BACHELOR OF TECHNOLOGY (CBCS) (2020 COURSE)
B.Tech.Sem - VI Computer Science & Engineering : WINTER : 2023
SUBJECT : NATURAL LANGUAGE PROCESSING

Day : Tuesday
Date : 21-11-2023

Time : 10:00 AM-01:00 PM
W-24330-2023
Max. Marks : 60

N.B.:
1) All questions are **COMPULSORY**.
2) Figures to the right indicate **FULL** marks.
3) Use of non-programmable **CALCULATOR** is allowed.

Q.1 Describe various models and algorithms of NLP. **(10)**

OR

Define perplexity? How it is calculated for below example. **(10)**
<S> I am Henry</S>
<S> I like college </S>
<S> Do Henry like college </s>
<S> Henry I am </S>
<S> Do I like Henry </S>
<S> Do I like college </S>
<S> I do like Henry </S>
1) Calculate perplexity for bigram model
<S> I like college </S>.
2) Calculate perplexity for trigram model
<S> I like college </S>

Q.2 Discuss various stages of text preprocessing. **(10)**

OR

Define Finite Automata? Design DFA for **(10)**
1) Q ={a, b, c}, q_0 = {a}, Σ= {0. 1}. F={C}
2) Σ= {0, 1} accept those strings which starts with 1 and ends with 0.

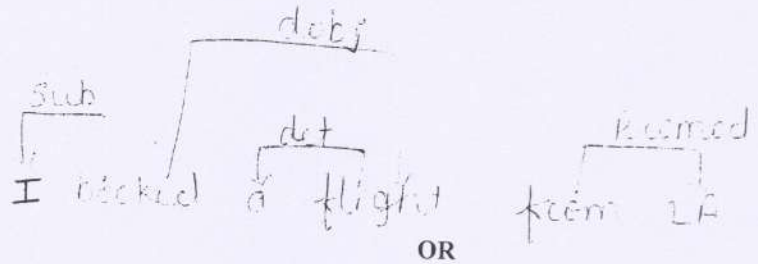
Q.3 Discuss in detail architecture of speech recognition system. **(10)**

OR

Determine the correct POS tag for word using HMM model. (Assume random values of word for checking). **(10)**

Sacretariat	is	expected	to	Race	tomorrow
NNP	VBZ	VBN	To	VB	NP
				NN	

Q.4 Define concept of Parsing? Also explain concept of Transition based dependency parser. Solve below example using Transition based dependency parser. **(10)**



OR

Use cky algorithm for PCFG to find most probable parse tree. (10)
Sentence: "A Pilot likes flying plane."
CFG Rules with the probabilities-

S → NP VP	[1.0]
VP → VBG NNS	[0.20]
VP → VBZ VP	[0.05]
VP → VBZ VP	[0.30]
NP → DT NN	[0.02]
NP → JJ NNS	[0.01]
DT → a	[0.20]
VBZ → likes	[0.10]
NN → pilot	[0.005]
VBG → flying	[0.02]
JJ → flying	[0.001]
NNS → planes	[0.10]

Q.5 Discuss the concept of Matrix factorization. Also explain the concept of singular value decomposition with suitable example. (10)

OR

Define concept of vector space model. Also solve below example to find Cosine similarity of document (10)
D_1= (0.8, 0.7, 0.2)
D_2= (0.5, 0.6, 0.4)
Q= (1.3, 1.0, 0)

Q.6 How linguistic data is managed with the help of GATE. Explain it in detail. (10)

OR

Write in detail about NLTK library of Python. (10)

* * * * *