

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK



VII SEMESTER

CS3761 – NATURAL LANGUAGE PROCESSING

Regulation – 2023

Academic Year 2026-2027 (Odd Semester)

Prepared by

Dr.C.Pabitha, Associate professor

Ms. Shanthi. S, Assistant Professor (Sel.G.)

Ms. S. Anslam Sibi, Assistant Professor (Sr. G.)



SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203.



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

SUBJECT : CS3761 -Natural Language Processing

YEAR/SEM: IV/VII

UNIT I INTRODUCTION

Origins and challenges of NLP – 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

Q. No.	Questions	BT Level	Competence
1	List the origin of NLP.	BTL 1	Remembering
2	Define Natural Language processing.	BTL 1	Remembering
3	List the challenges of NLP.	BTL 1	Remembering
4	What are the challenges of NLP for Artificial Intelligence?	BTL 1	Remembering
5	Differentiate between human and machine language processing.	BTL 2	Understanding
6	Explain the preprocessing of text.	BTL 2	Understanding
7	Define Language modelling	BTL 1	Remembering
8	What is pattern?	BTL 1	Remembering
9	Explain the importance of language modelling.	BTL 2	Understanding
10	What is a regular expression?	BTL 1	Remembering
11	What is Finite State Automata?	BTL 1	Remembering
12	Give how the morphological rules differentiates from the orthographic rules?	BTL 2	Understanding
13	Explain the relationship between Finite Automata, Regular Grammars, and Regular Expressions.	BTL 2	Understanding
14	Explain the concept of minimum edit distance using the words INTENTION and EXECUTION.	BTL 2	Understanding
15	What are the three problems in detecting and correcting spelling errors.	BTL 2	Understanding
16	List the seven parameters of Finite State Transducers (FST).	BTL 1	Remembering
17	Define the term Lexicon	BTL 1	Remembering
18	Compare and contrast the Substitution, Capture Groups, and ELIZA	BTL 2	Understanding
19	Explain the concepts of disjunction, grouping, and precedence.	BTL 2	Understanding
20	What are the types of text normalization?	BTL 2	Understanding
21	Explain Finite State Transducers as generators.	BTL 2	Understanding
22	What is sentence segmentation?	BTL 1	Remembering
23	Explain the concept of minimum edit distance.	BTL 2	Understanding
24	List the English verbal inflections.	BTL 1	Remembering

PART – B

Q. No.	Questions	Marks	BT Level	Competence
1	(i) Apply the steps involved in text pre-processing to a given text.	8	BTL3	Applying
	(ii) Illustrate the importance and applications of language modelling with suitable examples.	8		
2	(i) Apply the concept of language modelling in NLP.	8	BTL3	Applying
	(ii) Illustrate how language modelling works using a suitable example.	8		
3	Analyze grammar-based language modelling and its role in NLP.	16	BTL4	Analyzing

4	Analyze the role of Artificial Intelligence and the challenges of NLP in AI applications.	16	BTL4	Analyzing
5	Analyze the various types of statistical-based language modelling techniques.	16	BTL4	Analyzing
6	(i) Analyze Regular Expressions, Disjunction, Grouping, and Precedence.	8	BTL4	Analyzing
	(ii) Apply Kleene Closure using suitable examples of Regular Expressions.	8		
7	Analyze the various types of Finite Automata.	16	BTL4	Analyzing
8	Apply the three tasks commonly used in text normalization with suitable examples.	16	BTL3	Applying
9	Evaluate the Minimum Edit Distance algorithm and justify its effectiveness using a suitable example.	16	BTL5	Evaluating
10	Construct regular expressions using appropriate operators for given patterns with suitable examples.	16	BTL6	Creating
11	Evaluate Inflectional Morphology and its significance in Natural Language Processing.	16	BTL5	Evaluating
12	(i) Apply Morphological Parsing to analyze given words.	8	BTL3	Applying
	(ii) Illustrate the different types of morphemes with suitable examples.	8		
13	Evaluate the process of finding the Minimum Edit Distance and compute the edit distance between INTENTION and EXECUTION.	16	BTL5	Evaluating
14	Apply the concepts of Inflectional and Derivational Morphology to distinguish between them with examples.	16	BTL3	Applying
15	Design an approach for Word Tokenization and Sentence Tokenization with suitable examples.	16	BTL6	Creating
16	Evaluate Sequential Transducers and Determinism in NLP with suitable examples.	16	BTL5	Evaluating
17	Evaluate parsing using FST Lexicon and Rules, highlighting its advantages and limitations.	16	BTL5	Evaluating

	UNIT II WORD LEVEL ANALYSIS			
	Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.			
	PART – A			
Q. No.	Questions	BT Level	Competence	
1	Define language modelling.	BTL 1	Remembering	
2	Describe n-gram language models.	BTL 2	Understanding	
3	Illustrate how to determine the probability of the next word for the history "its water is so transparent that" using an n-gram language model.	BTL 2	Understanding	
4	Explain the Markov assumption.	BTL 2	Understanding	
5	State the chain rule of probability.	BTL 1	Remembering	
6	Define Maximum Likelihood Estimation (MLE).	BTL 1	Remembering	
7	Explain why Part-of-Speech (POS) tagging is difficult.	BTL 2	Understanding	
8	Differentiate between Hidden Markov Model (HMM) and Maximum Entropy Markov Model (MEMM).	BTL 2	Understanding	
9	State the Maximum Likelihood formula.	BTL 1	Remembering	
10	Define log probability.	BTL 1	Remembering	
11	Define Perplexity	BTL 1	Remembering	
12	Differentiate between open-class and closed-class words.	BTL 2	Understanding	
13	Differentiate between word likelihood probabilities and tag transition probabilities.	BTL 2	Understanding	
14	Explain Part-of-Speech (POS) tagging.	BTL 2	Understanding	
15	Explain the accuracy of a POS tagging algorithm.	BTL 2	Understanding	
16	Define a capture group with an example.	BTL 1	Remembering	
17	Define a tag set with an example.	BTL 1	Remembering	

18	Describe the various types of smoothing.	BTL 2	Understanding
19	Discuss smoothing in n-gram counts.	BTL 2	Understanding
20	Outline the objective of HMM.	BTL 2	Understanding
21	Assess the absolute discounting	BTL 2	Understanding
22	Illustrate the Viterbi algorithm.	BTL 2	Understanding
23	Differentiate between bigram and trigram models.	BTL 2	Understanding
24	Summarize the Bigram assumption.	BTL 2	Understanding

PART – B

Q. No.	Questions	Mark	BT Level	Competence
1	Apply the concept of N-gram language models to construct unigram, bigram, and trigram models using a suitable example.	16	BTL3	Applying
2	Compare and evaluate the intrinsic and extrinsic evaluation methods of language models with suitable examples.	16	BTL5	Evaluating
3	Analyze perplexity in language modelling and examine its relationship with the information-theoretic concept of entropy.	16	BTL4	Analyzing
4	Differentiate the various N-gram smoothing techniques and determine their effectiveness in language modelling.	16	BTL4	Analyzing
5	Apply Laplace smoothing and Add-k smoothing to estimate probabilities for unseen N-grams using suitable examples.	16	BTL3	Applying
6	Analyze Interpolation and Back-off techniques and compare their performance in handling sparse data.	16	BTL4	Analyzing
7	Classify the different word classes and examine their significance in Natural Language Processing.	16	BTL4	Analyzing
8	Apply Part-of-Speech (POS) tagging to annotate a sentence and justify the assigned tags.	16	BTL3	Applying
9	Analyze the Rule-Based POS Tagging approach and assess its advantages and limitations.	16	BTL5	Evaluating
10	Apply stochastic POS tagging to determine tag sequence probabilities for a given sentence.	16	BTL3	Applying
11	Evaluate the techniques used for tagging unknown words and justify their effectiveness in POS tagging.	16	BTL5	Evaluating
12	Analyze the Transformation-Based Learning (TBL) tagger and evaluate its advantages and disadvantages.	16	BTL5	Evaluating
13	Analyze the Hidden Markov Model (HMM) for Part-of-Speech tagging and examine its working with a suitable example.	16	BTL4	Analyzing
14	Design a Maximum Entropy Markov Model (MEMM) for sequence labeling and demonstrate its working using an example.	16	BTL6	Creating
15	Examine how the Maximum Entropy Markov Model (MEMM) is formulated using the log-linear model and analyze its characteristics.	16	BTL4	Analyzing
16	Develop a language modelling framework by incorporating Maximum Likelihood, Log-linear Model, and Log-likelihood Function for probability estimation.	16	BTL6	Creating
17	Construct a bigram language model for the Berkeley Restaurant Project corpus and evaluate it using the First-order Markov Assumption, POS tagging accuracy, sequence inference, and large-scale language modelling concepts.	16	BTL6	Creating

UNIT III SYNTACTIC ANALYSIS

	Context-Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing, Ambiguity, Dynamic Programming parsing – Shallow parsing – Probabilistic CFG, Probabilistic CYK, Probabilistic Lexicalized CFGs - Feature structures, Unification of feature structures.
--	--

PART – A

Q. No.	Questions	BT Level	Competence
1	Define Context-Free Grammar (CFG).	BTL1	Remembering
2	Differentiate between Context-Free Grammar (CFG) and Regular Grammar.	BTL2	Understanding

3	What is the purpose of a Treebank in Natural Language Processing?		BTL1	Remembering
4	Differentiate between Dependency Grammar and Phrase Structure Grammar.		BTL2	Understanding
5	Describe syntactic parsing.		BTL2	Understanding
6	What is syntactic ambiguity?		BTL1	Remembering
7	Describe the role of dynamic programming in syntactic parsing.		BTL2	Understanding
8	What is the role of probability in Probabilistic Context-Free Grammar (PCFG)?		BTL1	Remembering
9	Describe feature structures in syntactic analysis.		BTL2	Understanding
10	Describe the purpose of unification of feature structures in linguistic analysis.		BTL2	Understanding
11	What is left recursion in Context-Free Grammars?		BTL1	Remembering
12	Differentiate the Earley parser from other parsing algorithms.		BTL2	Understanding
13	Define LALR parsing.		BTL1	Remembering
14	List the basic steps of the CYK algorithm.		BTL1	Remembering
15	Describe the applications of probabilistic parsing in Natural Language Processing.		BTL2	Understanding
16	Describe lexical ambiguity and its challenges in syntactic analysis.		BTL2	Understanding
17	Describe the role of probabilistic parsing in speech recognition systems.		BTL2	Understanding
18	Define chart parsing.		BTL1	Remembering
19	Describe common strategies for resolving parsing ambiguity in Natural Language Processing.		BTL2	Understanding
20	Describe the metrics used to evaluate probabilistic parsing algorithms.		BTL2	Understanding
21	Define shallow parsing.		BTL1	Remembering
22	Describe the significance of probabilistic parsing.		BTL2	Understanding
23	Describe the basic steps of the CYK algorithm and state its time complexity.		BTL2	Understanding
24	Differentiate between Top-down Parsing and Bottom-up Parsing.		BTL2	Understanding
PART – B				
1	Apply Context-Free Grammars (CFGs) to represent the syntax of natural language sentences and construct a suitable CFG for a given example.	16	BTL3	Applying
2	Analyze the grammar rules used in English language processing and demonstrate their representation using Context-Free Grammars.	16	BTL4	Analyzing
3	Evaluate the role of Treebanks in syntactic analysis and assess the information captured during Treebank construction.	16	BTL5	Evaluating
4	Analyze the transformation of grammars into Chomsky Normal Form (CNF) and Greibach Normal Form (GNF), and examine their impact on parsing efficiency.	16	BTL4	Analyzing
5	Compare the different types of Dependency Graphs and analyze their significance in dependency parsing.	16	BTL4	Analyzing
6	Apply syntactic parsing techniques to parse sentences and demonstrate the working of different parsing algorithms.	16	BTL3	Applying
7	Evaluate the contribution of dynamic programming to syntactic parsing and justify its effectiveness in improving parsing efficiency.	16	BTL5	Evaluating
8	Apply shallow parsing to identify syntactic constituents in a given text and illustrate its applications in Natural Language Processing.	16	BTL3	Applying
9	Analyze Probabilistic Context-Free Grammars (PCFGs) and examine how probability enhances syntactic parsing performance.	16	BTL4	Analyzing
10	Analyze the Probabilistic Cocke–Younger–Kasami (CYK) parsing algorithm and compare it with the traditional CYK algorithm.	16	BTL4	Analyzing
11	Apply feature structures to represent linguistic information and demonstrate the process of feature unification with suitable examples.	16	BTL3	Applying
12	Evaluate Probabilistic Lexicalized Context-Free Grammars	16	BTL5	Evaluating

	(PLCFGs) and assess their effectiveness in syntactic analysis.			
13	Compare and evaluate probabilistic parsing strategies, including Viterbi Parsing and Earley Parsing, highlighting their advantages and limitations.	16	BTL5	Evaluating
14	Design Context-Free Grammar rules to define the syntax of programming language constructs with suitable examples.	16	BTL6	Creating
15	Apply techniques for resolving lexical ambiguity during syntactic analysis and demonstrate their effect on parsing accuracy.	16	BTL3	Applying
16	Design a syntactic analysis framework that integrates syntax and semantics to improve Natural Language Processing tasks.	16	BTL6	Creating
17	Construct the equivalent Chomsky Normal Form (CNF) grammar for the following Context-Free Grammars (CFGs) by applying the appropriate transformation rules. $G_1 = \{ S \rightarrow AB \mid c, A \rightarrow a, B \rightarrow b \}$ $G_2 = \{ S \rightarrow aA, A \rightarrow a, B \rightarrow c \}$	16	BTL6	Creating

UNIT IV SEMANTICS AND PRAGMATICS

Requirements for representation, Description Logics-Syntax-Driven Semantic analysis, Semantic attachments -Word Senses, Relations between Senses, Thematic Roles, Selectional Restrictions-Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods -Word Similarity using Thesaurus and Distributional methods.

PART – A

Q. No.	Questions	BT Level	Competence
1	What is semantic analysis?	BTL1	Remembering
2	State why Semantic Analysis is difficult?	BTL2	Understanding
3	Differentiate WordNet and Wordsense?	BTL2	Understanding
4	What are the basic elements of First Order logic?	BTL1	Remembering
5	Define Lambda notation.	BTL1	Remembering
6	Give the ways in which Modus Ponens can be used?	BTL2	Understanding
7	What are the two operators of FOL?	BTL2	Understanding
8	Difference between Polysemy and Homonymy.	BTL2	Understanding
9	How does word senses can be related?	BTL2	Understanding
10	What is WSD?	BTL2	Understanding
11	What is First-Order Logic, and how is it utilized in natural language processing?	BTL1	Remembering
12	Discuss how semantic attachments aid in disambiguating word senses?	BTL2	Understanding
13	What is the role of Description Logics in semantic analysis?	BTL1	Remembering
14	Discuss how senses contribute to a more comprehensive semantic understanding?	BTL2	Understanding
15	Discuss how dictionaries and thesauri be utilized in Word Sense Disambiguation?	BTL2	Understanding
16	Differentiate word similarity using thesaurus and distributional methods.	BTL2	Understanding
17	What are the key requirements for semantic representation in natural language processing?	BTL1	Remembering
18	What challenges may arise in supervised WSD, especially when dealing with new or rare word senses?	BTL1	Remembering
19	What advantages does bootstrapping offer in the context of WSD, and how does it address the data labeling challenge?	BTL1	Remembering
20	What are potential drawbacks of relying solely on a thesaurus for measuring word similarity?	BTL1	Remembering
21	How do distributional methods measure word similarity, and what	BTL2	Understanding

	information do they rely on?			
22	What is Pragmatics?		BTL1	Remembering
23	Write the WSD algorithms.		BTL2	Understanding
24	Define the term word sense.		BTL6	Remembering
PART – B				
1	(i)What is WordNet? How is "sense" defined in WordNet? Examine with example. (ii)Explain the fundamental requirements for representation in natural language processing.	8 8	BTL5	Evaluating
2	Explain Lesk Algorithm for WSD with suitable example. A knowledge dictionary based approach	16	BTL5	Evaluating
3	What do you mean by word sense disambiguation (WSD)? Analyse the machine learning based (Navie based) approach for WSD	16	BTL4	Analyzing
4	Explain with suitable example following relationships between word meanings: Homonymy, Polysemy, Synonymy, Antonymy, Hypernymy, Hyponymy. Discuss in detail attachments for fragments of English sentences.	16	BTL5	Evaluating
5	What is Word sense disambiguation? Explain different ways to identify correct sense of an ambiguous word.	16	BTL5	Evaluating
6	Explain various approaches to semantic analysis. Discuss different semantic relationships between the words.	16	BTL5	Evaluating
7	Analyze Pragmatics with suitable example. And discuss the implementation phases of language processing	16	BTL4	Analyzing
8	Explain how a supervised learning algorithm can be applied for word sense disambiguation.	16	BTL4	Analyzing
9	(i)Write short notes on Word Senses. (ii)Give FOL translations for the following sentences: 1. Vegetarians do not eat meat. 2. Not all vegetarians eat eggs.	8 8	BTL3	Applying
10	(i)Explain the relationships between different word senses. (ii)Define and elaborate on thematic roles in the context of semantics.	6 7	BTL5	Evaluating
11	(i)Define Word Sense Disambiguation (WSD) and its importance in NLP. (ii)Discuss the challenges of WSD and potential applications. Provide examples of successful WSD techniques.	16	BTL4	Analyzing
12	List out the various algorithms used for WSD. Evaluate with example.	16	BTL5	Evaluating
13	Define distributional methods for measuring word similarity. Discuss the theoretical foundations of distributional semantics.	16	BTL3	Applying
14	(i)Explore the relationships between different word senses. (ii)Define and elaborate on thematic roles in the context of semantics.	8 8	BTL6	Creating
15	Analyze how thematic roles contribute to a deeper understanding of linguistic meaning.	16	BTL4	Analyzing
16	Define and explain the principles of First-Order Logic (FOL).Discuss the role of FOL in semantic analysis and representation. Provide examples illustrating the application of FOL in natural language understanding.	16	BTL3	Applying
17	What is Word Sense Disambiguation (WSD)? Explain its importance and describe the dictionary-based approach, including the Lesk algorithm, with an example	16	BTL4	Analyzing
UNIT V DISCOURSE ANALYSIS AND LEXICAL RESOURCES				
Discourse segmentation - Coherence –Reference Phenomena - Anaphora Resolution using Hobbs and Centering Algorithm –Coreference Resolution –Resources: Porter Stemmer - Lemmatizer, Penn Treebank, Brill's Tagger,				

	Word Net, Prop Bank, Frame Net, Brown Corpus, British National Corpus (BNC).			
	PART – A			
Q. No.	Questions		BT Level	Competence
1	Define discourse model.		BTL1	Remembering
2	What are the types of referring expression?		BTL1	Remembering
3	What is reference resolution?		BTL2	Understanding
4	Define discourse processing?		BTL1	Remembering
5	Name two algorithms for Anaphora Resolution.		BTL2	Understanding
6	What do you mean by anaphora and anaphor?		BTL2	Understanding
7	Give the properties of Coreference relation.		BTL2	Understanding
8	Discuss how Penn Treebank contribute to NLP research?		BTL2	Understanding
9	Differentiate tasks and dataset?		BTL2	Understanding
10	What is the purpose of Lemmatization in NLP?		BTL2	Understanding
11	List the stages of RST parsing.		BTL1	Remembering
12	What is Co-reference Resolution, and why is it important in NLP?		BTL1	Remembering
13	List the examples of resources used in NLP for text processing.		BTL1	Remembering
14	Discuss how does Anaphora Resolution contribute to text understanding?		BTL2	Understanding
15	What is the significance of Coherence in discourse analysis?		BTL1	Remembering
16	Define Discourse Segmentation in Natural Language Processing (NLP).What is the purpose of Lemmatization in NLP?		BTL1	Remembering
17	What is a Porter Stemmer?		BTL1	Remembering
18	Differentiate between Written and Spoken Data in the British National Corpus		BTL2	Understanding
19	Define semantic role labeling and its relevance in natural language processing.		BTL1	Remembering
20	Discuss how FrameNet contribute to capturing meaning beyond individual words?		BTL2	Understanding
21	What is Entity based Coherence?		BTL1	Remembering
22	Define the terms: pathos, ethos and logos		BTL2	Understanding
23	What is the significance of Brill's tagger?		BTL1	Remembering
24	Define Prop Bank.		BTL1	Remembering
	PART – B			
1	Analyze the different types of referring expressions and classify them with suitable examples..	16	BTL4	Analyzing
2	Evaluate Hobbs' Algorithm for anaphora resolution by applying it to a sample discourse..	16	BTL5	Evaluating
3	(i) Analyze the process of discourse segmentation in NLP. (ii) Examine the challenges associated with discourse segmentation using suitable examples.	8 8	BTL4	Analyzing
4	(i) Analyze different strategies employed to achieve coherence within a text. (ii) Evaluate an incoherent discourse and propose suitable improvements to enhance coherence.	8 8	BTL4	Analyzing
5	Apply the Centering Algorithm to resolve anaphora in a given discourse and analyze the results.	16	BTL3	Applying
6	Evaluate the strengths and limitations of the Hobbs and Centering algorithms for anaphora resolution.	16	BTL5	Evaluating
7	Design a discourse processing framework incorporating Penn Treebank as a lexical resource, and justify its role in NLP applications.	16	BTL6	Creating
8	(i)Analyze the functionality of Brill's Tagger in part-of-speech tagging (ii)Compare the strengths and weaknesses of Penn Treebank and Brill's Tagger.	8 8	BTL4	Analyzing

9	Define Coreference resolution. Describe how the reference expressions are used to evoke and access entities in the discourse model.	16	BTL3	Applying
10	Discuss real-world applications of anaphora resolution using Hobbs and Centering algorithms.	16	BTL3	Applying
11	Discuss how Word Net enhances semantic understanding in natural language processing.	16	BTL4	Analyzing
12	Explain the role of speech processing in the broader context of natural language processing.	16	BTL5	Evaluating
13	(i)Explore the intersections between discourse analysis and lexical resources in natural language processing. (ii)Discuss how the insights gained from discourse analysis can enhance the development and utilization of lexical resources.	8 8	BTL6	Creating
14	(i)Discuss challenges in co-reference resolution and propose solutions. (ii)Provide examples where co-reference resolution plays a crucial role in text understanding.	8 8	BTL5	Evaluating
15	(i)Explore evaluation measures for lexical resources, including precision, recall, and F1-score. (ii)Compare and contrast the challenges in evaluating discourse analysis and lexical resource systems.	16	BTL5	Evaluating
16	Provide examples of how lexical resources contribute to discourse-level understanding and coherence.	16	BTL5	Evaluating
17	Describe the Hobbs algorithm for anaphora resolution. Provide step-by-step details.	16	BTL3	Applying