

**Duration: 3hrs**

**[Max Marks: 80]**

- N.B.:** (1) Question No 1 is Compulsory.  
(2) Attempt any three questions out of the remaining five.  
(3) All questions carry equal marks.  
(4) Assume suitable data, if required, and state it clearly.

- Q1a)** Explain the applications of Natural Language processing. **5M**
- Q1b)** Illustrate the concept of tokenization and stemming in Natural Language processing. **5M**
- Q1c)** Discuss the challenges in part of speech tagging. **5M**
- Q1d)** Describe the semantic analysis in Natural Language processing. **5M**
- Q2a)** Explain inflectional and derivational morphology with an example **10M**
- Q2b)** Illustrate the working of Porter stemmer algorithm **10M**
- Q3a)** Explain hidden markov model for POS based tagging. **10M**
- Q3b)** Demonstrate the concept of conditional Random field in NLP **10M**
- Q4a)** Explain the Lesk algorithm for Word Sense Disambiguation. **10M**
- Q4b)** Demonstrate lexical semantic analysis using an example **10M**
- Q5a)** Illustrate the reference phenomena for solving the pronoun problem **10M**
- Q5b)** Explain Anaphora Resolution using Hobbs and Canterling Algorithm **10M**
- Q6a)** Demonstrate the working of machine translation systems **10M**
- Q6b)** Explain the Information retrieval system **10M**
-

**Duration: 3hrs**

**[Max Marks: 80]**

- N.B.:** (1) Question No 1 is Compulsory.  
 (2) Attempt any three questions out of the remaining five.  
 (3) All questions carry equal marks.  
 (4) Assume suitable data, if required and state it clearly.

- 1 Attempt any FOUR [20]
  - a What is the rule-based and stochastic part of speech taggers?
  - b Explain Good Turing Discounting?
  - c Explain statistical approach for machine translation.
  - d Explain with suitable example the following relationships between word meanings: Hyponymy, Hypernymy, Meronymy, Holonymy
  - e What is reference resolution?
- 2 a Explain FSA for nouns and verbs. Also Design a Finite State Automata (FSA) for the words of English numbers 1-99. [10]
  - b Discuss the challenges in various stages of natural language processing. [10]
- 3 a Consider the following corpus [10]

`<s> the/DT students/NN pass/V the/DT test/NN</s>`  
`<s> the/DT students/NN wait/V for/P the/DT result/NN</s>`  
`<s> teachers/NN test/V students/NN</s>`

Compute the emission and transition probabilities for a bigram HMM. Also decode the following sentence using Viterbi algorithm.  
**“The students wait for the test”**

  - b What are five types of referring expressions? Explain with the help of example. [10]
- 4 a Explain dictionary-based approach (Lesk algorithm) for word sense disambiguation (WSD) with suitable example. [10]
  - b Explain the various challenges in POS tagging. [10]
- 5 a Explain Porter Stemming algorithm in detail. [10]
  - b Explain the use of Probabilistic Context Free Grammar (PCFG) in natural language processing with example. [10]
- 6 a Explain Question Answering system (QAS) in detail. [10]
  - b Explain how Conditional Random Field (CRF) is used for sequence labeling. [10]

\*\*\*\*\*

Time: 3 hours

Max. Marks: 80

N.B. (1) Question No. 1 is compulsory

(2) Assume suitable data if necessary

(3) Attempt any three questions from the remaining questions

**Q.1 Solve any Four out of Five**

**5 marks each**

- Explain the challenges of Natural Language processing.
- Explain how N-gram model is used in spelling correction
- Explain three types of referents that complicate the reference resolution problem.
- Explain Machine Translation Approaches used in NLP.
- Explain the various stages of Natural Language processing.

**Q.2 10 marks each**

- What is Word Sense Disambiguation (WSD)? Explain the dictionary based approach to Word Sense Disambiguation.
- Represent output of morphological analysis for Regular verb, Irregular verb, singular noun, plural noun Also Explain Role of FST in Morphological Parsing with an example

**Q.3 10 marks each**

- Explain the ambiguities associated at each level with example for Natural Language processing.
- Explain Discourse reference resolution in detail.

**Q.4 10 marks each**

a	<S>	Martin	Justin	can	watch	Will	<E>
	<S>	Spot	will	watch	Martin	<E>	
	<S>	Will	Justin	spot	Martin	<E>	
	<S>	Martin	will	pat	Spot	<E>	

For given above corpus,

N: Noun [Martin, Justin, Will, Spot, Pat]

M: Modal verb [can, will]

V: Verb [watch, spot, pat]

Create Transition Matrix & Emission Probability Matrix

Statement is "Justin will spot Will"

Apply Hidden Markov Model and do POS tagging for given statements

- b Describe in detail Centering Algorithm for reference resolution.

**Q.5 10 marks each**

- a For a given grammar using CYK or CKY algorithm parse the statement

"The man read this book"

Rules:

$S \rightarrow NP VP$	$Det \rightarrow that   this   a   the$
$S \rightarrow Aux NP VP$	$Noun \rightarrow book   flight   meal   man$
$S \rightarrow VP$	$Verb \rightarrow book   include   read$
$NP \rightarrow Det NOM$	$Aux \rightarrow does$
$NOM \rightarrow Noun$	
$NOM \rightarrow Noun NOM$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb NP$	

- b Explain Porter Stemmer algorithm with rules

**Q.6 10 marks each**

- a Explain information retrieval versus Information extraction systems  
b Explain Maximum Entropy Model for POS Tagging



Time: 3 Hours

Max. Marks: 80

- N.B. (1) Question No. 1 is compulsory  
 (2) Assume suitable data if necessary  
 (3) Attempt any three questions from remaining questions

- Q.1** Any Four **20[M]**
- a** Differentiate between Syntactic ambiguity and Lexical Ambiguity. **[5M]**
  - b** Define affixes. Explain the types of affixes. **[5M]**
  - c** Describe open class words and closed class words in English with examples. **[5M]**
  - d** What is rule base machine translation? **[5M]**
  - e** Explain with suitable example following relationships between word meanings. **[5M]**  
 Homonymy, Polysemy, Synonymy, Antonymy
  - f** Explain perplexity of any language model. **[5M]**
- Q.2 a)** Explain the role of FSA in morphological analysis? **[5M]**
- Q.2 b)** Explain Different stage involved in NLP process with suitable example. **[10M]**
- Q.3 a)** Consider the following corpus **[5M]**
- <s> I tell you to sleep and rest </s>  
 <s> I would like to sleep for an hour </s>  
 <s> Sleep helps one to relax </s>
- List all possible bigrams. Compute conditional probabilities and predict the next ord for the word “to”.
- Q.3 b)** Explain Yarowsky bootstrapping approach of semi supervised learning **[5M]**
- Q.3 c)** What is POS tagging? Discuss various challenges faced by POS tagging. **[10M]**
- Q.4 a)** What are the limitations of Hidden Markov Model? **[5M]**
- Q.4 b)** Explain the different steps in text processing for Information Retrieval **[5M]**
- Q.4 c)** Compare top-down and bottom-up approach of parsing with example. **[10M]**
- Q.5 a)** What do you mean by word sense disambiguation (WSD)? Discuss dictionary based approach for WSD. **[10M]**
- Q.5 b)** Explain Hobbs algorithm for pronoun resolution. **[10M]**
- Q.6 a)** Explain Text summarization in detail. **[10M]**
- Q.6 b)** Explain Porter Stemming algorithm in detail **[10M]**

\*\*\*\*\*