Batch: B1 Roll No.: 16010121045

Experiment / assignment / tutorial No. 10

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of the Staff In-charge with date

Title: Self Learning Experiment using Vlab in NLP

Objective: To study NLP to introduce concepts like word analysis, word generation and building POS Tagger.

Expected Outcome of Experiment:

| Course Outcome | After successful completion of the course students should be able to |
|-------------------|--|
| | |

Books/ Journals/ Websites referred:

http://nlp-iiith.vlabs.ac.in/List%20 of%20 experiments.html?domain=Computer Science

Pre Lab/ Prior Concepts:

Knowledge Engineering

Historical Profile:

What is Natural Language Processing?

Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages.



New Concepts learnt: (point wise)

- 1. Learnt about types of Morphology
 - **a. Inflectional Morphology -** Deals with word forms of a root, where there is no change in lexical category. For example, 'played' is an inflection of the root word 'play'. Here, both 'played' and 'play' are verbs.
 - **b. Derivational Morphology -** Deals with word forms of a root, where there is a change in the lexical category. For example, the word form 'happiness' is a derivation of the word 'happy'. Here, 'happiness' is a derived noun form of the adjective 'happy'.
- 2. Learnt about various Morphological Features
 - a. All words will have their lexical category attested during morphological analysis.
 - b. A noun and pronoun can take suffixes of the following features: gender, number, person, case
 - c. A verb can take suffixes of the following features: tense, aspect, modality, gender, number, person
 - d. The value of tense can be present, past or future. This feature is applicable for verbs.
 - e. The value of aspect can be perfect (pft), continuous (cont) or habitual (hab). This feature is not applicable for verbs.
- 3. Two broad classes of morphemes: stems (main meaning) and affixes (additional)
- 4. Analysis of a word into root and affix(es) is called as Morphological analysis of a word.
- 5. It is mandatory to identify root of a word for any natural language processing task.
- 6. A root word can have various forms.
- 7. For example, the word 'play' in English has the following forms: 'play', 'plays', 'played' and 'playing'.
- 8. Hindi shows a greater number of forms for the word 'खेल' (khela) which is equivalent to 'play'. The forms of 'खेल'(khela) are the following: खेल(khela), खेला(khelaa), खेली(khelii), खेल्ंगा(kheluungaa), खेल्ंगी(kheluungii), खेलेगा(khelegaa), खेलेगी(khelegii), खेलते(khelate), खेलती(khelatii), खेलने(khelane), खेलकर(khelakar)
- 9. Thus, we understand that the morphological richness of one language might vary from one language to another.
- 10. Indian languages are generally morphologically rich languages and therefore morphological analysis of words becomes a very significant task for Indian languages.



Types of Morphology

Morphology is of two types,

1. Inflectional morphology

Deals with word forms of a root, where there is no change in lexical category. For example, 'played' is an inflection of the root word 'play'. Here, both 'played' and 'play' are verbs.

2. Derivational morphology

Deals with word forms of a root, where there is a change in the lexical category. For example, the word form 'happiness' is a derivation of the word 'happy'. Here, 'happiness' is a derived noun form of the adjective 'happy'.

Morphological Features:

All words will have their lexical category attested during morphological analysis.

A noun and pronoun can take suffixes of the following features: gender, number, person, case

For example, morphological analysis of a few words is given below:

| Language | input:word | output:analysis |
|----------|------------------|--|
| Hindi | लंडके (ladake) | rt=ল্ডিকা(ladakaa), cat=n, gen=m, num=sg, case=obl |
| Hindi | लंडके (ladake) | rt=ল্ড্ৰা(ladakaa), cat=n, gen=m, num=pl, case=dir |
| Hindi | लड़कों (ladakoM) | rt=ল্ড্ৰা(ladakaa), cat=n, gen=m, num=pl, case=obl |
| English | boy | rt=boy, cat=n, gen=m, num=sg |
| English | boys | rt=boy, cat=n, gen=m, num=pl |

A verb can take suffixes of the following features: tense, aspect, modality, gender, number and person.

Language input:word output:analysis



Hindi हँसी(hansii) rt=हँस(hans), cat=v, gen=fem, num=sg/pl, per=1/2/3 tense=past, aspect=pft

English toys rt=toy, cat=n, num=pl, per=3

- 'rt' stands for root. 'cat' stands for lexical category. The value of lexical category can be noun, verb, adjective, pronoun, adverb, preposition.
- 'gen' stands for gender. The value of gender can be masculine or feminine.
- 'num' stands for number. The value of number can be singular (sg) or plural (pl).
- 'per' stands for person. The value of person can be 1, 2 or 3
- The value of tense can be present, past or future. This feature is applicable for verbs. The value of aspect can be perfect (pft), continuous (cont) or habitual (hab). This feature is not applicable for verbs.
- 'case' can be direct or oblique. This feature is applicable for nouns. A case is an oblique case when a postposition occurs after noun. If no postposition can occur after noun, then the case is a direct case. This is applicable for hindi but not english as it doesn't have any postpositions. Some of the postpositions in hindi are: কা(kaa), কা(kii), ক(ke), কা(ko), में(meM)

Chosen Experiment (Title with reference):

Title – Word Analysis

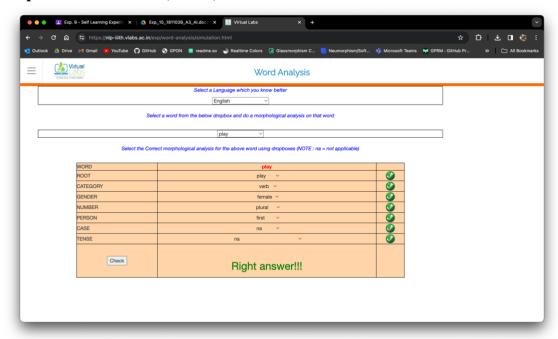
Link: http://nlp-

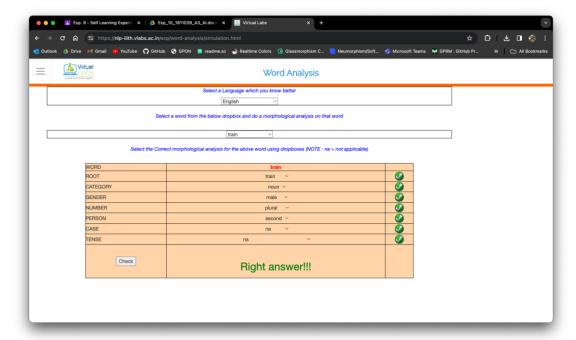
 $\underline{iiith.vlabs.ac.in/exp1/Introduction.html?domain=ComputerScience\&lab=Natural}$

LanguageProcessingLab

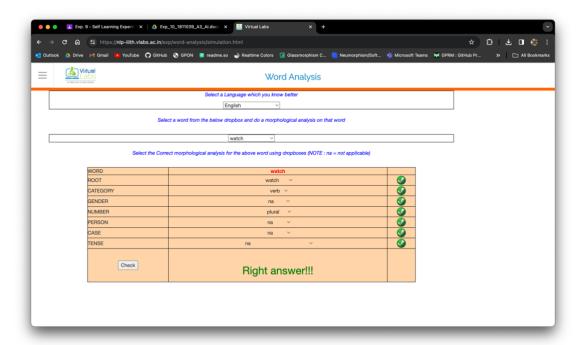


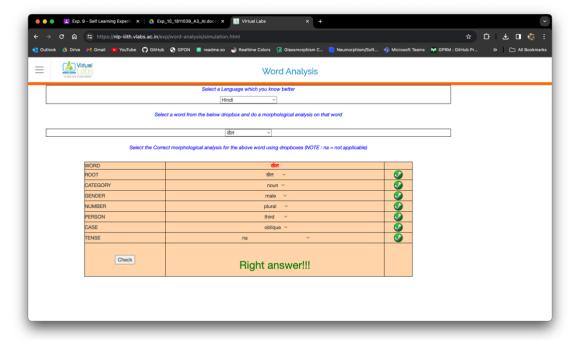
Snapshots of Execution (Min 3):



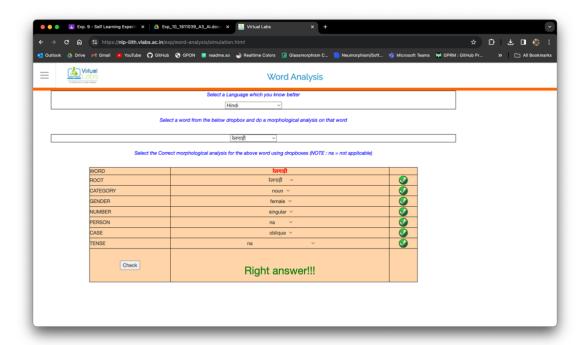


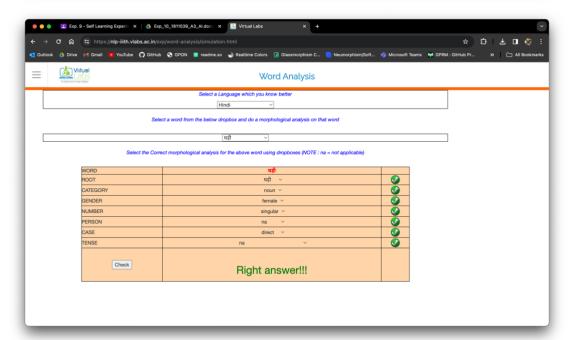














Findings/Analysis/Conclusion:-

- **3.** Analysis of a word into root and affix(es) is called as Morphological analysis of a word. It is mandatory to identify root of a word for any natural language processing task. A root word can have various forms.
- **4.** Thus, I understand that the morphological richness of one language might vary from one language to another. Indian languages are generally morphologically rich languages and therefore morphological analysis of words becomes a very significant task for Indian languages.
- 5. Also, I understood how to perform word analysis.