Semester : _____ Subject : _____ Academic Year: 20 - 20

What is POS Tagging?

↳ Tagging sentences with the respective parts of speech.
eg noun, verb, adjective etc.

John has a small dog.

Noun | Verb | determiner | adjective | noun.

Here, we will be trying to predict the parts of speech / word class (the sentence is already given with that), in the given input sentence

↳ Grammatical Analysis, Spell Checker.

⇒ Significance of POS Tagging.
↳ POS used NLP Pipeline, used in grammarly that does grammar check & spell check.

Scanned with OKEN Scanner

POS looking for:- (in a sentence)

① noun (name) :- John, car, India, apple, dog, house

② modal verb eg :- must, will, would, can, may

③ verb eg :- run, swim, talk, eat, speak etc.

Mary Saw Will	Mary saw Jane
Noun Verb Noun.	N V N

① Collect data (Labelled Data) .... (here we have as above)

② Create a lookup table
⇒ Here in this step, we tag each word with the most common part of speech.

③ Tag our target sentence
⇒ while going word by word, tag the most common part of speech associated with it

④ Done ☺

Scanned with OKEN Scanner

Semester : _____ Subject : _____ Academic Year: 20 - 20

Lookup Table

Q → Mary Saw Will

Given:- Jane saw Will
noun verb noun

Mary saw Jane
noun verb noun

	N	V
Mary	1	0
Saw	0	2
Jane	2	0
Will	1	0

Looking at the lookup table, we can find the most occuring part of speech for each and every word.

Mary saw Will.

noun + verb + noun (occurs as a ')

Scanned with OKEN Scanner

Labelled Data
(Our Data!)

Mary will see Jane
noun modal verb noun

Will will see Mary
noun modal verb noun

Jane will see Will
noun modal verb noun

Lookup Table

	N	V	M
Mary	2	0	0
see	0	3	0
Jane	2	0	0
will	2	0	3

"noun"

Tag this Q. Mary will see Will
noun modal verb ?? ← Lookup table says it is a modal verb even though the context is different

Scanned with OKEN Scanner

Semester : _____ Subject : _____ Academic Year: 20 - 20

This is because, lookup tables wont work if a single word can have multiple tags i.e. it wont consider any context.

Therefore, this wont work all the time.

So, how can we consider the context?

One way is using Bigrams.

It takes into consideration how the word was used / looking at the neighbour.

Bi Grams

For same above, labelled data,

	N-M	M-V	V-N
mary-will	1	0	0
will-see	0	3	0
see-jane	0	0	1
will-will	1	0	0
see-mary	0	0	1
jane-will	1	0	0
see-will	0	0	1

Scanned with OKEN Scanner

Mary will ~~to~~ will
noun modal verb noun

Q3] Mary Jane can see Will
Labelled data ⇒ Spot will see Mary
Will Jane spot Mary ?
Mary will pat Spot ?

Testdata →

Jane will Spot Will
? ? ? ?

Using Bigrams,

Jane . will

Jane followed by will ~~is~~ cannot be determined using bigram model.
Hence the question wont be solved by bigram model.
∴ Hidden Markov Model comes into picture.

Scanned with OKEN Scanner