Name: **Enrolment No:**	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2018

Course: CSAI7006 (Natural Language Processing) **Semester: II**
Program: M.Tech.-CS –AI
Time: 03 hrs. **Max. Marks: 100**

Instructions: Attempt all the questions. Clearly state any assumption if you made.

SECTION A

S. No.		Marks	CO
Q1	Distinguish the term "anaphora" and "cataphora" with example.	**4**	**CO1**
Q2	What is the difference between homonymy and polysemy? Give an example of each that illustrates your point.	**4**	**CO3**
Q3	Write a regular expression to find all instances of the determiner "the" in a given sentence.	**4**	**CO2**
Q4	Discuss the various challenges involved in syntactic processing with suitable examples.	**4**	**CO2**
Q5	Illustrate the difference between Generative model and Discriminative model.	**4**	**CO3**
	SECTION B		
Q6	Discuss the application of Hidden Markov Model to solve the Tagging problem.	**10**	**CO4**
Q7	Illustrate the following term:- "Perspectives of NLP: Areas of AI and their inter dependencies"	**10**	**CO1**
Q8	Comprehend the major application area of NLP. And also discuss the major challenges in each of the application area.	**10**	**CO1**
Q9	Consider the following grammar : S → NP VP, VP → Verb NP, VP → Verb PP, VP → VP PP, NP → NP and NP, PP → P NP, NP → Ram, NP → Delhi, NP → Dehradun, NP → May, Verb → flew, P → in, P → to, CONJ → and Show the various steps of the CKY parser that generates the sentence "Ram flew to Dehradun and Delhi in May".	**10**	**CO4**
	SECTION-C		
Q10	a) Consider the weight term vectors of two documents as: $D_1 = 2T_1 + 3T_2 + 5T_3$, $D_2 = 3T_1 + 7T_2 + 1T_3$ For a query $Q = 5T_1+5T_2+2T_3$, compute the similarity using inner product and cosine similarity metrics. With respect to this problem, which one is better measurement? b) Given a document containing terms with given frequencies: A(3), B(2), C(1).	**20**	**CO5**

	Assume that collection contains 10,000 documents and document frequencies of these terms are: A(50), B(1300), C(250). Computer the term frequency–inverse document frequency of the document collection.		
Q11	Discuss the challenges of Name Entity Recognition (NER)? Show with examples how global knowledge helps in better classification of named entity. How does NER and Coreference resolution interact other? Or, Write short notes on the following term. a) Co-reference resolution b)Discourse c) Pragmatics d) POS Tagging e)Perplexity	**20**	**CO1/CO2/CO3**