DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING AND INFORMATION TECHNOLOGY
YEAR/SEM: III/VI
23CSX503, NATURAL LANGUAGE PROCESSING FUNDAMENTALS
UNIT I
INTRODUCTION

QUESTION BANK

Natural Language Processing tasks in syntax, semantics, and pragmatics **-** challenges of NLP - NLP Phases - Language Modeling: Grammar-based LM, Statistical LM - Regular Expressions, Finite-State Automata – English Morphology, Transducers for lexicon and rules, Tokenization, Detecting and Correcting Spelling Errors, Minimum Edit Distance.

PART - A (2 MARKS)

Introduction
1.	Define Natural Language Processing (NLP).	K1
2.	State the applications of NLP.	K1
3.	Name the two main components of NLP.	K1

Natural Language Processing Tasks in Syntax, Semantics, and Pragmatics
4.	What is meant by pragmatics?	K1
5.	Why is POS tagging important in syntactic analysis?	K2
6.	How does lack of training data limit NLP performance?	K2

Challenges of NLP
7.	List any two challenges faced by NLP systems.	K1
8.	Why does ambiguity affect the accuracy of NLP systems?	K2

NLP Phases
9.	Why can a sentence be syntactically correct but semantically incorrect?	K2
10.	List the five phases of NLP.	K1

Language Modeling
11.	Define language modeling.	K1
12.	How does a statistical language model predict the next word?	K2
13.	Why do grammar-based language models have limited coverage?	K2

Regular Expression-Finite-State Automaton
14.	Why are regular expressions considered greedy?	K2
15.	How do character classes simplify pattern matching?	K2
16.	How does an FSA recognize a valid string?	K2

English Morphology
17.	What is a morpheme?	K1
18.	List the types of affixes.	K1
19.	How does derivational morphology differ from inflectional morphology?	K2

Tokenization
20.	Define tokenization.	K1
21.	What is text normalization?	K1

VISION: To produce demand driven, quality conscious and globally recognized computer professionals through education, innovation and collaborative research.

| 22. | Why must dates and numbers be treated as single tokens? | | K2 |
| 23. | How does subword tokenization reduce the OOV problem? | | K2 |

Detecting and Correcting Spelling Errors

24.	Define spelling error detection in NLP.		K1
25.	List any two causes of spelling errors.		K2
26.	How do non-word errors differ from real-word errors?		K2

Minimum Edit Distance

| 27. | List the three basic edit operations used in MED. | | K1 |
| 28. | Why is dynamic programming used to compute Minimum Edit Distance? | | K2 |

Transducers for Lexicon and Rules

| 29. | What is the role of a lexicon in NLP? | | K1 |
| 30. | How does a finite-state transducer differ from a finite-state automaton? | | K2 |

PART - B (16 Marks)

1.(i)	Discuss the tasks of syntax, semantics and pragmatics in NLP.	6	K2
1.(ii)	Examine the challenges of NLP and discuss their impact on system performance.	10	K2
2.(i)	Elaborate on the different phases of NLP and their significance in language processing.	10	K2
2.(ii)	Describe the applications of NLP.	6	K2
3.(i)	Apply grammar-based language modeling to parse a sentence using CFG rules.	8	K3
3.(ii)	Apply a bigram language model to calculate the probability of the sentence: "I like to eat apples" using the provided corpus.	8	K3
4.	Elaborate on morphological analysis and its importance in NLP applications.	16	K2
5.	Explain different tokenization approaches used in NLP systems.	16	K2
6.(i)	Evaluate how finite-state automata and regular expressions correspond to each other in language representation.	8	K4
6.(ii)	Analyze the different types of spelling error detection and correction problems in NLP and explain how each requires different handling techniques.	8	K4
7.(i)	Construct a finite-state automaton (FSA) to recognize the regular expression /baa+!/ and explain how it accepts valid strings.	8	K3
7.(ii)	Apply the Normalization techniques for the given text and comment on the vocabulary before and after the normalization: Raj and Vijay are best friends. They play together with other friends. Raj likes to play football but Vijay prefers to play online games. Raj wants to be a footballer. Vijay wants to become an online gamer.	8	K3
8.	Analyze the working of the Edit Minimum Distance algorithm and explain why dynamic programming is suitable for solving it.	16	K4

VISION: To produce demand driven, quality conscious and globally recognized computer professionals through education, innovation and collaborative research.