USN					



 $Internal\ Assessment\ Test\ 1-September\ 2025$ 

Sub	Natural Language	Processing			Sub Code:	BCS714B Branch: CSE			CSE			
Date:	29/09/2025	Duration:	50	Sem /Sec:	VII Profess	OBE						
	Answer any FIVE FULL Questions											RBT
1	Define NLP. List and explain applications of NLP.										CO1	L2
2	Consider the follo	wing corpus of	three sentences	S					[:	[01	CO1	L3
	There is a big											
	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.											
3	Explain Minimum Edit Distance algorithm. Compute minimum edit distance between "Tutor" and "Tumor".								[1	10]	CO2	L3
4	Derive a top down and bottom up parse tree for the given sentence								[1	10]	CO2	L3
	"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											
5	Explain design fe	atures of IR wit	h a neat diagrai	m.					[:	10]	CO4	L2
6	Explain WordNet	with its applica	tions.						[	10]	CO4	L2

CI CCI HOD-CSE



USN

Internal Assessment Test 1 – September 2025

Sub	Natural Language	Processing	Sub Code:	BCS714B	B Branch: CSE							
Date:	29/09/2025	Duration:	90 mins	50	Sem /Sec:	VII Profess	OBE					
		MA	RKS	CO	RBT							
1	Define NLP. List		[	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.										CO1	L3
3	Explain Minimum Edit Distance algorithm. Compute minimum edit distance between "Tutor" and "Tumor".										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 fe	atures of IR wit	th a neat diagra	m.					[	10]	CO4	L2
6	Explain WordNet	with its applica	ations.						[	10]	CO4	L2

CI CCI HOD-CSE

2/2	TAT Solution
1>	NLP- Natural Longuage Browsing:
Caro	NIP is a field of AI and Computation linguistics that consist with the development of a model for processing natural human larging.  It enables computers to interpret, understood and process the human language.  The main aims of sulp are:  - Durulap automated tools for natural language processing.  - Understanding the count of how human against a store.
363	- Understanding the concept of how human against store , and process the lines memory (longuage)
	i) Machine Translation:
	· Used to convert tent from one longuess into another.  (En Josef Translate)
	ii) Spach identification:
	· Used to convert speech into text form.
	· Exi siri, alens etc

	CMR
	iii) Speech synthesies:
	closed to connect appearts the tent to human readable
	Ex: Screen Recoder
	iv) Natural Longuege Interfoce to delaboses:
	· Database quireging in natural or simple english sentences
	· Ex Show Steeled from × 7th som
	v) Duta Information Redrival (IR):
	· Reterives the relevent information or documents.
	vi) Information Entraction:
	Entroits the structural foods (nomes, data, relation) from
	unstrudured information.
1)	given Sentenas:
	<s>There is a big gorden .</s> <s>Children play in a gorden .</s>
	< >> They play inside beautiful gardon < 15>
	Test sentence:
	They play in a big garden.



Probability wing Bi-gram modal
P(They play in a big garden) = P(They 1 < 2>) * P(play 1Thuy)  * P(in (play) * P(alian)  * P(big 1a) * P(gorden 1big)
3 x d x d x d x d x d
= 1 x 1 x 2 x 1 x 2 x 1
3 x 2 x 2
= 0.033 × 0.5 × 0.5
= 0.0825 P(They play in a big gordon) = 0.0825
Minimeson Edit Distoner Algorithmi
The minimum edit distance algorithm states that the minimum edit distance between two string is the minimum required operations to conveit one string into another.
· Operations Allocaed:
- Insertion (odd a letter) - Deletion (remove a letter)
- Substitution (oreplace a letter with mother)

3)



· Algorithm:
Inped: Two strings x and y
output: minimum edit distance blew n and y
$m \leftarrow length(x)$
n - length (g)
and the second s
Las iso to a 1
for i=0 to m do
$dist \Gamma i, OJ = i$
for j=0 to n do
dist Lo, i] - i
DE TRACTOR OF PRINTED AND PRINTED TO THE PRINTED TO
for i=a to m do
for 's-o tondo
dist [1,j] = min { dist [1-1,j] + insert_cost,
dist [i-1, j-1] + subset cool [ai, yi]
dist Li, j-1] + dehet cost ]
AL AND ONE OF THE PROPERTY OF
- The dit of its of the
· The distance is calculated using:
NEW MEMORIAL DESCRIPTION OF THE PROPERTY OF TH
did [i,j] = men & dist[t-1,j] + insert-cost,
dist[ [-1,j-1] 1 subset_cost[ni,yi],
did Li, j-1] + deld-cost 3
· Escople:
· Escople:
Escople:  Testos Vs Tiespos

		#	T	U	m	0	77	- disadisas
	#		1	12	3	4	5	
	7	1	0	1	2	3	4	
**********		2	1	0	I	2	3	
	1	3	2	1	1	2	3	
************	2	4	3	2	2	1	2	A STATE OF THE STA
	91	5	4	3	3	2	0	> minimeen edet
	371	3						distance
								a stracturate
						. //		" . 9 1 (Termos" - 1
The n	nenime	96	dit	distan	nelb	2Lw	Leven	" & 1( Tumos " = 2,
Tutor	-> rep	loce	7 1	oith.		-> L	umor	)
Top dou	20 L	Botto	202	p_po	use T	reo		
							<u>.</u>	
Sentense	The Th	e an	greg	bear.	chas	ed 14	o pe	ightened little oquired
	. ir	dit.I	Lalei	7611	1.00	200 = 2		
Gramma	Edm	+5-	1,1-	ink.ka	Ja			
	5		1	inti	Je			
SAM	YP+ Vt	)		D	wt ->	the		
NP -> 1	Del + M	om		A	li->	little	1 onas	ey I pright and
VP->	VANE	)					all be	
Nom->	oditi	<b>Y</b>			1-> 4			
i) Top de			tree					P+VP)
	V-	0					7	THE REPORT OF THE PARTY OF THE
	********			******				
	****					*******		



NP VP dt Nom v NP 1 The adj N chased det Nom angry bear the adj N ii) Bottom up pase tree (stails at ody > little longry / frightened) NP Nom Nom 5) Design features of IR systems: IR stands for information reteined system. - Given a database on library of documents, IR allows



as to retire information from those documents
- IR only retries the documents which contain this information
- IR systems use several methods to feth documents;
i) Clarical methods (mathematical model based):  a) Boolean methods: Class logical AND 10R1NOT operations to match keywords or other aspects of the document.
Link and little and little and little
b) vector space andersis: creates a vector space of the
b) vestor space and essis: creates a vestor space of the documents and the cretical information.
C) Probabalistie analyses: Uses probabalistic determination to
ii) Non: classical methods (data driven)  a) Clustering: groups similar documents together and searches  any the group or cleester that contains the  documents of required texpo resiner algorithms  libs k-means
b) Fuzzy logis: Assigns took ranging from 0 to 1 bosed on the relevance of the document.
Design Evateures
i) IR septems contain a work key matching to find relevance of the documents to the ordrieval organizements.



ii) Each word count in the downed is matched with the
Key in the retrieval and
key in the ordrieval query
III Determinated and stop words are climinated from
iii) Determiners and stop words are eliminated from the
iv) The documents with more occurences of the information's keywords
are assumed to be more relevant to the required infonction
(nt. 1)
( Relieval ) ( word matching)
(Retrieval) & word matching)
<u> </u>
Eliminate stop
words
Land Market & Committee of the Committee
Marie Daniel Marie Marie Marie Land - Special and a series of the second
[mth quite king]
(Document) match quiry kuys ) Library / Store ) to documents
Library Istory
Constant to the second of the
(Relevant documents)
WordNet:
- Walaixes
1 1 1 1 1 2 1 P 1 1 developed and
· WordNet is a longe database of English, developed and
main tained by at the Cognillal Science haboratory truncators
University, ander, Greary A. Miller
. It is inspired by psycholinguistic theories of human
La i I memorale
lexical memory.

6)



est of synanyms known as
- 2t groups words as a set of synonyms known as
sepasets"
· Each grown represent a disting meaning.
Each group represent a disting meaning.
Key Concepts:
key Concopts:
* Seports:
and have some means
I es a set of segmentyms, the most
· It is a set of synonyms, more means have some means.
* Relation between Synsets:
· Synonymy - Word, have the same meaning (Ex: big / huge)
· And onemy - word, with apposite meaning ( Ex: heat I cold)
· Hypernymy / Hyponymy - Generalization 1 Spainligation
(Ex: animal -> cat)
· morani / V / · · · · · · · · · · · · · · · · ·
· meronymy / Holonemy - ports/whole (Ex: what -> car)
" Islan: Dictionary words with their usage as example.
Example: over (verb) - more fost on Fad
( ) -
- Second Lafelet
run (noun) - a sequence of operations
t the County Have Selected to the transfer of
* Applications of Wadnet:
* Applications of Wadnet:
i) // d c c c D) bi do f
i) Word Suose Disambiguation (WSD):
· lyinon ( byle the correct meaning of the words coning content.
(a) me (an) end to





ii) Concept 3dentification:	
· Idetaples concept related to a word from the document, not	
only the meaning of the words.	
Til aug Expansion in IR:	
· Exilitates expossion of queries for better search results.	
in) MLP Applications:	
State of the state	
· Comentie analysies test seemmarisation, Occession answering	
· Semantic analysies, test seemmarization, Occession answering,	