Sardar Vallabhbhai National Institute of Technology, Surat
## Computer Science and Engineering Department
B Tech III(CSE) - Sixth Semester
CS324 - Natural Language Processing
End Semester Examination, May 2022

Time : 12.00pm to 3.00 pm   Date : 05-05-2022   Marks: 100

Q1 **a.** Write regular expressions for the set of all strings with two consecutive repeated words (e.g., "Humbert Humbert" and "the the" but not "the bug" or "the big bug"). **[8 Marks]**

b. Give examples of one noun and one verb in the above figure, and find Two exceptions to the rules. Eg. verbj : detect accepts -ive but it becomes a noun, not an adjective
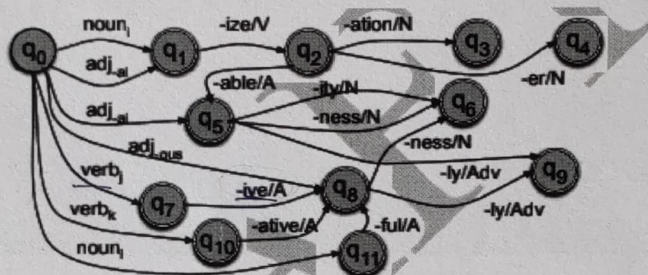
Q2 Answer any **Four**.   **[40 Marks]**

a. Compute minimum edit distances, figure out whether drive is closer to brief or to divers and what the edit distance is using 1-insertion, 1-deletion, 2-substitution costs.

b. Consider the mini-corpus as below:

\<s> I am working as a Data scientist in my company\</s>
\<s> My company provides lots of benefits \</s>
\<s> As a Data scientist there is tremendous growth in my company. \</s>

Compute the Bigram probability of sentence "Working as a Data Scientist benefits my growth in company"

c.

Transition probability matrix

	Winter	Summer	Monsoon
\<s>	0.4	0.2	0.4
Winter	0.6	0.3	0.1
Summer	0.4	0.4	0.2
Monsoon	0.1	0.4	0.5

Emission probability matrix

	Cold	Hot	Rain
Winter	0.8	0.1	0.10
Summer	0.10	0.7	0.2
Monsoon	0.10	0.2	0.70

The above said matrix consists of emission probability values represented as bi(ot). bi(ot) is the probability of an observation ot generated from a state bi. For example, P(Cold | Winter) = 0.8, P(Rain | Monsoon) = 0.70 etc.

π = [π1, π2, ..., πN] = set of prior probabilities = [0.6, 0.3, 0.1]. Here, the values refer to the probabilities P(Winter) = 0.6, P(Summer) = 0.3, and P(Monsoon) = 0.1. What would be the state sequence for observation (Hot Rain) ? Do the calculations using Vertribi algorithm.

d. Use the below tagset to tag each word in the following sentences. You may ignore punctuation.

1. Nobody ever takes the newspapers she sells
2. I am sitting in Jenny's restaurant putting on the dessert, which I am very fond of.

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	and, but, or	SYM	Symbol	+,%, &
CD	Cardinal number	one, two, three	TO	"to"	to
DT	Determiner	a, the	UH	Interjection	ah, oops
EX	Existential 'there'	there	VB	Verb, base form	eat
FW	Foreign word	mea culpa	VBD	Verb, past tense	ate
IN	Preposition/sub-conj	of, in, by	VBG	Verb, gerund	eating
JJ	Adjective	yellow	VBN	Verb, past participle	eaten
JJR	Adj., comparative	bigger	VBP	Verb, non-3sg pres	eat
JJS	Adj., superlative	wildest	VBZ	Verb, 3sg pres	eats
LS	List item marker	1, 2, One	WDT	Wh-determiner	which, that
MD	Modal	can, should	WP	Wh-pronoun	what, who
NN	Noun, sing. or mass	llama	WP$	Possessive wh-	whose
NNS	Noun, plural	llamas	WRB	Wh-adverb	how, where
NNP	Proper noun, singular	IBM	$	Dollar sign	$
NNPS	Proper noun, plural	Carolinas	#	Pound sign	#
PDT	Predeterminer	all, both	"	Left quote	' or "
POS	Possessive ending	's	"	Right quote	' or "
PRP	Personal pronoun	I, you, he	(	Left parenthesis	[, (, {, <
PRP$	Possessive pronoun	your, one's	)	Right parenthesis	],), }, >
RB	Adverb	quickly, never	,	Comma	,
RBR	Adverb, comparative	faster	.	Sentence-final punc	. ! ?
RBS	Adverb, superlative	fastest	:	Mid-sentence punc	: ; ... -- -
RP	Particle	up, off			

e. Consider the dataset. Using Naïve Bayes Model, Decide whether a person with Medium Income would be able to buy a laptop or not. Do all the calculations.

	Income	Buys Computer
0	High	No
1	Low	Yes
2	Medium	No
3	Medium	Yes
4	Low	No
5	High	Yes
6	Low	Yes
7	Medium	Yes
8	High	No
9	Low	No
10	Low	Yes
11	High	No
12	Medium	Yes
13	Medium	Yes

Q3 ① Explain the difference between coreference resolution and pronominal anaphora resolution with an example. [12 Marks]

2. Which Discourse relation is present in the below sentences.
   a. John was from Kansas. He lived on the Kansas prairies.
   b. Ram was stuck in the rain. He caught cold.

3. What is Rule based Machine translation? What problems are faced during it? List down benefits of Phrase-based SMT.

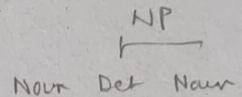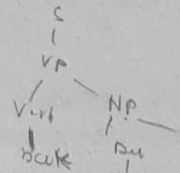**Q4** ⓐ Define the notation: Constituency with examples of Constituent Phrases. Justify the existence of Constituency. [18 Marks]
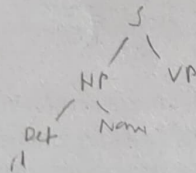
ⓑ What is parsing in the syntax analysis phase. Use the following grammar to parse the sentence "book the flight" using a top down approach.

Grammar		Lexicon	
S	→ NP VP	Det	→ the | a | that | this
S	→ Aux NP VP	Noun	→ book | flight | meal
S	→ VP	Verb	→ book | include | prefer
NP	→ Pronoun	Pronoun	→ I |he | she | me
NP	→ Proper-Noun	Proper-Noun	→ Houston | NWA
NP	→ Det Nominal	Aux	→ does
Nominal	→ Noun	Prep	→ from | to | on | through
Nominal	→ Nominal Noun		
Nominal	→ Nominal PP		
VP	→ Verb		
VP	→ Verb NP		
VP	→ VP PP		
PP	→ Prep NP		

ⓒ Explain Probabilistic Context Free Grammar(PCFG). Find out the probability for the given sentence "astronomers saw stars with ears" using a simple PCFG.

S	→ NP VP	1.0		NP	→ NP PP	0.4
VP	→ V NP	0.7		NP	→ astronomers	0.1
VP	→ VP PP	0.3		NP	→ ears	0.18
PP	→ P NP	1.0		NP	→ saw	0.04
P	→ with	1.0		NP	→ stars	0.18
V	→ saw	1.0		NP	→ telescope	0.1
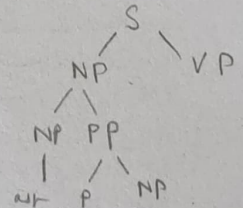
Q.5 Discuss Semantic Analysis Phase of Natural Language Processing. Explain Distributional semantics method and compute similarities between the following target words using simple vector products for the given word space. [8 Marks]
   1. automobile . car
   2. automobile . soccer
   3. automobile . football
   4. car . soccer
   5. car . football
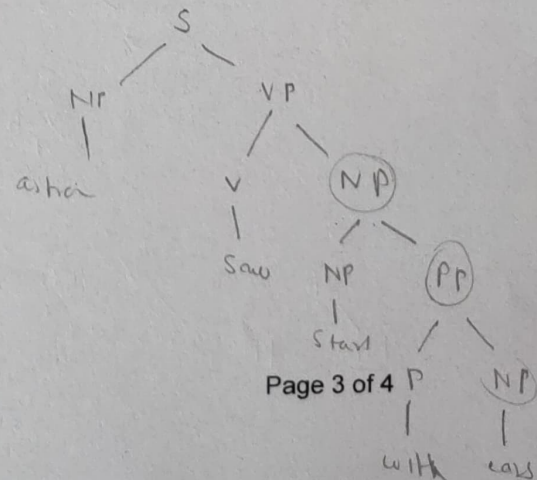
6. soccer . football

	wheel	transport	passenger	tournament	London	goal	match
automobile	1	1	1	0	0	0	0
car	1	2	1	0	1	0	0
soccer	0	0	0	1	1	1	1
football	0	0	1	1	1	2	1

Q6 Define Inside and Outside Probabilities for Probabilistic Context Free Grammars. Consider the sentence "The gunman sprayed the building with bullets", and define inside probability $\alpha_{NP}(4,5)$ and outside probability for $\beta_{NP}(4,5)$. Calculate inside probability for $\beta_{NN}(5,5)$. [Assume the rule: NN → building (0.5)] **[6 Marks]**

N^1

NP

| The 1 | gunman 2 | sprayed 3 | the 4 | building 5 | with 6 | bullets 7 |
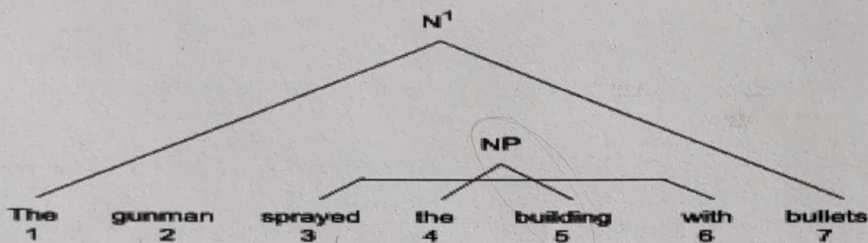
Q7 List out Dynamic Programming Parsing methods. Consider the given grammar, convert into Chomsky Normal Form and generate possible parse trees using CKY Parser method for the sentence "Book the flight through Houston". **[8 Marks]**

Grammar

S	→ NP VP	$S \to X_1\ VP$
~~S~~	→ Aux NP VP	$X_1 \to Aux\ NP$
~~S~~	→ VP	$S \to VP\ PP$
NP	→ Pronoun	
NP	→ Proper-Noun	
NP	→ Det Nominal	
Nominal	→ Noun book	flight	
Nominal	→ Nominal Noun	
Nominal	→ Nominal PP	
VP	→ Verb → a	book	include	prefer
VP	→ Verb NP ✓	
VP	→ VP PP ✓	
PP	→ Prep NP	

Lexicon

Det	→ the	a	that	this
Noun	→ book	flight	meal
Verb ✓	→ book	include	prefer
Pronoun	→ I	he	she	me
Proper-Noun	→ Houston	NWA
Aux	→ does
Prep	→ from	to	on	through