Continuous Assessment Test (CAT) – I (August 2024)

Programme	:	B. Tech. Computer Science and Engineering	Semester	:	**Fall Sem 24-25**
Course Code & Course Title	:	BCSE409L Natural Language Processing	Class Number	:	CH2024250101611 CH2024250101619 CH2024250101621 CH2024250101636 CH2024250101638
Faculty	:	**Dr.Manjula.D** **Dr.R.Krithiga** **Dr. Lakshmi Harika Palivela** **Dr.S.Sharmiladevi** **Dr.Gayathri.R**	Slot	:	B2+TB2
Duration	:	1½ Hours	Max. Mark		50

General Instructions:
- Write only your Reg. No. on the question paper in the box provided and do not write other information.
- Use statistical tables supplied from the exam cell as necessary
- Use graph sheets supplied from the exam cell as necessary
- Only non-programmable calculator without storage is permitted

Answer all questions

Q. No	Sub Sec.	Description	Marks
1		Read the following sentences and illustrate the different types of ambiguities present in each: a) *The bank was crowded.* b) *Visiting relatives can be annoying.* c) *John said he would give a ride to Tom, but he didn't show up.* d) *The professor discussed the student's paper in the office.* e) *The chicken is ready to eat.*	10
2		Assume that you are developing a text analysis system for a historical research application. This system needs to process and interpret texts related to significant events and figures from history or mythology. One such text is about the historical figure Krishna and his actions. Consider the following historical narrative: *In the epic Mahabharata, Krishna played a crucial role in the Kurukshetra battle and is renowned for killing the warrior Karna, a pivotal moment in the conflict.* Illustrate the various phases of Natural Language Processing (NLP) applied to the given scenario.	10

3		Consider the following sentences: • *It is going to sunny today.* • *I am going to watch the season premiere.* • *Today I am not going Chennai.* • *Chennai is very hot today.* • *"The quick brown fox jumps over the lazy dog. The dog, which was very lazy, slept all day."* List the steps required to tokenize these sentences and obtain the tokens. Calculate the type-token ratio for each Sentence.	10
4		Consider the following text corpus: • *The cat eats fish.* • *The dog barks.* • *A cat plays.* • *A dog eats.* • *The cat sleeps.* • *The dog runs.* • *A cat jumps.* Compute the emission and transition probability for the given corpus.	10
5		For the following transformations, draw a two-level transducer diagram illustrating the process: a. *Take → Taking* b. *Fire → Fires* c. *Large → Larger* d. *Bake → Baked* For each word transformation, identify which rule is applied from the base form. Specifically, explain the rule for e-insertion or retention, and describe how it affects the transformation process for the chosen example.	10

**********All the best ************

Continuous Assessment Test (CAT) – II (October 2024)

Programme	:	B. Tech. Computer Science and Engineering	Semester	:	Fall Sem 24-25
Course Code & Course Title	:	BCSE409L Natural Language Processing	Class Number	:	CH2024250101603 CH2024250101604 CH2024250101605 CH2024250101608
Faculty	:	Dr.Manjula.D Dr.R.Krithiga Dr. Lakshmi Harika Palivela Dr.Gayathri.R	Slot	:	B1 + TB1
Duration	:	1½ Hours	Max. Mark		50

General Instructions:
- Write only your Reg. No. on the question paper in the box provided and do not write other information.
- Use statistical tables supplied from the exam cell as necessary
- Use graph sheets supplied from the exam cell as necessary
- Only non-programmable calculator without storage is permitted

Answer all questions

Q. No	Sub Sec.	Description	Marks
1		Consider the following sentences. Det Adj N Det Adj N **Sentence 1:** The big dog and the small cat. [5 Marks] Det Adj N P Det N **Sentence 2:** A small cat in the box. [5Marks] Use the PCFG algorithm to find the probability for generating the sentence. NP -> Det (Adj)* N (PP)* | Det (Adj)* N (and NP)* | Quantifier NP Det -> the | a | an Adj -> big | small | old | new N -> dog | cat | book | table |box PP -> P NP P -> in | on | at Quantifier -> every | some | all NP -> Det (Adj)* N (PP)* [0.6] NP -> Det (Adj)* N (and NP)* [0.3] NP -> Quantifier NP [0.1] Det -> the [0.4];Det -> a [0.3];Det -> an [0.3] Adj -> big [0.3];Adj -> small [0.2];Adj -> old [0.3];Adj -> new [0.2] N -> dog [0.2];N -> cat [0.2];N -> book [0.3];N -> table [0.2]; N ->box[0.1] PP -> P NP [1]; P -> in [0.3];P -> on [0.3]; P -> at [0.4] Quantifier -> every [0.3];Quantifier -> some [0.4];Quantifier -> all [0.3]	10

2	a) b) c)	You're given the following sentence: " Kevin throws the ball in the basket." N V D N P Det N Identify the **head** of the sentence and list all the **dependent words** for the head. Describe the relationship each dependent has with the head.[2Marks] **Draw a dependency tree** and a **constituency tree** to visually represent how the words in the sentence depend on one another.[5 Marks] Compare both trees by highlighting the key differences between them. [3 Marks]	10

3		Given the following corpus, calculate the TF-IDF score for the given sentences: Sentence 1: "Software architecture is key to creating scalable systems." Sentence 2: "User interface design patterns improve usability and user experience". Sentence 3: "Books on software design often cover architectural patterns in detail". Sentence 4: "Effective software architecture includes patterns for both backend and frontend design".	10

4	a) b)	Identify the semantic roles for all the nouns in the sentences and also give justification. [3*2=6 Marks] 1. The teacher thinks we should draw trees in pencil. 2. Have you heard? The tornado last night destroyed the bridge to school 3. Can you give this note to Jim for me? Construct the Predicate argument structure for PRP N V D] N PRP N " The mam bought the silk in Kanchipuram " [4 Marks]	10

5		Consider the tables which shows the Unigram and bigram counts of a document also it contains a total of 2641 vocabularies. Estimate the bigram probability for the word sequence and also demonstrate add-one smoothing techniques to normalize the word. Calculate the probability of the occurrence of *"VIT students are very sharp and Intelligent"*.	10

Unigram counts are given here:

VIT	students	are	very	sharp	and	Intelligent
2545	794	1089	189	256	1221	257

Bigram counts are given below:

	VIT	students	are	very	Sharp	and	Intelligent
VIT	5	447	0	9	0	0	0
students	1	0	386	9	1	7	6
are	9	0	18	286	3	0	11
very	0	0	4	0	28	4	29
Sharp	5	0	0	0	0	46	1
and	16	0	16	0	1	5	0
Intelligent	2	0	0	0	0	1	0

**********All the best ************