

VLAB Activity Report

Lab: Natural Language Processing

Name: Arshadali Firoj Pathan

Experiment: Word Analysis

Roll Number: 20CO088

Date of: 12/03/2024

Aim:

The aim of the experiment is to learn about morphological features of a word by analysing it.

Theory:

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. For example, the word 'play' in English has the following forms: 'play', 'plays', 'played' and 'playing'.

Thus we 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.

Types of Morphology:

- **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.
- **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
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, person. '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.

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:


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

bookings ▼

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

WORD	bookings	
ROOT	booking ▼	✓
CATEGORY	noun ▼	✓
GENDER	na ▼	✓
NUMBER	plural ▼	✓
PERSON	na ▼	✓
CASE	na ▼	✓
TENSE	na ▼	✓
<input type="button" value="Check"/>	Right answer!!!	