

IMSc DS Sem.-8 Examination

CC 414

Advanced NLP

Time : 2-30 Hours]

April-2025

[Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

- Q.1**
- (a) What is Natural Language Processing (NLP)? What are the key goals and objectives of NLP? (07)
- (b) What are the steps of text Normalization? Explain them in brief. (07)
- OR**
- (a) Discuss various challenges of NLP with proper justification. (07)
- (b) What are the main components and tasks involved in NLP? (07)
- Q.2**
- (a) What is the role of smoothing in language models? Explain why smoothing is necessary and describe some common smoothing techniques used in language modeling. (07)
- (b) Given corpus: (07)
1. "Ramesh likes coding"
 2. "Will Ramesh search coding"
 3. "Can Ankit search coding"
 4. "Ankit likes Will"
- Construct POS tagging using HMM (Hidden Markov Model) for the sentence: "Will will search coding".
- OR**
- (a) Given Sentences: (07)
1. "I like natural language processing."
 2. "Natural language processing is exciting."
 3. "I like learning new things."
- What is the probability of the sentence "I like natural language" using a bigram model? Assume that a start token (<s>) is used at the beginning of each sentence.
- (b) Differentiate between CRF (Conditional Random Field) and HMM (Hidden Markov Model). (07)
- Q.3**
- (a) Explain TFIDF with its applications. (07)
- (b) Explain Skip Gram with diagram. (07)
- OR**
- (a) What do you mean by word sense disambiguation (WSD)? Explain Lesk Algorithm for WSD with suitable example. (07)
- (b) Create a document vector table for the given corpus: (07)
- Document 1: We are going to Mumbai
- Document 2: Mumbai is a famous place.
- Document 3: We are going to a famous place.
- Document 4: I am famous in Mumbai.

- Q.4** (a) Compare and contrast IBM Model 1 Vs. IBM Model 2. (07)
- (b) How does a phrase-based model improve translation accuracy compared to a word-based model? (07)

OR

- (a) Explain the types of Machine Translation System. (07)
- (b) Write alignment function for foreign language words $f = (f_1, f_2, \dots, f_5)$ and English language words $e = (e_1, e_2, \dots, e_9)$ from the following figure. Also, show which and how many phrase pairs can be extracted? (07)

	f_1	f_2	f_3	f_4	f_5
e_1					
e_2					
e_3					
e_4					
e_5					
e_6					
e_7					
e_8					
e_9					

- Q.5** Attempt any **SEVEN** out of **TWELVE**: (14)
- (1) Explain lemmatization with an example.
 - (2) What is the purpose of removing stopwords?
 - (3) Explain Homonym with an example.
 - (4) What is the purpose of named entity recognition in NLP?
 - (5) Give an example of a word with multiple senses along with their glosses.
 - (6) Assuming that a language model assigns the following conditional probabilities to a 4-word sentence $(S)=0.01212$. What is the perplexity?
 - (7) Explain adequacy and fluency in the context of machine translation.
 - (8) Explain Hypernyms with an example.
 - (9) Explain Polysemy with an example.
 - (10) Explain syntactic ambiguity with example.
 - (11) Write the log-linear model for a phrase-based machine translation system.
 - (12) Draw Vauquois diagram stating various approaches to machine translation.
