



**S. B. JAIN INSTITUTE OF TECHNOLOGY,
MANAGEMENT & RESEARCH, NAGPUR.**

Practical No. 2

Aim: Perform the syntactic parsing for feature extraction on the text data

Name of Student: _____

Roll No.: _____

Semester/Year: IV/VII

Academic Session: 2025-2026

Date of Performance: _____

Date of Submission: _____

AIM: Perform the syntactic parsing for feature extraction on the text data

OBJECTIVE/EXPECTED LEARNING OUTCOME:

- Understanding syntactic parsing
- Able to apply feature extraction on the text data

HARDWARE AND SOFTWARE REQUIREMENTS:

Hardware Requirement:

Software Requirement:

THEORY:

To analyse a preprocessed data, it needs to be converted into features. Depending upon the usage, text features can be constructed using assorted techniques – Syntactical Parsing, Entities / N-grams / word-based features, Statistical features, and word embeddings.

Syntactical parsing involves the analysis of words in the sentence for grammar and their arrangement in a manner that shows the relationships among the words. Dependency Grammar and Part of Speech tags are the important attributes of text syntactics

Entities are defined as the most important chunks of a sentence – noun phrases, verb phrases or both. Entity Detection algorithms are generally ensemble models of rule based parsing, dictionary lookups, pos tagging and dependency parsing. The applicability of entity detection can be seen in the automated chat bots, content analyzers and consumer insights.

Word embedding is the modern way of representing words as vectors. The aim of word embedding is to redefine the high dimensional word features into low dimensional feature vectors by preserving the contextual similarity in the corpus.

CODE:

```
import spacy
import pandas as pd

# Load spaCy English model
nlp = spacy.load("en_core_web_sm")

# Take sentence as input from the user
text = input("Enter a sentence: ")

# Process the text
doc = nlp(text)

# Create a list to hold feature data
data = []

# Mapping for full part-of-speech names
pos_full_form = {
    "NOUN": "Noun",
    "VERB": "Verb",
    "AUX": "Auxiliary Verb",
    "PRON": "Pronoun",
    "DET": "Determiner",
    "PUNCT": "Punctuation",
    "PROPN": "Proper Noun",
    "ADP": "Adposition",
    "CONJ": "Conjunction",
    "SCONJ": "Subordinating Conjunction",
    "PART": "Particle",
    "INTJ": "Interjection",
    "NUM": "Numeral",
    "SYM": "Symbol",
    "X": "Other"
}

# Extract features from each token
for token in doc:
    part_of_speech = pos_full_form.get(token.pos_, token.pos_)
    data.append({
        "Word": token.text,
        "Root Word (Lemma)": token.lemma_,
        "Part of Speech": part_of_speech,
        "Grammatical Tense": ', '.join(token.morph.get("Tense")) if token.morph.get("Tense") else "None",
        "Gender": ', '.join(token.morph.get("Gender")) if token.morph.get("Gender") else "None"
    })

# Convert to DataFrame for table format
df = pd.DataFrame(data)
print("\nExtracted Features:\n")
print(df.to_string(index=False))
```

OUTPUT (SCREENSHOT):

Enter a sentence: He eats an Apple.

Extracted Features:

Word	Root Word (Lemma)	Part of Speech	Grammatical Tense	Gender
He	he	Pronoun	None	Masc
eats	eat	Verb	Pres	None
an	an	Determiner	None	None
Apple	Apple	Proper Noun	None	None
.	.	Punctuation	None	None

Morphology

Select a Root Word

बच्चा

Fill the add delete table here:

Delete	Add	Number	Case	Correction
आ	आ	sing	dr	✓
आ	ए	plu	dr	✓
आ	ए	sing	ob	✓
आ	औ	plu	ob	✓

For Example for लड़का:

Delete	Add	Number	Case
आ	आ	sing	dr
आ	ए	plu	dr
आ	ए	sing	ob
आ	औ	plu	ob

Submit

Correct Answer!

Morphology

Select a Root Word

बच्चा

Fill the add delete table here:

Delete	Add	Number	Case	Correction
आ	ए	sing	dr	✗
आ	ए	plu	dr	✓
आ	आ	sing	ob	✗
आ	आ	plu	ob	✗

For Example for लड़का:

Delete	Add	Number	Case
आ	आ	sing	dr
आ	ए	plu	dr
आ	ए	sing	ob
आ	औ	plu	ob

Submit

Error in your Add-Delete table!

Get Answer

CONCLUSION:

DISCUSSION AND VIVA VOCE:

- What is feature Extraction?
- What are the types of text Extraction?
- What is Word Embedding?

REFERENCE:

- www.w3schools.com
- www.tutorialsmade.com
- <https://www.javatpoint.com/>