

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	: 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		<p>Given the following context-free grammar (CFG):</p> <ol style="list-style-type: none"> 1. $S \rightarrow NP VP$ 2. $VP \rightarrow V NP \mid V NP PP$ 3. $NP \rightarrow Det N \mid Det Adj N$ 4. $PP \rightarrow P NP$ 5. $Det \rightarrow "the" \mid "a"$ 6. $N \rightarrow "dog" \mid "cat" \mid "park"$ 7. $Adj \rightarrow "big"$ 8. $V \rightarrow "saw" \mid "chased"$ 9. $P \rightarrow "in" \mid "with"$ <p>Use the Chomsky Normal Form Parser to construct a parse tree for the below sentence: "The big dog saw the cat in the park"</p>	10
2	a) b)	<p>Discuss the use of Conditional Random Fields (CRFs) in natural language processing tasks like Named Entity Recognition (NER), part-of-speech tagging, and chunking. [5 Marks]</p> <p>Explain important challenges while building a dependency parse tree with suitable examples. [5 Marks]</p>	10

You have a dataset consisting of text documents that mention various countries, along with their known development statuses. However, some country names may appear in different contexts, leading to ambiguities in classification. For example, the word "Canada" could refer to a country known for its high development status or be part of a phrase that describes a collaborative project in a developing context. Apply a suitable text classification approach to predict the class.

Documents	Words	Class
1	USA Canada Germany Japan	Developed
2	India Brazil Africa China	Developed
3	Canada Australia UK France	Developing
4	Brazil India Mexico Argentina	Developing
5	Nigeria Ethiopia Afghanistan Pakistan	Underdeveloped
6	Ethiopia Sudan Yemen Bangladesh	Underdeveloped
7	Germany France Japan Sweden	Developing
8	Canada India Brazil Africa Mexico	?

Explain how concepts are represented in a Wordnet. Illustrate how it can be used for word sense disambiguation with suitable example.

Calculate the bigram and trigram of the following sequences,

<S>Data science drives modern industries </S>

<S>Machine learning powers data models </S>

<S>Artificial intelligence enhances user experience </S>

<S>Deep learning improves prediction accuracy</S>

<S>Data analytics informs business decisions</S>