



PES Institute of Technology, Bangalore
(Autonomous Institute under VTU, Belgaum)

DECEMBER 2011 SEMESTER END EXAMINATION (SEE) B. E. 7th SEMESTER CS & IS

08CS423 - Natural Language Processing

Time: 3 Hrs

Answer All Questions

Max Marks: 100

1.	a)	What are the three basic needs to build a morphological parser?	05
	b)	Derive the Add-One smoothed bigram probability.	06
	c)	What is deleted interpretation technique for computing trigrams?	06
	d)	What are the practical use cases of POS tagging?	03
2.	a)	Consider the experiment of four successive tosses of a fair coin. Let X be the random variable that denotes the number of heads in each experiment. How many bits (averaged minimum) are required to transmit X, if the experiment is repeated forever?	04
	b)	Augment $S \rightarrow Aux NP VP$, to handle "AGREEMENT" feature for the verbs "do/does/did"	04
	c)	Provide precise pseudo code for Predictor, Scanner and Completer procedures in Earley's parser.	06
	d)	Draw the parse tree for the sentence(Assume a standard English grammar fragment) : 'Purchase a ticket from Bengaluru to Chennai in Brindavan Express'	06
3.	a)	Explain the lambda notation and its utility in FOPC with a couple of examples.	04
	b)	Construct an FOPC sentence for the sentence "I told Harry to go to the rail station yesterday".	04
	c)	A travel related website has information on 250 hill stations. Of these 50 are close to Kudremukh. The website also has information on 20 resorts owned and run by a company M/S Hill Station County Resorts Pvt Ltd. For a user query "Which are the hill stations close to Kudremukh?", Bing and Google return 65 & 72 distinct results that also contain 32 & 38 inappropriate ones respectively. Assume a Beta value of 2.0. Evaluate Google and Bing search engines on the four standard measures.	06
	d)	Clearly define the four types of relations between lexemes and their senses?	06
4.	a)	What are the four preprocessing steps to carry out robust word sense disambiguation?	04
	b)	What are the four approaches to word sense disambiguation?	06
	c)	For the terms (Computing, Multicore, Algorithms, Processor) in documents Doc-A, Doc-B and Doc-C the absolute term frequencies are (4, 5, 6, 3), (4, 20, 2, 18) and (2, 21, 27, 8) respectively. Compute the cosine similarity measures and find out the Document retrieved for the query vectors: i) (Computing, Multicore , Algorithms , Processor) ii) (Multicore , Algorithms) iii)(Computing, Multicore, Processor)	07
	d)	What is the standard relevance feedback update formula for document retrieval?	03
5.	a)	What are the salience factors that drive the Lappin-Leass Algorithm for discourse level pronoun resolution? Write Clear pseudo-code for Lappin-Leass Algorithm.	08
	b)	Construct salience tables to resolve pronouns in the following referring expressions R1: Ram bought it in Belgium. R2: He showed the gem to Ravi when he returned to India. R3: Ravi said, he has never seen a gem like that before.	08
	c)	Write brief notes on the following discourse reference phenomena with proper examples: i)Anaphora, ii) Cataphora iii) Discontinuous Sets iii) Definite Noun Phrase	04