

Pune Vidyarthi Griha's College of Engineering and Technology & G.K. Pate (Wani) Institute of Management, Pune- 411009.

(Affiliated to Savitribai Phule Pune University)



A Mini Project Report On

“POS Tagger for Indian Languages”

Submitted in Partial Fulfillment for the Term-work of Fourth year in Computer Engineering of Savitribai Phule Pune University.

BY

1	Bhagyashree Kande	9028
2	Aditi Wankhade	9083
3	Apurva Ingale	0071
4	Shivani Kulkarni	0072

**Under The Guidance of
Prof. M. V. Marathe**

Department of Computer Engineering

Academic Year: - 2022-2023

**Pune Vidyarthi Griha's College of Engineering and
Technology & G.K. Pate (Wani) Institute of
Management, Pune- 411009.**

(Affiliated to Savitribai Phule Pune University)



CERTIFICATE

**This is to certify that the Project report entitled “POS Taggers for Indian Languages”
submitted by,**

Students Name	Roll No
1. Bhagyashree Kande	9028
2. Aditi Wankhade	9083
3. Apurva Ingale	0071
4. Shivani Kulkarni	0072

is a record of bonafide work carried out by him/her, in the partial fulfilment of the Term-work of fourth year in Computer Engineering of Savitribai Phule Pune University at Pune Vidyarthi Griha's College of Engineering and Technology & G.K. Pate (Wani) Institute of Management, Pune under Savitribai Phule Pune University, Pune. This work is done during the academic year 2022-23.

Date: - 10/05/23

Place: - Pune

**Prof. M. V. Marathe
Subject Teacher**

**Prof. D. D. Sapkal
H.O.D. (Computer Engg.)**

Acknowledgment

It is our privilege to acknowledge with deep sense of gratitude to our Project guide Prof. M. V. Marathe Ma'am for her valuable suggestions and guidance throughout our course of study and timely help given to us in completion of our project.

We express our gratitude to Prof. Deepak Sapkal Sir (Head of Computer Engineering) for his kind help and co-operation. We are highly obliged to the entire staff of Computer Engineering Department for their kind co-operation and help. I would like to express my gratitude towards my fellow teammates for supporting me and working wonderfully as a team.

Bhagyashree Kande 9028

Aditi Wankahde 9083

Apurva Ingale 0071

Shivani Kulkarni 0072

Title – POS Tagger for Indian Language

Problem Statement –

Create a POS tagger on Hindi Language and display part of Speech as output.

Objectives -

3. Understand the importance of text preprocessing for NLP tasks.
4. Learn how to perform POS tagging using Python.
5. Learn how to create POS tags for regional languages.

Outcomes -

- Y Gain hands-on experience in preprocessing textual data, including identifying and removing stop words and performing lemmatization.
- Y Understand the concept of POS tagging of text preprocessing.

Theory –

1. POS tagging:

The process of assigning one of the parts of speech to the given word. It is generally called POS tagging. In simple words, we can say that POS tagging is a task of labelling each word in a sentence with its appropriate part of speech. We already know that parts of speech include nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.

Natural language processing (NLP) tools have sparked a great deal of interest due to rapid improvements in information and communications technologies. As a result, many different NLP tools are being produced. However, there are many challenges for developing efficient and effective NLP tools that accurately process natural languages. One such tool is part of speech (POS) tagging, which tags a particular sentence or words in a paragraph by looking at the context of the sentence/words inside the paragraph.

It is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form (*word, tag*)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.

Part of Speech Tagging using NLTK in python:

The Natural Language Toolkit (NLTK) is a platform used for building programs for text analysis. One of the more powerful aspects of the NLTK module is the Part of Speech tagging.

NLTK POS tag is the practice of marking up the words in text format for a specific segment of a speech context is known as POS Tagging (Parts of Speech Tagging). It is in charge of interpreting a language's text and associating each word with a specific token. Grammar tagging is another term for it. A part-of-speech tagger, also known as a POS-tagger, analyses a string of words.

In order to run the below python program you must have to install NLTK. Please follow the installation steps.

- X Open your terminal, run pip install nltk.
- X Write python in the command prompt so python Interactive Shell is ready to execute your code/Script.
- X Type import nltk
- X nltk.download()

A GUI will pop up then choose to download “all” for all packages, and then click ‘download’. This will give you all of the tokenizers, chunkers, other algorithms, and all of the corpora, so that’s why installation will take quite time.

```
import nltk
nltk.download()
```

Let’s knock out some quick vocabulary:

Corpus : Body of text, singular. Corpora is the plural of this.

Lexicon : Words and their meanings.

Token : Each “entity” that is a part of whatever was split up based on rules.

In corpus linguistics, part-of-speech tagging (POS tagging or PoS tagging or POST), also called grammatical tagging or word-category disambiguation.

Y NLTK POS Tag List

There is a multiple tag list available in nltk, tag list showing in output as per word.

- **CC**: It is the conjunction of coordinating
- **CD**: It is a digit of cardinal
- **DT**: It is the determiner
- **EX**: Existential
- **FW**: It is a foreign word
- **IN**: Preposition and conjunction
- **JJ**: Adjective
- **JJR and JJS**: Adjective and superlative
- **LS**: List marker
- **MD**: Modal
- **NN**: Singular noun
- **NNS, NNP, NNPS**: Proper and plural noun
- **PDT**: Predeterminer
- **WRB**: Adverb of wh
- **WP\$**: Possessive wh
- **WP**: Pronoun of wh
- **WDT**: Determiner of wp
- **VBZ**: Verb
- **VBP, VBN, VBG, VBD, VB**: Forms of verbs
- **UH**: Interjection
- **TO**: To go
- **RP**: Particle
- **RBS, RB, RBR**: Adverb
- **PRP, PRP\$**: Pronoun personal and professional

- **Procedure:**

Solution contains the use of Open-Source And widely used library NLTK.

The whole solution is divided into two parts;

- Training the Pos tagger to detect Hindi Part of speech
- Predicting the tags for a new phrase or sentence.

Training of the Pos Tagger

We use indian Corpus provided by NLTK, to train our tagger. The whole dataset contains 540 records out of which we use 90% dataset i.e 486 records to train and remaining to evaluate the accuracy

Our Model performs with 81.11 % accuracy on test dataset.

Prediction of Tags

Once the model is trained, the tagger is passed to a seperate function which takes care of prediction of tags, given a sentence.

Conclusion -

In this assignment, we perform fundamental text preprocessing techniques in NLP i.e. POS tagging for regional language Hindi. We created a tagger which identifies tags of a given sentence.

Code –

```
import nltk
from nltk.corpus import indian
from nltk.tag import tnt
import string

nltk.download("punkt")
nltk.download("indian")

def train():

    taggedSet = "hindi.pos"
    wordSet = indian.sents(taggedSet)
    count = 0
    print(wordSet[0])
    for sen in wordSet:
        count = count + 1
        sen = "".join(
            [
                " " + i if not i.startswith('"') and i not in string.punctuation else
                i for i in sen
            ]
        ).strip()
        print(count, sen, "sentences")
    print("Total sentences in the tagged file are", count)

    trainPerc = 0.9

    trainRows = int(trainPerc * count)
    testRows = trainRows + 1

    data = indian.tagged_sents(taggedSet)
    train_data = data[:trainRows]
    test_data = data[testRows:]

    print("Training dataset length: ", len(train_data))
    print("Testing dataset length: ", len(test_data))

    pos_tagger = tnt.TnT()
    pos_tagger.train(train_data)
    print("Accuracy: ", pos_tagger.evaluate(test_data))
    return pos_tagger

def tagger(pos_tagger, sentenceToBeTagged):
```

tokenized = nltk.word_tokenize(sentenceToBeTagged)

return pos_tagger.tag(tokenized)

if __name__ == "__main__":

pos_tagger = train()

sentence_to_be_tagged = "

"

print(tagger(pos_tagger, sentence_to_be_tagged))

Output:

```
pos_tagger = train()
sentence_to_be_tagged = "३९ में द चौकों और एक छक्के की मदद से ३४ रन बनाने वाले पर २ अौत तक आउट नहीं हुए ।"
print(tagger(pos_tagger, sentence_to_be_tagged))

nltk_data] Downloading package punkt to /root/nltk_data...
nltk_data] Package punkt is already up-to-date!
nltk_data] Downloading package indian to /root/nltk_data...
nltk_data] Package indian is already up-to-date!
पूर्ण, प्रतिबंध, हटाओ, ':', 'इराक']
पूर्ण प्रतिबंध हटाओ: इराक sentences
संयुक्त राष्ट्र । sentences
इराक के विदेश मंत्री ने अमरीका के उस प्रस्ताव का मजाक उड़ाया है, जिसमें अमरीका ने संयुक्त राष्ट्र के प्रतिबंधों को इराकी नागरिकों के लिए कम हानिकारक बनाने के लिए कहा है । sentences
विदेश मंत्री का कहना है कि चूंकि बग़दाद संयुक्त राष्ट्र की मांगों का पालन करते हुए अपने भारी विनाशकारी हथियारों को नष्ट कर रहा है । sentences
सिंहाना प्रतिबंधों को पूर्ण रूप से उठा दिया जाना चाहिए । sentences
विदेश मंत्री मोहम्मद सईद का कहना है कि वे इसे 'सुव्यवस्थित प्रतिबंध' कह कर आम राय और सुरक्षा परिषद को छल रहे हैं । sentences
बेनलौर की सुनवाई स्थगित sentences
कराची । sentences
पाकिस्तान की पूर्व प्रधानमंत्री बेनजीर भुट्टो पर लगे भ्रष्टाचार के आरोपों के खिलाफ भुट्टो द्वारा दायर की गई याचिका की सुनवाई मंगलवार को वकीलों की हड़ताल के कारण स्थगित कर दी गई । sentences
० सिंह हाईकोर्ट बार एसोसिएशन के अध्यक्ष रसीद रिजवी के मृताबिक यह हड़ताल उच्च न्यायालय और निचली अदालतों के स्तर पर सफल रही । sentences
१ देश में पुनः प्रजातंत्र की स्थापना की मांग को लेकर यह हड़ताल की गई थी । sentences
२ सुप्रीम कोर्ट में भुट्टो के उक्त मामले की सुनवाई सोमवार से शुरू हुई, जो फिलहाल बुधवार तक स्थगित है । sentences
३ मुशर्रफ सऊदी अरब को मनाएंगे sentences
४ इस्लामाबाद । sentences
५ पाकिस्तानी सेन्य प्रशासक जनरल परवेज मुशर्रफ द्वारा सऊदी अरब को कश्मीर विवाद मुद्दे पर भारत और पाकिस्तान की मध्यस्थता के लिए मनाने की संभावना व्यक्त की गई है । sentences
६ आधिकारिक सूत्रों के हवाले से कहा गया है कि मुशर्रफ की बुधवार से शुरू हो रही हड़ यात्रा इस मायने में काफी महत्वपूर्ण हो सकती है । sentences
७ मुशर्रफ अपने इस हड़ यात्रा के दौरान कश्मीर के अलावा फिलीस्तिन, अमरीका द्वारा ओसमा बिन लादेन के प्रत्यर्पण की मांग आदि पर चर्चा करेंगे । sentences
८ इस दौरान मुशर्रफ सऊदी राजा फाहद, उप प्रधानमंत्री सहित कई बड़ी हस्तियों से मिलेंगे । sentences
९ पाकिस्तान के बर्खस्त पूर्व प्रधानमंत्री नवाज शरीफ के मामले में भी यहां चर्चा होगी । sentences
० पत्रकारों के लिए कल्याण कोष sentences
१ नई दिल्ली । sentences
२ वित्तमंत्री यशवंत सिन्हा ने अपने बजट में पत्रकारों के लिए कल्याण कोष स्थापित करने की घोषणा की है । sentences
३ कोष में सूचना एवं प्रसारण मंत्रालय १ करोड़ की सहायता देगा । sentences
४ बजट प्रस्तुत करते हुए सिन्हा ने कहा कि पत्रकार कई अवसरों पर कवरेज के लिए जोखिम उठाते हैं यह कोष उनकी सेवा एवं संघर्ष के लिए होगा । sentences
५ उन्होंने कहा कि पांच वर्ष में एक बार मान्यता प्राप्त पत्रकार को करमुक्त आयात द्वारा व्यक्तिगत उपयोग के लिए सामग्री मंगाने की सुविधा अब दो वर्ष कर दी गई है । sentences
६ फिलहाल मान्यता प्राप्त पत्रकार एवं कैमरामैन कैमरा, फैक्स, कम्प्यूटर आदि जिनकी अधिकतम लागत १ लाख रुपए तक हो बिना कर्टम छूटी चुकाए पांच साल में एक बार मंगा सकते थे । sentences
७ बजट विकास को गति देगा: वाजपेयी sentences

८६ दोनों ही प्राचीनतम इस्लामी देश हैं । sentences
८७ क्या वहां लोकतंत्र है? sentences
८८ यदि नहीं तो फिर पाकिस्तान में क्या आवश्यकता है । sentences
८९ इसलाम और लोकतंत्र का कोई मेल नहीं यह स्पष्ट कहा जाता है । sentences
९० आयोग का कहना है कि धार्मिक कानूनों का दुरुूप्योग बढ़ता जा रहा है । sentences
९१ मानव अधिकार आयोग पाकिस्तान के आतंकवादी संगठनों को पाकिस्तान सरकार द्वारा संरक्षण दिये जाने का भी विरोधी है । sentences
९२ उस का कहना है कि सैनिक शासन में आतंकवाद को प्रशिक्षण देने वाले मदरसे अधिक सक्रिय हो गए हैं । sentences
९३ उनकी इस सक्रियता से पाकिस्तान में मानव अधिकारों की स्थिति खराब होती जा रही है । sentences
९४ आयोग का आंकलन है कि इस से पाकिस्तान के आसपास जितने देश हैं उनमें आतंकवाद विकसित होगा । sentences
९५ जहां कहीं आतंकवाद चलेगा वहां मानव अधिकार के घटने और उस पर रोक लगने के अवसर अधिक बढ़ जाएंगे । sentences
९६ मानव अधिकार आयोग ने खुल्लम खुल्ला पाक सरकार को कहा है कि वह उन मदरसों को बंद कर दें जहां आतंकवाद सिखाया जाता है । sentences
९७ देश के धर्म गुरूओं से भी कहा कि वे धर्म को हिंसा और अराजकता के लिये नहीं अपितु मानव के कल्याण के लिए उपयोगी बनाएं । sentences
९८ मानव आयोग ने अपनी रपट में पाकिस्तान में बढ़ती बेरोजगारी का भी उल्लेख किया है । sentences
९९ सरकारी आंकड़ों के अनुसार पाकिस्तान में बेकारी की दर ३८ प्रतिशत है लेकिन आयोग का कहना है कि यह दर ४५ तक पहुंच सकती है । sentences
१०० बेकारी अनेक दुर्गुणों की जन्मदाता है । sentences
१०१ अपराध और अनैतिकता के फैलाव में वह महत्व का भाग बनती है । sentences
१०२ जब किसी देश से पुंजी और बौद्धिक कोशल का पलायन होता है तो उस समाज और देश को दीवालिया होने से कोई रोक नहीं सकता । sentences
१०३ मानव अधिकार आयोग की रपट को पाकिस्तान के गृह मंत्री ने झूठ और बकवास के पुलिंदे की संज्ञा दी है और आयोग की रपट को स्वीकार करने से े इनकार कर दिया है । sentences
१०४ रेल बजट पर बीजद नाराज sentences
१०५ भुवनेश्वर । sentences
१०६ राष्ट्रीय जनतांत्रिक गठबंधन की सहयोगी पार्टी बीजू जनता दल ने रेल बजट में उड़ीसा की उपेक्षा पर अपनी नाराजगी जताई है । sentences
१०७ दल में राजग से समर्थन वापस लेने की आवाजें भी उठने लगी हैं । sentences
१०८ बीजू जनता दल के अध्यक्ष एवं उड़ीसा के मुख्यमंत्री नवीन पटनायक ने प्रधानमंत्री अटलबिहारी वाजपेयी को लिखे पत्र में इस मुद्दे पर अपनी नाराजगी जताते हुए कहा है कि बजट में सिर्फ बंगाल का ख्याल रखा गया है । sentences
१०९ इसके साथ ही बीजद के सांसद ममता से इस्तीफे की मांग उठा रहे हैं । sentences
११० उन्होंने कहा है कि 'बीजद-राजग का सहयोगी है अतः अपनी मांगों पर हम चर्चा करना चाहते हैं । sentences
१११ हम इन जायज मांगों को लेकर लड़ते रहेंगे । sentences
११२ ' नवीन ने कहा कि हमारे राज्य को सिर्फ एक एक्सप्रेस ट्रेन दी गई । sentences
११३ उन्होंने कहा कि उड़ीसा के साथ अन्याय हुआ है । sentences
११४ बताया जाता है कि पार्टी के भीतर इस मामले पर राजग सरकार से समर्थन वापस लेने का नवीन पटनायक पर खासा दबाव है । sentences
११५ इसके चलते ही नवीन मुखर होकर इस मामले पर आगे आए हैं । sentences
११६ हालांकि बीजद नेता प्रसन्न पटसानी कहते हैं कि-'समर्थन वापस लेना समस्या का हल नहीं है । sentences
११७ इसलिए हम प्रधानमंत्री से अपील कर रहे हैं कि वे इस ओर ध्यान दें ।' sentences
११८ बजट पर सिर्फ राजग जमा
```


441 पटेल छबीलदास मेहता की सरकार में उपमुख्यमंत्री के रूप में भी दायित्व निभा चुके थे । sentences
442 राजीव गांधी और सीताराम केसरी की अध्यक्षता में वे अखिल भारतीय कांग्रेस कमेटी में संविध थे । sentences
443 १९९८ में उन्हें प्रदेश कांग्रेस का अध्यक्ष बनाया गया था । sentences
444 कश्मीर में युद्ध विराम की अवधि ३ माह बढ़ी sentences
445 नई दिल्ली । sentences
446 प्रधानमंत्री अटलबिहारी वाजपेयी ने गुरुवार को लोकसभा में कश्मीर में जारी युद्ध विराम की अवधि ३ माह के लिए बढ़ाने की घोषणा की । sentences
447 प्रधानमंत्री ने लोकसभा में सरकार के इस निर्णय की जानकारी देते हुए कहा कि युद्ध विराम की अवधि में शांति प्रक्रिया को आगे बढ़ाने के लिए सैन्य बलों द्वारा आतंकवादियों के खिलाफ कार्रवाई रुकी रहेगी । sentences
448 उन्होंने पाकिस्तान से कहा कि वह शांति प्रक्रिया को आगे बढ़ाने में सहयोग करे और आतंकवादी गुर्गों को मदद देना बंद जो कश्मीर घाटी में दहशत फैला रहे हैं । sentences
449 प्रधानमंत्री ने कहा कि उन्हें उम्मीद है कि पड़ोसी देश पाकिस्तान, भारत विरोधी दुश्मचार रोक कर सीमा पार आतंकवाद को रोकने में सहयोग करेगा । sentences
450 इसके पूर्व बुधवार राति सर्वदलीय बैठक में कश्मीर में शांति प्रक्रिया को आगे बढ़ाने के लिए फैसला ले लिया गया था । sentences
451 लगभग ८० मिनट लंबी चली इस बैठक में आमसहमति से कश्मीर में युद्ध विराम को जारी रखने का फैसला हुआ । sentences
452 प्रधानमंत्री ने इसकी औपचारिक घोषणा आज लोकसभा में की । sentences
453 पंजाब राष्ट्रीय खेलों की तैयारी में sentences
454 चंडीगढ़ । sentences
455 सितम्बर में पंजाब में आयोजित होनेवाले राष्ट्रीय खेलों की तैयारी जोरों पर है । sentences
456 खेलों से संबंधित आधारीक संरचनाओं के नवीनीकरण किये जा रहे हैं । sentences
457 राज्य के खेल विभाग के एक प्रवक्ता ने गुरुवार को कहा कि मरम्मत और नवीनीकरण के काम में २४ करोड़ रुपये खर्च किये जा रहे हैं । sentences
458 छह करोड़ कीमत की खेलों से संबंधित उपकरण खरीदे या आयात किये जा रहे हैं । sentences
459 सभी खेल कॉम्प्लेक्स अगले दो माह के अंदर तैयार हो जायेंगे । sentences
460 प्रवक्ता ने कहा कि राष्ट्रीय खेल पांच भिन्न स्थलों पर आयोजित किये जायेंगे । sentences
461 ये स्थल पटियाला स्थित पोलो ग्राउंड एंड नेशनल इंस्टीट्यूट ऑफ स्पोर्ट्स, लुधियाना स्थित गुरुनानक स्टेडियम एंड पंजाब एथ्रीकल्चर युनिवर्सिटी, जालंधर स्थित हंस राज स्टेडियम एंड सुरजीत हॉकी स्टेडियम, आनंदपुर साहिब और चंडीगढ़ है । sentences
462 प्रवक्ता ने बताया कि जालंधर में सिंथेटिक एथलेटिक ट्रैक और लुधियाना में एस्ट्रो टर्फ हॉकी मैदान का निर्माण पांच करोड़ की लागत से किया जा चुका है । sentences
463 युगल मैचों में नहीं खेलेगी हिंगिस sentences
464 दुबई । sentences
465 मोर्टिना हिंगिस ने इस सीजन में युगल मैच नहीं खेलने की घोषणा कर दी है । sentences
466 विश्व की नंबर एक टेनिस खिलाड़ी हिंगिस इस वर्ष मोनिका सेलेस के साथ मिलकर कई युगल टेनिस मैचों में खेलना चाहती थीं । sentences
467 लेकिन अब वह केवल एकल मैचों में भाग लेने की बात कर रही है । sentences
468 हिंगिस ने कहा, "मे युगल मैचों में खेलकर थक गई हूं । मुझे केवल एकल मैचों में खेलकर अच्छा लग रहा है । sentences
469 पिछले सप्ताह दोहा में मैं प्रायोजकों के साथ कुछ पार्टियों में शरीक हुई । sentences
470 इससे मुझे थोड़ा खाली समय मिलने के साथ-साथ एकान्त भी मिला । sentences
471 अतः अब मैं उन सभी पार्टियों में शामिल होऊंगी जिनमें सामान्य तौर पर मैं शामिल नहीं हुआ करती थी ।" sentences
472 हिंगिस ने कहा, " इस साल मैं युगल मैचों में खेलना नहीं चाहती । sentences
473 लेकिन यदि मुझे ऐसा लग कि मैं लक्जरे बैचर बन जाऊंगी तो मैं उन भी लक्जरे बैचर बन जाऊंगी ।"

✓ 1s completed at 4:40 PM

515 वार्न की गेंदबाजी प्रभावशाली नहीं रही । मैचमा ने ४६ रन देकर तीन और पाॉटिंग ने १० रन देकर दो बल्लेबाजों का आउट किया । sentences
516 तीसरा एकदिवसीय मैच पाक ने जीता sentences
517 वॉलेंटन । sentences
518 पाकिस्तान ने पंाच एकदिवसीय मैचों की श्रृंखला के तीसरे मैच में न्यू जीलैंड को २८ रनों से हरा दिया । sentences
519 न्यू जीलैंड को जीत के लिए २४४ रन बनाने थे लेकिन उनकी पारी ४८. २ ओवर में २१५ रन पर सिमट गई । sentences
520 मेकमिलन, विंसेट और पर्रो को छोड़ टीम का कोई भी बल्लेबाज महत्वपूर्ण योगदान नहीं दे सका । sentences
521 तीन विकेट लेने वाले ऑफ स्पिनर सक लेन मूस्ताक को मेन ऑफ द मैच घोषित किया गया । sentences
522 न्यू जीलैंड की शुरुआत अच्छी नहीं रही जब उनके तीन बल्लेबाज फ्लेमिंग, टुज और एस्टल को टीम के १८ रन के योग पर ही पेवेलियन लौट जाना पड़ा । sentences
523 अकरम ने फ्लेमिंग को एक रन पर और एस्टल को दस रन पर बोल्ल किया । sentences
524 टुज को वकार युनुस ने विकेट के पीछे मोह्न खान के हाथों एक रन पर लपकवाया । sentences
525 लेकिन फिर विंसेट और मेकमिलन के बीच चौथे विकेट के लिए १०० रनों की भागीदारी हुई । sentences
526 विंसेट ने ६१ गेंदों में तीन चौकों की मदद से ४३ रनों की पारी खेली । sentences
527 उन्हें सकलेन ने मोह्न के हाथों स्टम्प कराया । sentences
528 मेकमिलन ने ८३ गेंदों में छह चौकों और एक छक्के की मदद से ६४ रनों की पारी खेली । sentences
529 न्यू जीलैंड के १५२ रन के योग पर मेकमिलन को अब्दुर रज्जाक ने यूसुफ योहाना के हाथों लपकवाया । sentences
530 छोटा विकेट भी १५२ रन पर ही गिरा जब क्रिस हैरिस को अब्दुर रज्जाक ने ११ रन पर बोल्ल कर दिया । sentences
531 विटोरी, टफी, फ्रेकलीन और मार्टिन तेज बल्लेबाजी नहीं कर पाये और न्यू जीलैंड की पारी २१५ रनों पर सिमट गई । sentences
532 ३९ गेंदों में दो चौकों और एक छक्के की मदद से ३४ रन बनाने वाले पर्रोरे अंत तक आउट नहीं हुए । sentences
533 अकरम और सकलेन ने तीन-तीन, रज्जाक ने दो और युनुस तथा महमूद ने एक-एक विकेट लिये । sentences
534 इससे पहले पाकिस्तान ने निर्धारित ५० ओवर में नौ विकेट खोकर २४३ रन बनाये । sentences
535 सईद अनवर ने ५७, इमरान नजीर ने ३२, सलीम ईलाही ने ३० और यूसुफ योहाना ने ४७ रनों की पारी खेली । sentences
536 अनवर और नजीर ने पहले विकेट के लिए ४५ और फिर दूसरे विकेट के लिए अनवर और ईलाही ने मिलकर ६० रनों की भागीदारी की । sentences
537 पाक के मुख्यले बल्लेबाजों में अजहर महमूद ने १७ रन बनाये । sentences
538 कप्तान मोह्न खान ने १५ रनों का योगदान दिया । sentences
539 न्यू जीलैंड की तरफ से टफी और फ्रेक्लीन ने तीन-तीन और हैरिस ने दो विकेट लिये । sentences
540 अनवर को विंसेट ने रन आउट किया । sentences
Total sentences in the tagged file are 540
Training dataset length: 486
Testing dataset length: 53
<ipython-input-3-2c4275bc2ca2>:36: DeprecationWarning:
Function evaluate() has been deprecated. Use accuracy(gold)
instead.
print("Accuracy: ", pos_tagger.evaluate(test_data))
Accuracy: 0.8111964873765993
[('३९', 'QFNUM'), ('गेंद', 'NN'), ('ों', 'Unk'), ('में', 'PREP'), ('द', 'NNPC'), ('चौक', 'Unk'), ('ों', 'Unk'), ('और', 'CC'), ('एक', 'QFNUM'), ('छक्के', 'QFNUM'), ('की', 'PREP'), ('मदद', 'NNPC')]