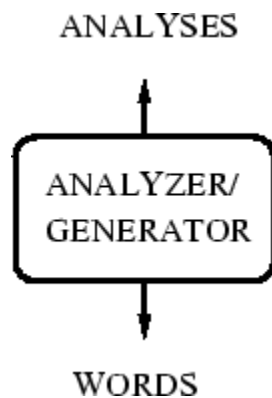


Name: Drashya
Roll number: 20BCE069
Course Name: Natural Language Processing
Course Code: 2CSDE70

Practical: 4
Title: Virtual Labs (<https://nlp-iiith.vlabs.ac.in/>)

Word Analysis

Aim: 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'.



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

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.

Output:



Virtual Labs
An IIT Madras Initiative

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

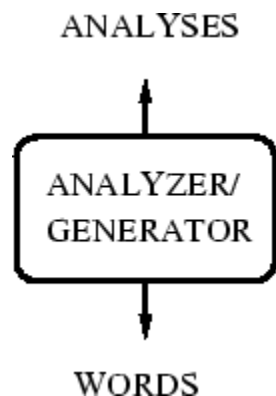
play ▼

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	simple-present ▼	✓
<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">Check</div>		Right answer!!!

Word Generation

Aim: 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'.



Objective: The objective of the experiment is to generate word forms from root and suffix information.

Procedure:

STEP 1: Select the language.

OUTPUT: Drop downs for selecting root and other features will appear.

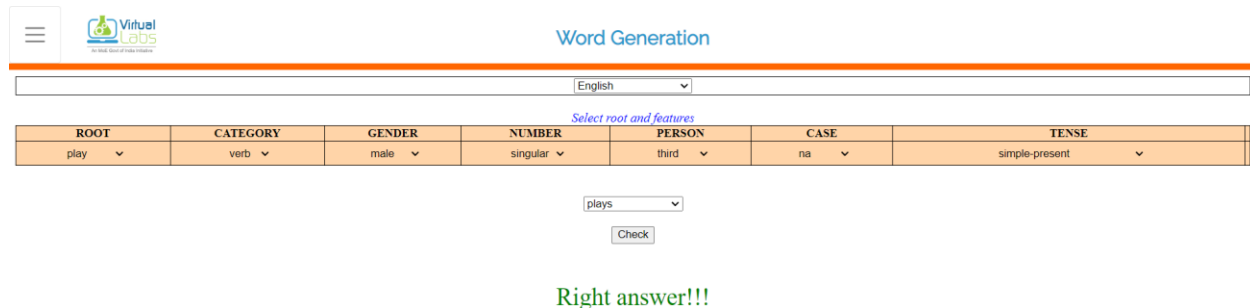
STEP 2: Select the root and other features.

STEP 3: After selecting all the features, select the word corresponding above features selected.

STEP 4: Click the check button to see whether right word is selected or not

OUTPUT: Output tells whether the word selected is right or wrong

Output:



Word Generation

English

Select root and features

ROOT	CATEGORY	GENDER	NUMBER	PERSON	CASE	TENSE
play	verb	male	singular	third	na	simple-present

plays

Check

Right answer!!!

N - grams

Aim: Probability of a sentence can be calculated by the probability of sequence of words occurring in it. We can use Markov assumption, that the probability of a word in a sentence depends on the probability of the word occurring just before it. Such a model is called first order Markov model or the bigram model.

$$P(W_n | W_{n-1}) = P(W_{n-1}, W_n) / P(W_{n-1})$$

Here, W_n refers to the word token corresponding to the n th word in a sequence.

Objective: The objective of this experiment is to learn to calculate bigrams from a given corpus and calculate probability of a sentence.

Procedure:

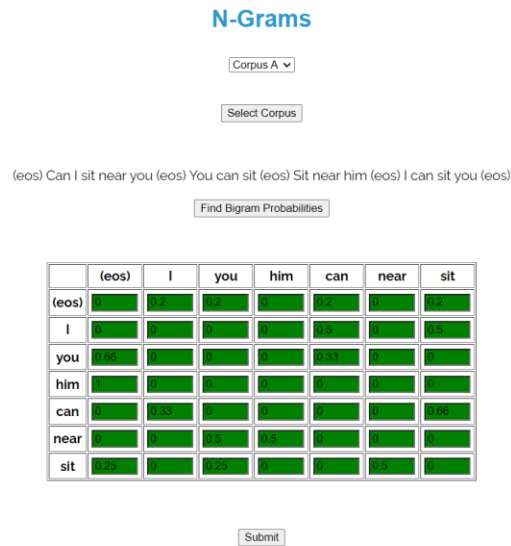
STEP 1: Select a corpus and click on Generate bigram table

STEP 2: Fill up the table that is generated and hit Submit

STEP 3: If incorrect (red), see the correct answer by clicking on show answer or repeat Step 2.

STEP 4: If correct (green), click on take a quiz and fill the correct answer

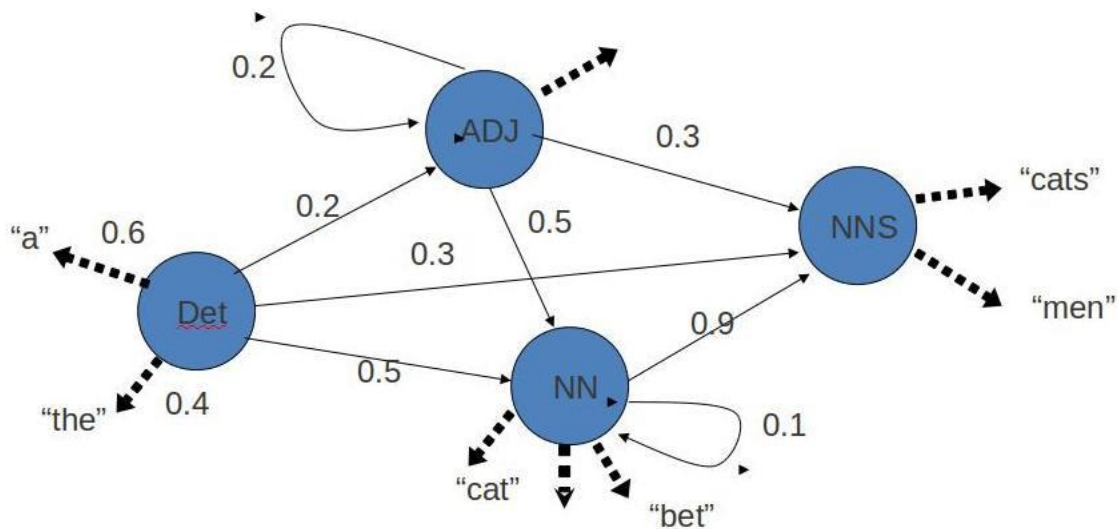
Output:



POS Tagging - Hidden Markov Model

Aim: POS tagging or part-of-speech tagging is the procedure of assigning a grammatical category like noun, verb, adjective etc. to a word. In this process both the lexical information and the context play an important role as the same lexical form can behave differently in a different context. For example the word "Park" can have two different lexical categories based on the context.

The boy is playing in the park. ('Park' is Noun)
Park the car. ('Park' is Verb)



Assigning part of speech to words by hand is a common exercise one can find in an elementary grammar class. But here we wish to build an automated tool which can assign the appropriate part-of-speech tag to the words of a given sentence. One can think of creating hand crafted rules by observing patterns in the language, but this would limit the system's performance to the quality and number of patterns identified by the rule crafter. Thus, this approach is not practically adopted for building POS Tagger. Instead, a large corpus annotated with correct POS tags for each word is given to the computer and algorithms then learn the patterns automatically from the data and store them in form of a trained model. Later this model can be used to POS tag new sentences.

In this experiment we will explore how such a model can be learned from the data.

Objective: The objective of the experiment is to calculate emission and transition matrix which will be helpful for tagging Parts of Speech using Hidden Markov Model.

Procedure:

STEP1: Select the corpus.

STEP2: For the given corpus fill the emission and transition matrix. Answers are rounded to 2 decimal digits.

STEP3: Press Check to check your answer.

Wrong answers are indicated by the red cell.

Output:



POS Tagging - Hidden Markov Model

Corpus A

EOS/eos **Book**/verb **a**/determiner **car**/noun EOS/eos **Park**/verb **the**/determiner **car**/noun EOS/eos **The**/determiner **book**/noun **is**/verb **in**/preposition **the**/determiner **car**/noun EOS/eos
The/determiner **car**/noun **is**/verb **in**/preposition **a**/determiner **park**/noun EOS/eos

Emission Matrix							
	book	park	car	is	in	a	the
determiner	0	0	0	0	0	1	1
noun	0.5	0.5	1	0	0	0	0
verb	0.5	0.5	0	1	0	0	0
preposition	0	0	0	0	1	0	0

Transition Matrix					
	eos	determiner	noun	verb	preposition
eos	0	0.33	0	0.5	0
determiner	0	0	1	0	0
noun	1	0	0	0.5	0
verb	0	0.33	0	0	1
preposition	0	0.33	0	0	0

Check

Right answer!!!