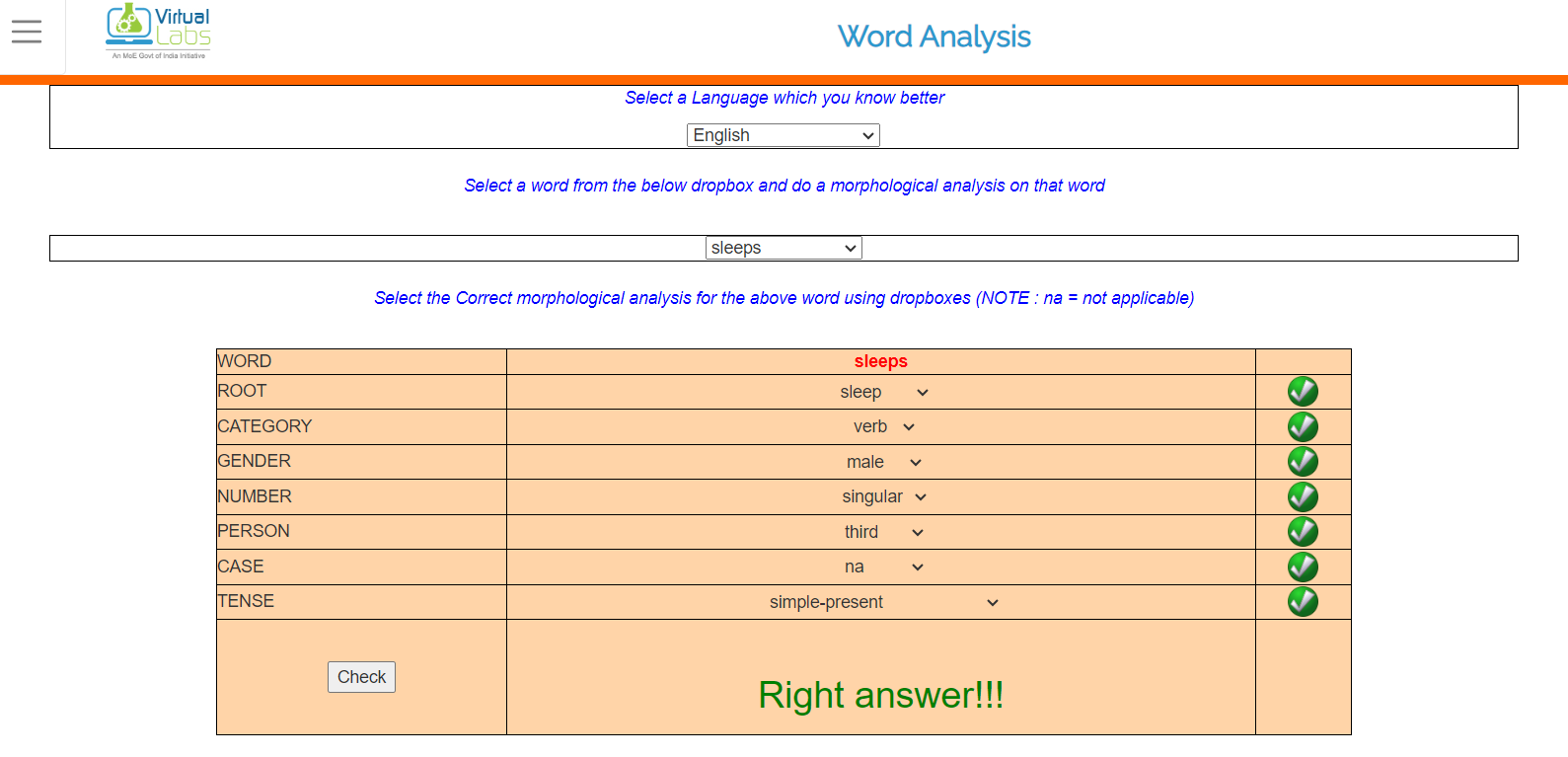
**OUTPUT**: Right features are marked by tick and wrong features are marked by cross.

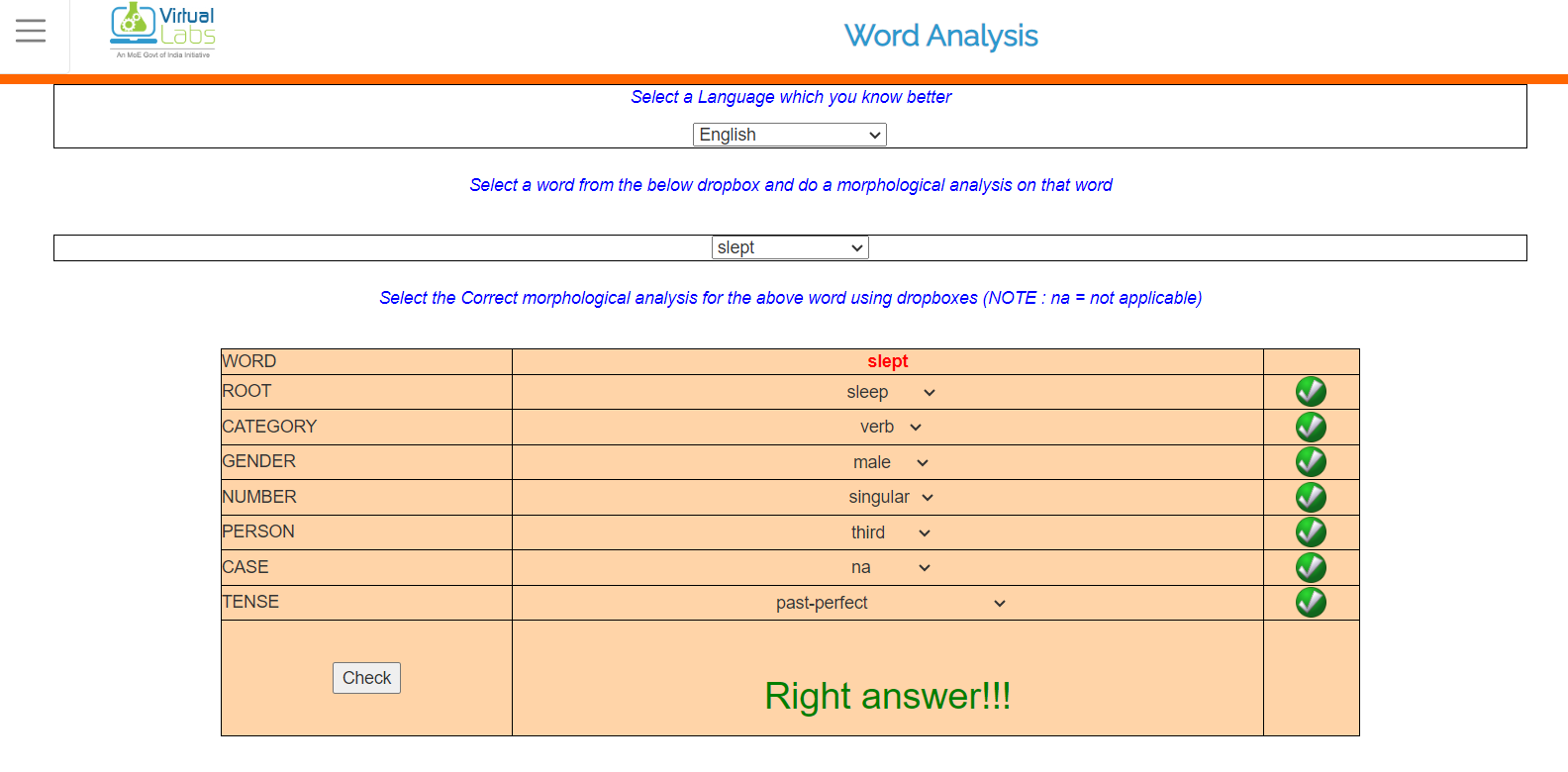
**Simulation:**

**Root word: Sleep**

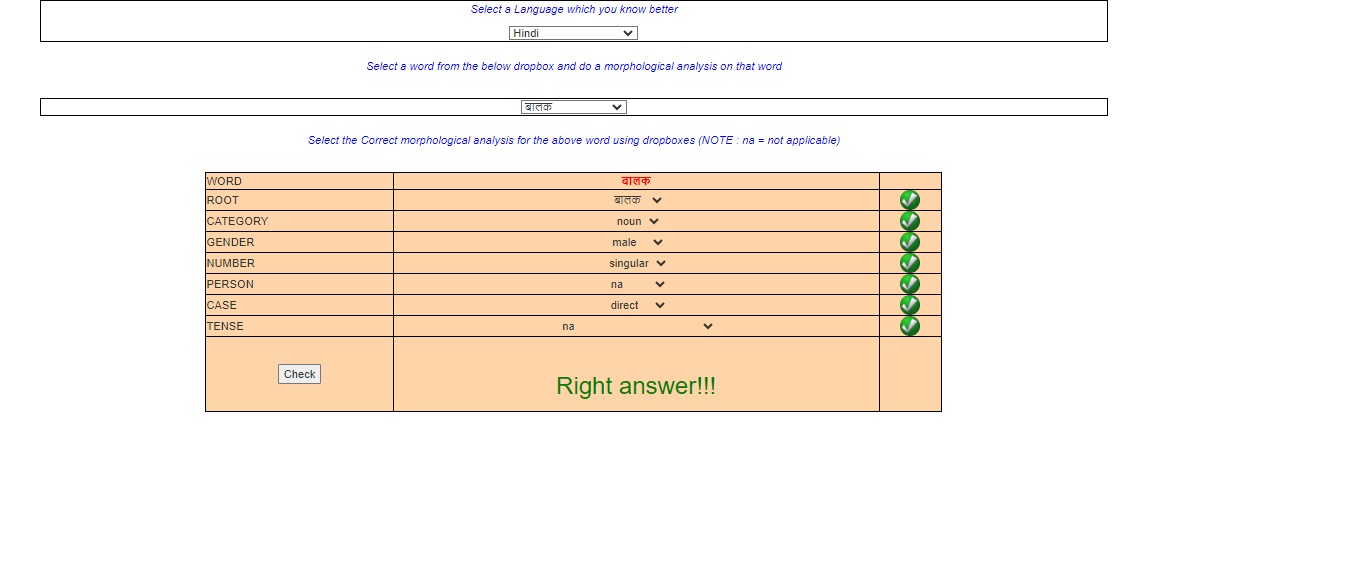


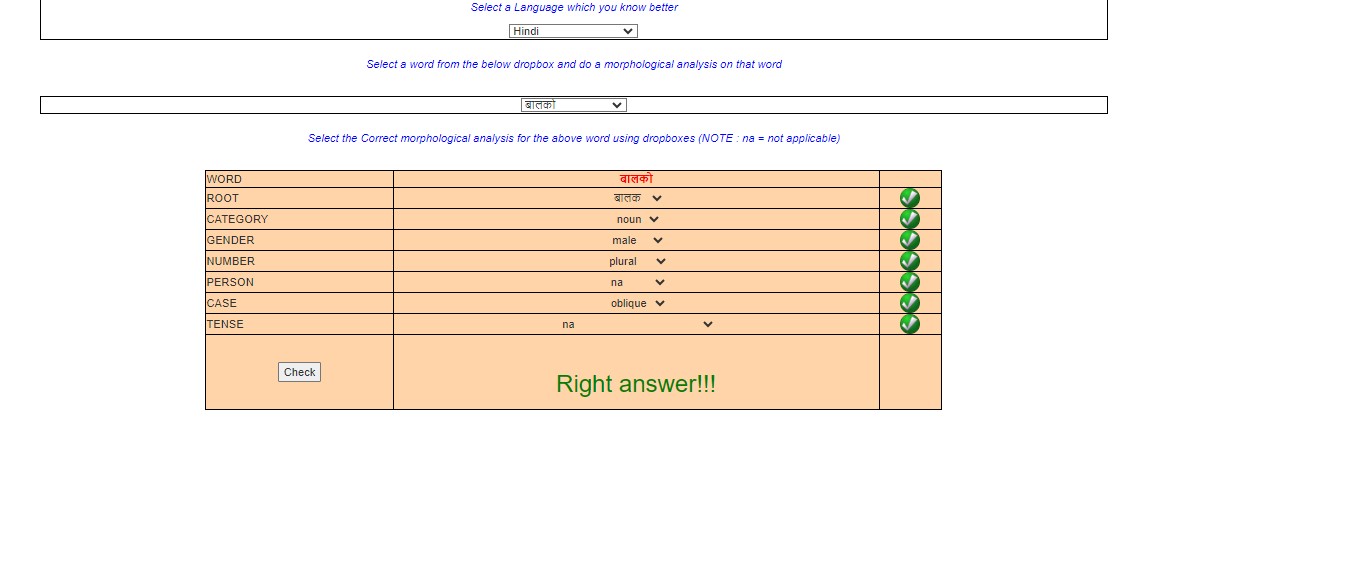






**Root word: बालक**





**Post Lab questions:**

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

Two real-time applications of Natural Language Processing (NLP) are:

1. Chatbots: Chatbots use NLP algorithms to understand and respond to human language in real-time, enabling them to engage in conversations, answer questions, and perform tasks such as customer support or information retrieval.
2. Sentiment Analysis: Sentiment analysis involves analyzing text data to determine the sentiment expressed within it, such as positive, negative, or neutral. This application is used in real-time to monitor social media feeds, customer reviews, or news articles to gauge public opinion or customer satisfaction.
3. **What are stop words?**

Stop words are commonly used words (such as "the," "is," "in," "and," etc.) that are filtered out before or after processing of natural language data. These words are often removed because they occur frequently and do not carry significant meaning in the context of text analysis, such as information retrieval or sentiment analysis.

1. **List the components of Natural Language Processing.**

Components of Natural Language Processing (NLP) typically include:

1. Tokenization: Breaking down text into smaller units, such as words or sentences, known as tokens.
2. Morphological Analysis: Analyzing the structure of words to understand their root forms, prefixes, suffixes, etc.
3. Part-of-Speech Tagging (POS): Assigning grammatical categories to words (e.g., noun, verb, adjective) within a sentence.
4. Named Entity Recognition (NER): Identifying and categorizing entities mentioned in text, such as names of people, organizations, locations, etc.
5. Syntax and Parsing: Analyzing the grammatical structure of sentences to understand relationships between words.
6. Semantic Analysis: Understanding the meaning of words, phrases, and sentences within a given context.
7. Sentiment Analysis: Determining the sentiment expressed in text data, whether it's positive, negative, or neutral.
8. Machine Translation: Translating text from one language to another automatically using computational methods.
9. Question Answering: Developing systems that can understand and answer questions posed in natural language.
10. Text Summarization: Generating concise summaries of longer texts while retaining key information.

**Conclusion:**

We successfully performed the virtual lab experiment and learnt about the morphological features of a word by analysing it.