USN | | | | | | | | | | |

Internal Assessment Test 1 – September 2025

Sub	Natural Language Processing					Sub Code:	BCS714B	Branch:	CSE	
Date:	29/09/2025	Duration:	90 mins	Max Marks:	50	Sem /Sec:	VII Professional Elective		OBE	

	Answer any FIVE FULL Questions	MARKS	CO	RBT
1	Define NLP. List and explain applications of NLP.	[10]	CO1	L2
2	Consider the following corpus of three sentences There is a big garden. Children play in a garden. They play inside beautiful garden. Calculate the probability P for the sentence "They play in a big garden" assuming a grammar language model.	[10]	CO1	L3
3	Explain Minimum Edit Distance algorithm. Compute minimum edit distance between "Tutor" and "Tumor".	[10]	CO2	L3
4	Derive a top down and bottom up parse tree for the given sentence "The angry bear chased the frightened little squirrel" Use the following grammar rule to create the parse tree S -> NP VP Det -> the NP -> Det Nom Adj -> little/angry/frightened VP -> V NP N -> Squirrel/bear Nom -> adj N V -> chased	[10]	CO2	L3
5	Explain design features of IR with a neat diagram.	[10]	CO4	L2
6	Explain WordNet with its applications.	[10]	CO4	L2

CI CCI HOD-CSE

USN | | | | | | | | | | |

Internal Assessment Test 1 – September 2025

Sub	Natural Language Processing					Sub Code:	BCS714B	Branch:	CSE	
Date:	29/09/2025	Duration:	90 mins	Max Marks:	50	Sem /Sec:	VII Professional Elective		OBE	

	Answer any FIVE FULL Questions	MARKS	CO	RBT
1	Define NLP. List and explain applications of NLP.	[10]	CO1	L2
2	Consider the following corpus of three sentences There is a big garden. Children play in a garden. They play inside beautiful garden. Calculate the probability P for the sentence "They play in a big garden" assuming a grammar language model.	[10]	CO1	L3
3	Explain Minimum Edit Distance algorithm. Compute minimum edit distance between "Tutor" and "Tumor".	[10]	CO2	L3
4	Derive a top down and bottom up parse tree for the given sentence "The angry bear chased the frightened little squirrel" Use the following grammar rule to create the parse tree S -> NP VP Det -> the NP -> Det Nom Adj -> little/angry/frightened VP -> V NP N -> Squirrel/bear Nom -> adj N V -> chased	[10]	CO2	L3
5	Explain design features of IR with a neat diagram.	[10]	CO4	L2
6	Explain WordNet with its applications.	[10]	CO4	L2

CI CCI HOD-CSE

IAT Solution

1) NLP - Natural Language Processing:

- NLP is a field of AI and Computational linguistics that concerned with the development of a model for processing natural human language.
- It enables computers to interpret, understand and process the human language.
- The main aims of nlp are:
  - Develop automated tools for natural language processing
  - Understanding the concept of how human acquire, store, and process the lexical memory (language).
- Applications of NLP:

i) Machine Translation:

- Used to convert text from one language into another. (Ex: Google Translate)

ii) Speech identification:

- Used to convert speech into text form.
- Ex: siri, alexa etc.

iii) Speech synthesis:

- Used to convert the text to human readable speech.
- Ex: Screen Recoder

iv) Natural Language Interface to databases:

- Database querying in natural or simple english sentences.
- Ex: Show Student from 7th sem

v) Data Information Retrival (IR):

- Retrieves the relevent information or documents.

vi) Information Extraction:

- Extracts the structured facts (names, date, relation) from unstructured information.

2) Given Sentences:

<s> There is a big garden. </s>
<s> Children plays in a garden. </s>
<s> They plays inside beautiful garden. </s>

Test sentence:

They plays in a big garden.

Probability using Bi-gram Model:

P(They play in a big garden) = P(They | <s>) * P(play | They)
* P(in | play) * P(a | in)
* P(big | a) * P(garden | big)

$= \frac{1}{3} \times \frac{1}{1} \times \frac{1}{2} \times \frac{1}{1} \times \frac{1}{2} \times \frac{1}{1}$

$= \frac{1}{3} \times 1 \times \frac{1}{2} \times 1 \times \frac{1}{2} \times 1$

$= \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}$

$= 0.33 \times 0.5 \times 0.5$

$= 0.0825$

∴ P(They play in a big garden) = 0.0825

3) Minimum Edit Distance Algorithm:

The minimum edit distance algorithm states that the minimum edit distance between two string is the minimum required operations to convert one string into another.

- Operations Allowed:

- Insertion (add a letter)
- Deletion (remove a letter)
- Substitution (replace a letter with another)

• Algorithm:

Input: Two strings x and y
Output: Minimum edit distance b/w x and y

```
m ← length(x)
n ← length(y)

for i=0 to m do
    dist[i,0] = i

for j=0 to n do
    dist[0,j] = j

for i=0 to m do
    for j=0 to n do
        dist[i,j] = min{ dist[i-1,j] + insert_cost,
                         dist[i-1,j-1] + subset_cost[x_i, y_j],
                         dist[i,j-1] + delet_cost }
```

• The distance is calculated using:

```
dist[i,j] = min{ dist[i-1,j] + insert_cost,
                 dist[i-1,j-1] + subset_cost[x_i, y_j],
                 dist[i,j-1] + delet_cost }
```

• Example:

Tutor Vs Tumor

	#	T	U	m	O	r
#	0	1	2	3	4	5
T	1	0	1	2	3	4
u	2	1	0	1	2	3
t	3	2	1	1	2	3
o	4	3	2	2	1	2
r	5	4	3	3	2	(1) → minimum edit distance

∴ The minimum edit distance b/w "Tutor" & "Tumor" = 1,

Tutor → replace 't' with 'm' → Tumor)

4) Top down & Bottom up parse Tree

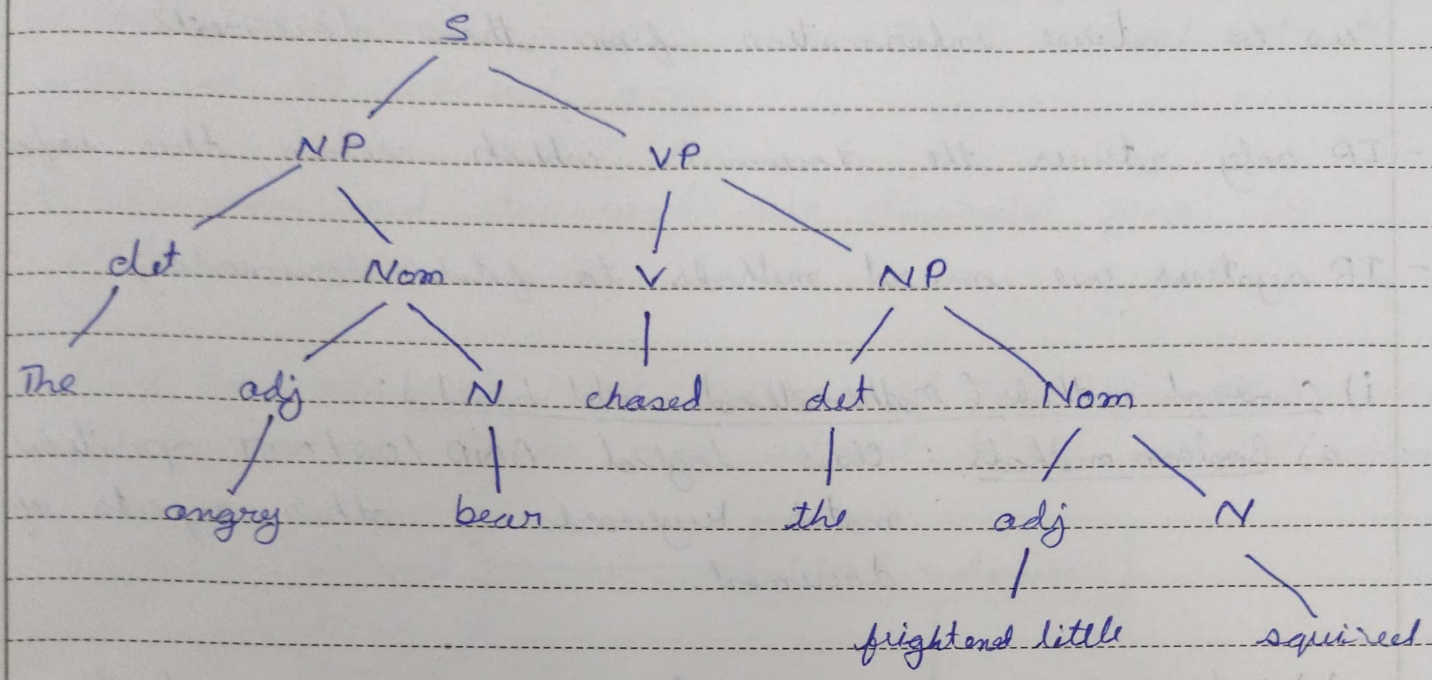
Sentence : "The angry bear chased the frightened little squirrel"

Grammar:

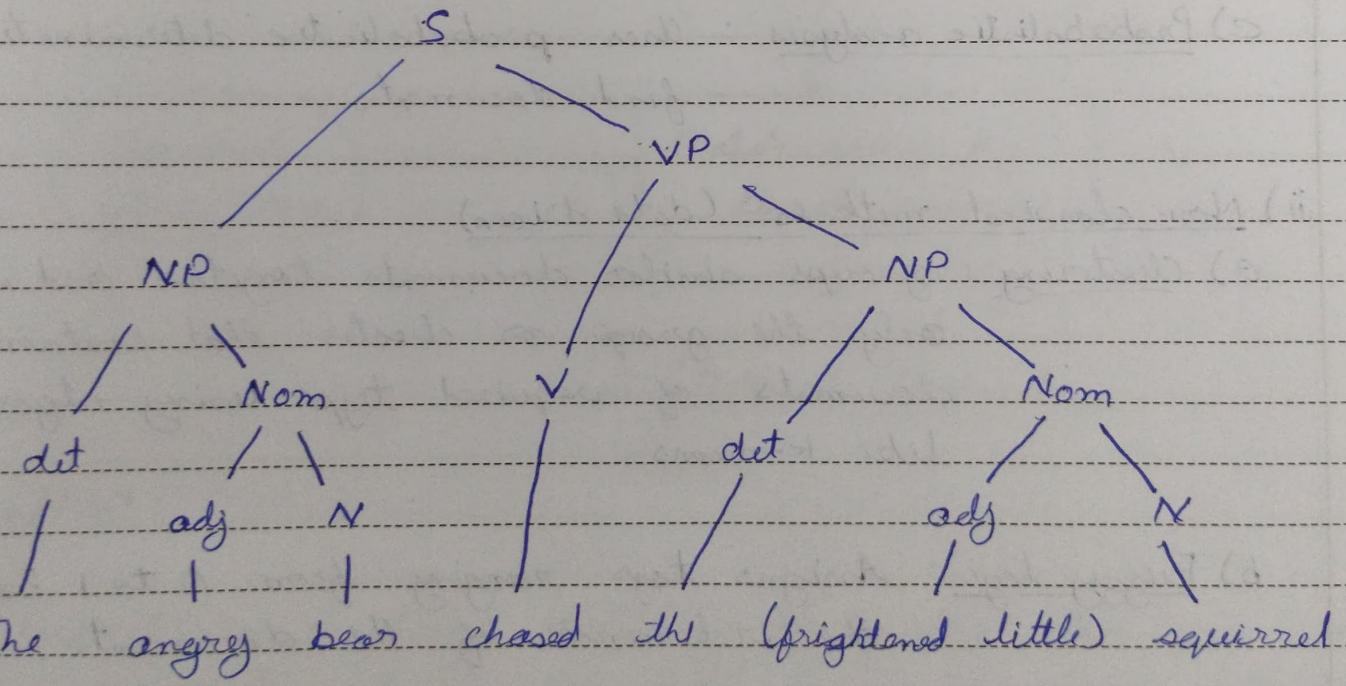
S → NP + VP
NP → Det + Nom
VP → V + NP
Nom → adj + N

Det → the
Adj → little / angry / frightened
N → squirrel / bear
V → chased

i) Top down parse tree (starts with S → NP + VP)



ii) Bottom up parse tree (starts at adj → little / angry / frightened)



5) Design features of IR systems:

- IR stands for information retrieval system.
- Given a database or library of documents, IR allows

us to retrieve information from those documents.

- IR only retrieves the documents which contain this information.
- IR systems use several methods to fetch documents;

i) Classical methods (Mathematical model based):

a) Boolean methods: Uses logical AND/OR/NOT operations to match keywords or other aspects of the document.

b) Vector space analysis: creates a vector space of the documents and the retrival information.

c) Probabalistic analysis: Uses probabalistic determination to find documents

ii) Non-classical methods (data driven)

a) Clustering: groups similar documents together and searches only the group or cluster that contains the documents of required type using algorithms like K-means

b) Fuzzy logic: Assigns tags ranging from 0 to 1 based on the relevance of the document.

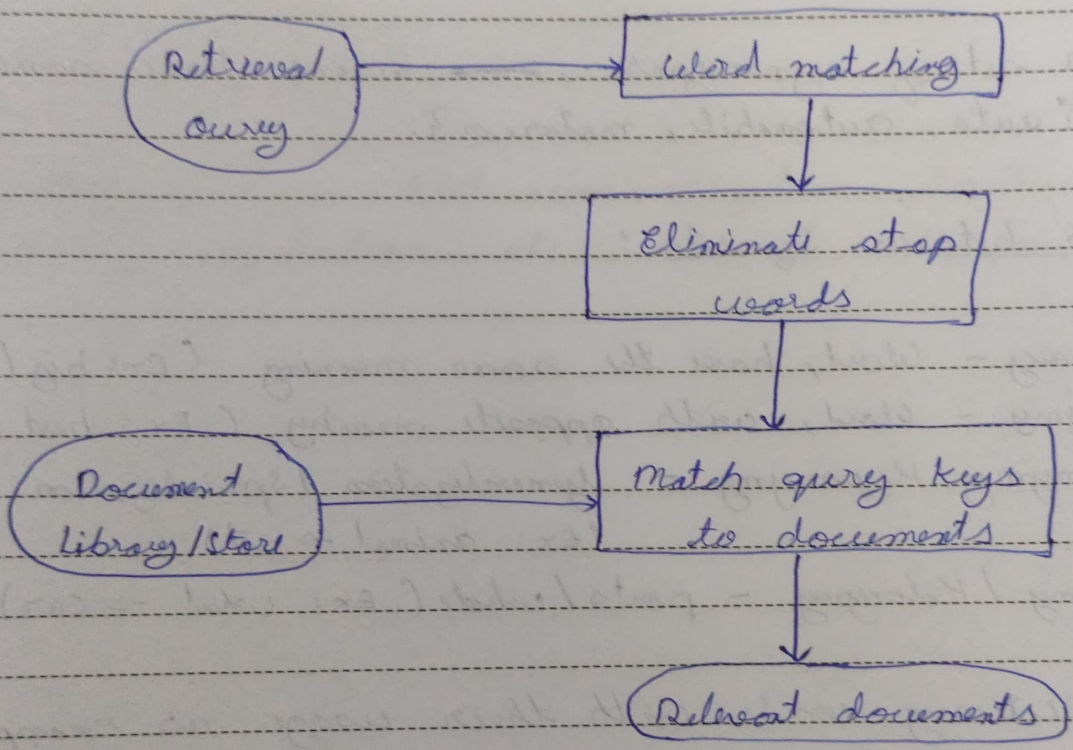
- Design Features

i) IR systems contain a word-key matching to find relevance of the documents to the retrieval requirements.

ii) Each word occurs in the document is matched with the key in the retrieval query.

iii) Determiners and stop words are eliminated from the retrieval query.

iv) The documents with more ocurenas of the information's keywords are assumed to be more relevant to the required information.



6) WordNet:

- WordNet is a large database of English, developed and maintained at the Cognitive Science Laboratory, Princeton University, under, Georg A. Miller.

- It is inspired by psycholinguistic theories of human lexical memory.

- It groups words as a set of synonyms known as "synsets".
- Each group represent a distinct meaning.

Key Concepts:

* Synsets:

- It is a set of synonyms, ~~mean~~ means have same meaning
- Ex: {auto, automobile, motorcar}.

* Relation between Synsets:

- Synonymy - Word, have the same meaning (Ex: big / huge)
- Antonymy - Word, with opposite meaning (Ex: heat / cold)
- Hypernymy / Hyponymy - Generalization / Specialization (Ex: animal → cat)
- Meronymy / Holonymy - parts / whole (Ex: wheel → car)

* Gloss: Dictionary words with their usage as example.

Example: run (verb) - move fast on Foot
run (verb) - run motors / parts
run (noun) - a sequence of operations

* Applications of WordNet:

i) Word Sense Disambiguation (WSD):

- Gives (gets the correct meaning of the words using context.

ii) Concept Identification:

- Identifies concept related to a word from the document, not only the meaning of the words.

iii) Query Expansion in IR:

- Facilitates expansion of queries for better search results.

iv) NLP Applications:

- Semantic analysis, text summarization, Question answering, ontology applications etc.